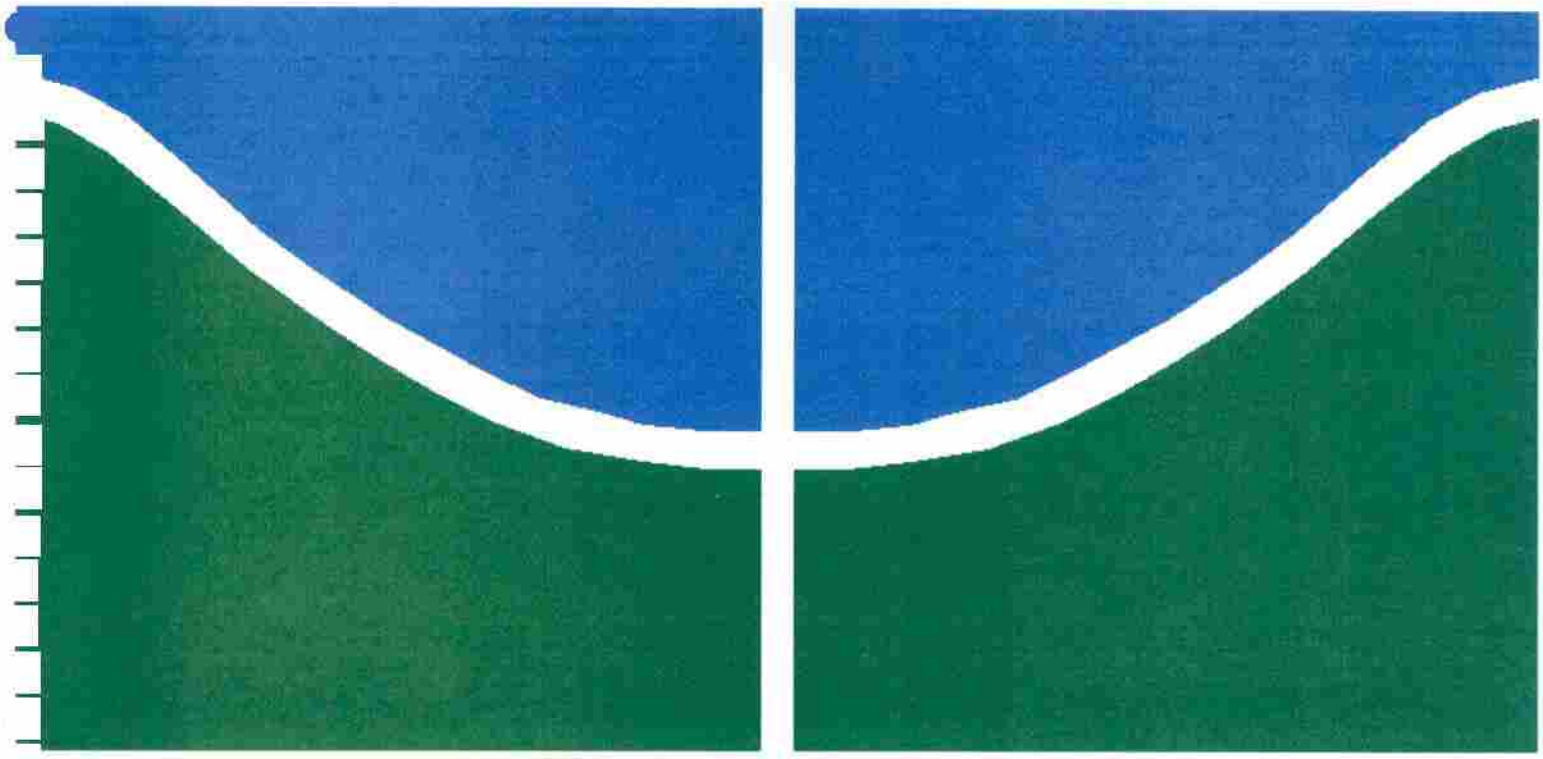


ESTUDO DE REDES DE ANONIMATO

SAMARA BITTENCOURT AMUI DE OLIVEIRA

DISSERTAÇÃO DE PÓS GRADUAÇÃO EM GESTÃO DE SEGURANÇA DA INFORMAÇÃO EM ENGENHARIA ELÉTRICA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

FACULDADE DE TECNOLOGIA

UNIVERSIDADE DE BRASÍLIA

UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

ESTUDO DE REDES DE ANONIMATO

SAMARA BITTENCOURT AMUI DE OLIVEIRA

Orientador: PROF. DR. LAERTE PEOTTA DE MELO, ENE/UNB

DISSERTAÇÃO DE PÓS GRADUAÇÃO EM GESTÃO DE SEGURANÇA DA INFORMAÇÃO EM ENGENHARIA ELÉTRICA

PUBLICAÇÃO PPGENE.DM - XXX/2017

BRASÍLIA-DF, XX DE JULHO DE 2017.

UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA ELÉTRICA

ESTUDO DE REDES DE ANONIMATO

SAMARA BITTENCOURT AMUI DE OLIVEIRA

DISSERTAÇÃO DE PÓS GRADUAÇÃO EM GESTÃO DE SEGURANÇA DA INFORMAÇÃO ACADÊMICO SUBMETIDA AO DEPARTAMENTO DE ENGENHARIA ELÉTRICA DA FACULDADE DE TECNOLOGIA DA UNIVERSIDADE DE BRASÍLIA, COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE ESPECIALISTA EM ENGENHARIA ELÉTRICA.

APROVADA POR:

Prof. Dr. Laerte Peotta de Melo, ENE/UnB
Orientador

Prof. Fulano de Tal 2, ENE/UnB
Examinador interno

Prof. Fulano de Tal 3, ENE/UnB
Examinador interno

Prof. Fulano de Tal 4, EESC/USP
Examinador externo

BRASÍLIA, XX DE JULHO DE 2017.

FICHA CATALOGRÁFICA

SAMARA BITTENCOURT AMUI DE OLIVEIRA

ESTUDO DE REDES DE ANONIMATO

2017xv, XXp., 201x297 mm

(ENE/FT/UnB, Especialista, Engenharia Elétrica, 2017)

Dissertação de Pós Graduação em Gestão de Segurança da Informação - Universidade de Brasília

Faculdade de Tecnologia - Departamento de Engenharia Elétrica

REFERÊNCIA BIBLIOGRÁFICA

SAMARA BITTENCOURT AMUI DE OLIVEIRA (2017) ESTUDO DE REDES DE ANONIMATO. Dissertação de Pós Graduação em Gestão de Segurança da Informação em Engenharia Elétrica, Publicação xxx/2017, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, XXp.

CESSÃO DE DIREITOS

AUTOR: SAMARA BITTENCOURT AMUI DE OLIVEIRA
TÍTULO: ESTUDO DE REDES DE ANONIMATO.
GRAU: Especialista ANO: 2017

SAMARA BITTENCOURT AMUI DE OLIVEIRA
BRASÍLIA/DF.

Agradecimentos

Agradeço ao meu orientador, pelo conhecimento e experiência compartilhados durante meus estudos. Meus agradecimentos também à Universidade de Brasília, em especial à Faculdade de Tecnologia, pela possibilidade de me especializar sem abandonar a carreira profissional, conseguindo conciliar o trabalho e o estudo. Por fim agradeço a minha família e ao meu companheiro pela compreensão por minha ausência em reuniões e festas familiares e pelo apoio para me dedicar à pós-graduação.

Resumo

Este trabalho apresenta uma comparação entre as redes de anonimato TOR e I2P com análise quantitativa e qualitativa do nível de anonimato que cada uma pode oferecer. São apresentadas técnicas de prevenção de análise de tráfego, como o padding e rerroteamento, utilizadas de maneira conjunta por ferramentas de anonimato conhecidas, como TOR e I2P. Definições de anonimato, classificações de anonimato em escala gradual, bem como definições matemáticas para o cálculo do nível de anonimato nas redes são apresentados. A entropia de Shannon é utilizada como referência para o cálculo do nível de anonimato nas redes, considerando como variáveis a quantidade de usuários e a probabilidade de que o usuário seja o remetente real de uma mensagem observada, ou seja, o modelo de ameaça é fundamental para o estudo de quão vulnerável é um usuário em uma rede de anonimato e quanto de privacidade um usuário pode esperar ao utilizar essas redes.

Abstract

This work presents a quantitative and qualitative analysis comparison between two anonimity networks, TOR and I2P, showing the level of anonimity each one can provide. There are presented traffic analysis prevention techniques, like padding and rerouting, used combined by anonimity tools as TOR or I2P. This work defines anonimity, degree classifications for anonimity, mathematical definitions for anonimity level estimation are shown. Shannon Entropy is used as math reference for calculate anonimity level in TOR and I2P networks, assuming as variables the number of users and user's probability of being a true sender or receiver of a given message. For this, threat model study is a key that allow the ananlysis of how vulnerable a user is in those anonymity networks and how much privacy he has when using those networks.

SUMÁRIO

LISTA DE FIGURAS

Capítulo 1

Introdução

A segurança da informação é baseada em três pilares principais: Confidencialidade, Integridade e Disponibilidade. A confidencialidade é a propriedade que define que apenas usuários permitidos possam ter acesso a uma determinada informação. Integridade significa dizer que uma informação é íntegra, isto é, nunca foi modificada ou destruída. Disponibilidade é a propriedade de manter uma informação disponível e acessível pelo tempo que for necessário. Esses três pilares são fundamentais para a compreensão dos métodos de tratamento de informações sensíveis e para o estudo da segurança dessas informações.

Outra propriedade da segurança da informação, não menos importante, é o anonimato. O anonimato está relacionado ao nível de privacidade que um usuário pode ter ao lidar com informações, seja fornecendo ou acessando essas informações. Nos dias atuais, muitas pessoas têm toda sua vida exposta em meios eletrônicos, seja de maneira deliberada ou não. Muitas atividades envolvem o uso de meios eletrônicos ou digitais que exigem o cadastro de dados pessoais de usuários do mundo todo, e a maneira como esses dados são tratados é extremamente relevante para que essas atividades tenham sucesso e credibilidade.

São tantos os serviços fonecidos no meio digital que algumas entidades possuem dados de milhares, milhões ou até bilhões de pessoas pelo mundo. Governos possuem informações de pessoas de todo o mundo obtidas de maneira legal ou até por meio de vigilância. Também existem criminosos que conseguem várias informações por meio de ferramentas de fácil uso disponibilizadas na Internet. Com tanta exposição e tantos olhares vigilantes, o usuário deve se perguntar quão seguro está ao navegar na Internet e quanto de privacidade ele tem.

Assim, se um usuário vive em um local em que há vigilância por parte do Governo e censura em vários níveis, ele é um dos usuários que precisa confiar em alguma rede de anonimato que lhe disponibilize o maior nível de anonimato e confiança possível. Da mesma forma jornalistas que investigam sistemas e pessoas poderosas precisam de anonimato para publicar ou até encontrar fontes que colaborem com seu trabalho. Entidades militares também possuem grande interesse em meios de comunicação anônima e que forneça mais privacidade. Empresas e corporações também podem utilizar redes de anonimato para evitar que

concorrentes consigam analisar e prever ideias ou parcerias em vista. E claro, um usuário comum, que apenas deseje ter mais privacidade ao utilizar a Internet, sem se aborrecer com a quantidade de propagandas e manipulação de resultados feitas por meio de análise de tráfego e de metadados por vários sites e empresas no meio digital.

Existem várias redes e iniciativas de anonimato pela Internet, algumas mais conhecidas e difundidas, outras menores, mas não menos efetivas. Mas para saber se, de fato, essas redes que procuram fornecer anonimato podem te manter anônimo com um nível suficiente para sua segurança, é necessário analisar como a rede funciona. Alguns estudos avaliam de maneira quantitativa o nível de anonimato que uma rede pode oferecer dependendo de algumas variáveis, como quantidade de usuários conectados na rede e o modelo de ameaça existente. Todavia, deve-se sempre lembrar que, para cada tipo de ameaça ou ataque, uma rede pode ser melhor do que outra, portanto, o usuário deve buscar conhecer as opções de redes existentes e para qual finalidade cada uma se destaca, bem como os tipos de ataques aos quais os usuários dessas redes estão mais suscetíveis e quão vulnerável é cada uma.

Neste trabalho são apresentadas duas redes de anonimato conhecidas, a rede TOR e a rede I2P. As características técnicas de ambas são apresentadas e uma análise técnica do nível de anonimato que elas oferecem é feita a partir de estudos existentes por meio de simulações.

1.1 Objetivo

Este trabalho tem como objetivo analisar o nível de privacidade do usuário no uso da Internet, apresentando um estudo comparativo de ferramentas de anonimato existentes. O intuito é entender se há anonimato e quais os níveis de anonimato em sistemas atuais de comunicações pela Internet, buscando avaliar o custo benefício desses sistemas, seja tecnicamente, financeiramente ou socialmente.

1.2 Objetivos Específicos

Analisar quantitativamente e qualitativamente quão anônimo é um dado sistema de comunicação, tendo como base estudos e análises existentes, a fim de concluir o grau de anonimato desse sistema e verificar se, neste contexto, a privacidade do usuário que busca o anonimato é preservada.

1.3 Justificativas

Atualmente, a privacidade do usuário que utiliza a Internet vem sendo bastante discutida, sendo o anonimato um dos pontos da discussão. A Internet pode ser utilizada para várias

atividades, como comércio eletrônico, correio eletrônico, Internet Banking, moedas virtuais, entre outras, o que faz com que os usuários se preocupem mais com a segurança de seus dados e, consequentemente, que os provedores dos serviços busquem técnicas que ofereçam aos usuários mais segurança no trato de suas informações.

Existem vários motivos que levam um usuário da Internet a querer utilizar ferramentas e técnicas que permitam seu anonimato. A vigilância por parte de Estados, que pode variar desde interesses políticos ao combate ao terrorismo, ou mesmo a vigilância e observação de empresas com interesses comerciais, que trocam informações entre si e compram informações com as grandes empresas de tecnologia, criando a cada dia técnicas mais robustas para análise do comportamento de um usuário enquanto consumidor, dentre vários outros fatores particulares que podem variar de usuário para usuário, como por exemplo o caso de criminosos, que preferem adotar medidas de proteção de sua identidade a fim de, por meio da Internet, cometer os mais diversos crimes ou infrações.

É fácil apontar várias razões para que usuários, sejam eles pessoas comuns, empresas ou governos, escolham formas anônimas para se comunicar pela Internet. Entretanto, é fundamental destacar que qualquer técnica ou ferramenta escolhida possui custos computacionais, econômicos e político-administrativos.

O termo privacidade pode ser classificado objetivamente de quatro maneiras: em relação ao ID (identificação do usuário); localização; comportamento e conteúdo. Com base nessa classificação, para os casos em que os dados do usuário são indispensáveis para o provimento de um serviço, deve-se saber se há permissão para o uso desses dados e então apresentar uma política de privacidade do provedor para o usuário. Para que seja aplicada uma técnica que permita ao usuário manter sua privacidade em diferentes níveis, deve-se entender como os dados trafegam na rede, se há utilização de protocolos de segurança de rede, se há uso de criptografia e buscar mecanismos que busquem garantir a privacidade sem prejudicar o fornecimento do serviço.

A mesma preocupação se dá em relação a intermediários ou mesmo adversários que buscam obter informações de outros usuários sem que lhes seja dada a devida permissão, ou seja, não só a privacidade do usuário em relação aos seus provedores de serviços está em jogo mas, também, em relação a qualquer parte que esteja conectada à Internet. Nesse contexto, algumas técnicas de anonimato serão estudadas e analisadas quanto ao nível de anonimato oferecido e o custo benefício de seu uso, que pode ser interpretado financeiramente, computacionalmente ou mesmo o custo benefício no âmbito social e político.

Para isso será utilizada a definição de anonimato que diz que o anonimato é aquilo que permite que atores envolvidos escondam sua relação com ações particulares e resultados [Danezis 2004]. Ainda, existem dois tipos de anonimato, o anonimato do remetente (sender anonymity) e o anonimato do destinatário (recipient anonymity) [Danezis 2004], em que o primeiro tipo tem como característica esconder qualquer correspondência entre a informação enviada e sua origem, e o segundo busca mascarar a correspondência entre o receptor

e a mensagem recebida. Dessa forma, este trabalho apresentará conceitos e definições necessários para a compreensão da análise quantitativa e qualitativa dos sistemas e técnicas de anonimato existentes que serão apresentados nos capítulos seguintes.

1.4 Organização do Trabalho

Este trabalho é divido em cinco capítulos. Este Capítulo 1 apresenta uma definição de anonimato, contextualizando os motivos que levam um usuário a buscar o anonimato na rede. Apresenta também a motivação do estudo de técnicas de privacidade e anonimato, destacando a metodologia a ser utilizada e o objetivo final do trabalho.

O Capítulo 2 apresenta fundamentos e conceitos relacionados a segurança da informação. Uma breve introdução sobre algumas propriedades de segurança da informação e criptografia é apresentada. É feita uma contextualização e apresentação de definição de anonimato.

No Capítulo 3 são apresentadas as técnicas de anonimato *padding* e rerroteamento, com uma análise probabilística sobre o nível de anonimato que cada uma dessas técnicas ou sua combinação podem oferecer. Também são apresentadas algumas ferramentas de anonimato conhecidas, o TOR, Freenet e I2P, e como cada uma delas funciona tecnicamente para prover anonimato, e que tipo de anonimato elas podem oferecer.

O Capítulo 4 apresenta uma análise comparativa das ferramentas de anonimato apresentadas no Capítulo 3, apresentando resultados de nível de anonimato que cada ferramenta pode oferecer e para qual tipo de finalidade podem ser recomendadas de maneira a otimizar os níveis de privacidade desejados.

Por fim, o Capítulo 5 mostra a conclusão do trabalho após os resultados obtidos e deixa sugestões de trabalhos futuros relacionados ao tema.

Capítulo 2

Fundamentos e Conceitos

A segurança da informação possui três pilares principais: Confidencialidade, Integridade e Disponibilidade. Além dessas, várias outras propriedades têm ganhado espaço em discussões e pesquisas relacionadas ao tratamento da informação, entre elas as propriedades de Autenticidade, Não-Repúdio, Autenticação e Anonimato. Com a crescente conscientização sobre o valor e importância da informação, empresas, governos e indivíduos vêm buscando meios de armazenarem suas informações valiosas de maneira segura e de se comunicarem com privacidade, contra o ataque de adversários interessados em obter essa informação sem autorização, seja para benefício financeiro, seja para vigilância.

2.1 Propriedades de Segurança da Informação

Para entender como a comunicação entre dois pontos deve ser implementada para que se dê de forma mais segura e confiável, deve-se conhecer os conceitos das propriedades de segurança da informação. No entanto, mesmo que uma comunicação seja confiável e segura em alguns aspectos, ela ainda pode ser vulnerável a certos tipos de ataques que permitem que um observador adquira dados suficientes para analisar o comportamento de um usuário e categorizar seu perfil de acordo com premissas estabelecidas. Como por exemplo a relação entre o usuário e seu provedor de serviços de Internet. O provedor, mesmo que seja uma fonte confiável para o usuário, sabe exatamente por quais sites o seu usuário trafega, com qual frequência e em quais horários, o que é utilizado em países em que alguns tipos de conteúdos são censurados pelo governo e punidos os cidadãos que os acessarem. Ou, por exemplo, sites de comércio eletrônico, que mesmo que utilizem protocolos para a transmissão de dados pessoais do usuário, número de cartão de crédito, senhas, etc., de maneira confidencial e íntegra, observam as atividades do usuário em seu site, coletando informações de quais *links* ele acessou e quantas vezes, bem como fazem controle da oferta de produtos e preços baseado na localização geográfica do usuário, adquirida por meio de observação e análise de tráfego, por meio de metadados.

A análise de tráfego também pode ser muito prejudicial para setores empresariais. Em-

presas investem em segurança para que seus segredos comerciais não sejam expostos, e, por isso, também têm muito interesse em prevenir análise de tráfego, que pode comprometer negociações. Com uma simples análise de tráfego é possível identificar possíveis clientes ou parceiros de negócio de uma empresa; pode-se identificar em qual empresa um determinado indivíduo trabalha devido ao provedor de e-mail específico que foi acessado, o que pode ser perigoso dependendo do local de onde ele está acessando o e-mail, se for em viagem fora do seu país de origem.

2.1.1 Anonimato

Em [Danezis 2004] anonimato é definido como a propriedade que permite que atores envolvidos escondam sua relação com ações particulares e seus resultados, e pode ser classificado em dois tipos, anonimato do remetente e anonimato do destinatário. Ainda, em um canal de comunicação que oferça anonimato bidirecional, tanto o remetene quanto o destinatário podem estar anônimos e ainda assim serão capazes de trocar mensagens anônimas.

Os tipos de anonimato podem ser selecionados baseados nos modelos de ameaça existentes. A ameaça advinda de um adversário passivo global é diferente da ameaça oferecida por um adversário ativo. Um adversário passivo global é aquele que apenas monitora, enquanto o ativo interfere na rede.

Considere uma escala qualitativa de graduação de anonimato, com os seguintes graus em nível decrescente de anonimato [Reiter and Rubin 1998]:

- Privacidade Absoluta (*absolute privacy*);
- Acima de Suspeita (*beyond suspicion*);
- Provável Inocente (*probable innocence*);
- Possível Inocente (*possible innocence*);
- Exposto (*exposed*);
- Provadamente Exposto (*provably exposed*).

Privacidade absoluta é quando um atacante não pode, de maneira alguma, provar que há relação entre uma mensagem e um usuário. O grau de anonimato Acima de Suspeita significa que para um atacante a relação entre uma mensagem e um usuário é equiprovável para qualquer usuário na rede. Perante um atacante, um usuário é considerado Provável Inocente se, para uma dada mensagem, a probabilidade de que ele tenha relação com essa mensagem é a mesma de ele não ter relação com a mensagem. O usuário é Possível Inocente aos olhos de um atacante se existir alguma probabilidade de que outro usuário tenha relação com uma mensagem específica. Um usuário é exposto se ele pode ser relacionado a uma mensagem por um atacante. Provadamente exposto é um nível em que o atacante consegue relacionar

uma mensagem a um usuário específico e provar essa relação para os outros participantes da rede.

Para o caso de redes com nós não confiáveis ou subvertidos (subverted nodes), caso o grafo que representa a troca de chaves seja dividido por um nó confiável, um adversário não pode dizer quantos participantes enviaram a mensagem com uma probabilidade melhor do que uma distribuição uniforme aleatória. No entanto, Chaum define o conjunto de anonimato de cada mensagem como o conjunto de todos os participantes da rede, o que pode ser referido como o tamanho do anonimato, e tem sido utilizado como medida de anonimato. Alguns ataques, porém, podem dividir o conjunto de anonimato em redes DC (*dining cryptographers*), e permitem o atacante encontrar de quem das duas divisões a mensagem foi originada. No pior caso o atacante gerencia a divisão de conjuntos de maneira que o verdadeiro emissor é deixado em um conjunto; se esse conjunto é formado por um único participante, então pode-se dizer que o sistema não provê anonimato.

Anonimato também pode ser definido como o estado de não ser identificado com um conjunto de assuntos, o conjunto de anonimato [Pfitzmann and Kohntopp 2001]. A qualidade de um conjunto de anonimato não seria medida apenas por sua cardinalidade, mas também quanto mais forte o anonimato, maior é o respectivo conjunto e mais uniformemente distribuídos o emissor ou o receptor, respectivamente, dos assuntos desse conjunto são [Pfitzmann and Kohntopp 2001]. Assim, o anonimato é máximo quando a probabilidade é igual se diferentes potenciais emissores ou receptores possuem diferentes probabilidades associadas a eles.

2.2 Entropia

O nível de anonimato oferecido por um sistema pode ser definido pela quantidade de informação que falta para que um atacante identifique exclusivamente a relação entre uma ação e um ator específico [Danezis 2004]. Assim, assuma que um conjunto de anonimato possua N participantes e que p_i do participante i é a probabilidade de que i seja o remetente real de uma mensagem, então a entropia do anonimato é dada por [Murdoch]:

$$H(S) = -\sum_{i=1}^{N} p_i \log_2(p_i), \tag{2.1}$$

em que S é uma variável aleatória discreta com espaço amostral $\omega = s_1 ... s_n$ e seja $p_i = Pr[S = s_i]$.

Se a entropia dada pela Equação 2.1 for normalizada, tem-se [Murdoch]:

$$D(S) = \frac{H(S)}{\log_2 N}. \tag{2.2}$$

Percebe-se que o grau de anonimato depende da quantidade de informação que um atacante possui e, portanto, do modelo de ameaça. Se um atacante não possui conhecimento prévio sobre os nós remetentes ou destinatários ou nem mesmo das conexões entre eles e é apenas um observador fazendo análise de tráfego, então para ele uma mensagem que chega a um nó pode ter sido encaminhada de qualquer nó vizinho com mesma probabilidade, ou seja, a probabilidade $p_(i)$ de que o usuário s_i tenha enviado a mensagem é igual a probabilidade $p_(j)$ de que o usuário s_j a tenha enviado. Portanto, para N nós na rede e n mensagens observadas, tem-se que:

$$\sum_{i}^{N} p_i = 1, \tag{2.3}$$

$$p_i = \frac{1}{N}. \tag{2.4}$$

Se, entretanto, for considerado um cenário em que um adversário tem informação sobre a relação entre alguns nós conseguida por meio de análise de tráfego, então a distribuição de probabilidade não será uniforme e, portanto, pode-se utilizar a Entropia de Rényi para quantificar o nível de anonimato neste sistema [Murdoch]:

$$H_\alpha(S) = \frac{1}{1-\alpha} \log_2(\sum_{i=1}^{N} p_i^\alpha) \tag{2.5}$$

Para $\alpha = 0$, $\alpha \to 1$ e $\alpha \to \infty$, casos especiais da entropia de Rényi, tem-se que H_0 representará a quantidade de perguntas que um adversário deve fazer para eliminar metade dos nós como candidatos a serem os remetentes de uma dada mensagem; H_1 também representará a quantidade de perguntas feitas por um adversário para eliminar uma quantidade arbitrária de possíveis remententes entre os nós no conjunto de anonimato observado; e $H_{\to\infty}$ representa a segurança quando um adversário pode investigar um dos nós [Murdoch].

2.3 Técnicas de Prevenção de Análise de Tráfego

Sistemas de prevenção de análise de tráfego podem ser utilizados para evitar que observadores, por meio dos vários nós intermediários em uma rede, analisem o tráfego entre um ponto e outro. Assim, o anonimato pode ser considerado em um contexto coletivo, isto é, pode-se estimar quanto anonimato uma rede oferece aos seus usuários por utilizar sistemas de prevenção de análise de tráfego.

Em um conjunto de nós numa rede, em que há entre eles adversários de todos os tipos (adversários passivos, passivos globais ou ativos), usuários podem utilizar técnicas de prevenção de análise de tráfego que possibilitem o anonimato conjunto desses nós. Dessa forma, a fim de confundir o adversário e evitar que seu ataque tenha sucesso, podem-se adotar técnicas como o *padding* ou o rerroteamento ou uma combinação de ambas, alterando a característica de tráfego na rede, dificultando que o adversário adquira informações reais das

comunicações que ocorrem entre os nós na rede. Alguns dos efeitos esperados como resultados da utilização de uma técnica de prevenção de análise de tráfego são o aumento do valor total de tráfego na rede; acréscimo na carga de processamento criptográfico nos nós envolvidos; o mascaramento da origem e destino de mensagens individuais e maior probabilidade de que um padrão de tráfego seja considerado verdadeiro pelo adversário, o que dificulta a diferenciação de tráfegos reais dos tráfegos adaptados por meio de técnicas de prevenção de análise de tráfego [Newman et al. 2003].

Assuma, para fins de análise, uma rede ponto-a-ponto, em que todos os nós podem ser considerados como transmissores, receptores ou nós intermediários e que todos os enalces diretos entre um nó e outro possuem a mesma capacidade. Considere, também, que a análise de tráfego na rede é feito por meio de matrizes de tráfego (MT) N x N não-negativa, tal que $T[i, j]$ representa a quantidade de mensagens trocadas entre os nós i e j de uma dada rede, tal que:

$$MT_{N,N} = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix}. \tag{2.6}$$

Para simplificar a análise, considere que um nó não envia mensagens para ele mesmo, isto é, os elementos $a_{1,1}$, $a_{2,2}$ e $a_{3,3}$ possuirão valor 0 (zero).

2.3.1 Padding

O método Padding adiciona tráfego na rede, ou seja, adiciona mensagens ao tráfego de maneira que um adversário não saiba diferenciar o que é o tráfego real de mensagens entre os nós e o que não é, pois, para ele, todo o tráfego observado parece real. Assim, uma MT observada por um adversário sempre possuirá elementos maiores que a MT real, isto é:

$$MT_{real}[i, j] \leq MT_{observada}[i, j], \tag{2.7}$$

em que a matriz observada é não-negativa.

Suponha que, dado um conjunto de nós em que um nó não envia mensagem para ele mesmo, a matriz de tráfego observada por um adversário seja dada por:

$$MT_{observada} = \begin{bmatrix} 0 & 2 & 1 \\ 1 & 0 & 3 \\ 1 & 1 & 0 \end{bmatrix} \tag{2.8}$$

e cada elemento representa a quantidade de mensagens enviada de um nó para o outro. Percebe-se que o nó 1 enviou duas mensagens para o nó 2 e uma mensagem para o nó 3. O nó 2 enviou uma mensagem para o nó 1 e três mensagens para o nó 3 e o nó 3 enviou uma mensagem para o nó 1 e uma mensagem para o nó 2.

2.3.2 Rerroteamento

O rerroteamento é um método que, como o próprio nome indica, tem como função modificar o caminho pelo qual a mensagem seguirá até seu destino, isto é, os pacotes e mensagens são roteados mais de uma vez na rede até chegarem a seu destino final por uma rota diferente da esperada. O tráfego resultante do rerroteamento será maior que o tráfego real, assim como ocorre com o *padding*, e pode ser representado por, [Newman et al. 2003]:

$$MT_{observada} = MT_{real} + D_{roteamento}, \tag{2.9}$$

em que $D_{roteamento}$ representa a matriz diferença do rerroteamento.

Uma matriz de rerroteamento unitário é a matriz U com N nós que representa o rerroteamento de uma unidade de tráfego entre o nó a e o nó N. A matriz U com três nós pode ser, então, representada da seguinte maneira:

$$U = \begin{bmatrix} 0 & 1 & -1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}. \tag{2.10}$$

Pode-se notar que, para uma unidade de tráfego entre os nós a e c sendo rerroteada por meio de b, a matriz unitária de tráfego U possui um elemento negativo, que representa um decréscimo no tráfego direto entre os nós a e c e dois elementos unitários positivos, que representam um acréscimo no tráfego entre os nós a e b e entre os nós b e c. Nota-se, ainda, que o resultado da soma entre os elementos de cada linha e coluna é igual a zero, exceto na linha e coluna que representa o nó intermediário. Portanto, a carga de tráfego adicionada devido a um rerroteamento é igual a 1.

A matriz U representa o caso de apenas uma unidade de rerroteamento. A quantidade de rerroteamento pode ser representada por um vetor tridimensional $r[a, b, c]$ que indica quantos pacotes foram rerroteados entre os nós a e b via nó intermediário. Dessa forma, a matriz D pode ser definida como:

$$D = \sum_{a,b,c \epsilon [1...N]} r[a, b, c]U, \tag{2.11}$$

e

$$\sum_{i=1}^{N} \sum_{j=1}^{N} D[i, j] \geq 0. \tag{2.12}$$

Por fim, se cada matriz de rerroteamento unitário U significa um acréscimo de uma unidade na carga de tráfego, então o acréscimo total na carga de tráfego será a soma de todos os rerroteamentos efetuados na rede.

Capítulo 3

Referencial Teórico

Sistemas e ferramentas de prevenção de análise de tráfego podem ser utilizados para que um conjunto de usuários possa se comunicar anonimamente em relação a terceiros. É o caso, por exemplo, de unidades militares, que por motivos de segurança buscam manter suas comunicações em sigilo em relação a qualquer parte não autorizada.

Esse tipo de sistema tem como objetivo impedir que terceiros coletem informações como: quantidade de informação que está sendo trafegada num determinado momento e, em um determinado conjunto de nós, de onde está indo e para onde está indo. Além disso, ferramentas utilizadas para análise de tráfego permitem analisar os pacotes trafegados, os protocolos utilizados e até senhas que possam passar em claro na rede. Assim, ferramentas vem sendo desenvolvidas e aprimoradas com o intuito de permitir que usuários desejosos de privacidade ou anonimato possam navegar pela Internet sem que observadores e adversários consigam informações sobre seu comportamento, localização ou opinião. O navegador TOR, Freenet e a rede I2P são exemplos de ferramentas desenvolvidas com esse propósito, cada dia mais difundidas e apoiadas por diferentes tipos de usuários ao redor do mundo.

3.1 Projeto TOR

Conhecido anteriormente como *The Onion Router*, TOR hoje significa mais do que um roteador cebola ou um navegador que esconde seu verdadeiro IP, é um projeto que defende o direito de um usuário na Internet ter privacidade, usufruir de sua liberdade de expressão, permitir o acesso a *sites* e conteúdos que são bloqueados em alguns países por motivos político-ideológicos. Em resumo, um projeto contra a censura e vigilância na Internet.

TOR é uma rede de comunicação anônima de baixa latência baseada em circuito, um modelo de distribuição e confiança [Conrad and Shirazi 2014] que foi construída em cima do projeto de roteamento cebola (*onion routing*. A rede TOR é formada por usuários que executam o software *Onion Proxy* (OP), o qual funciona como um gerenciador dos processos relacionados ao TOR. Para que cada usuário possa se comunicar pela rede, ele constroi um

circuito, selecionando um conjunto ordenado de roteadores cebola (*Onion Router*), ou OR, dentre todos os OR existentes na rede, em um processo conhecido como Seleção de Nós (*Node Selection*) [Conrad and Shirazi 2014]. A lista dos OR disponíveis é obtida de um conjunto de servidores diretório.

Servidores Diretório são servidores autorizados e conhecidos ou publicados em websites específicos que possuem uma lista de OR habilitados e disponíveis. O diretório contém informações sobre os ORs, como descrição do roteador, e um documento de *status* da rede, com medidas de largura de banda dos ORs. Para que um OR seja listado no Servidor Diretório, ele deve ser verificados por sua chave de identidade, ou então serão ignorados. Essa é uma maneira de prevenção contra ataques. Todos os servidores diretórios unem as informações que possuem sobre a topologia da rede e publicam em um diretório comum assinado de toda a rede. Cada OP possui uma lista padrão de servidores diretório, o que permite que consigam criar o circuito a partir dos ORs registrados nos diretórios [Conrad and Shirazi 2014]. Esses servidores diretório são constantemente atualizados pelos usuários OP e baixados via protocolo HTTP [Dingledine et al. 2004].

Roteadores cebola (OR) são o núcelo da rede, pois eles são indispensáveis para a construção dos circuitos. Políticas de saída descrevem quais servidores e portas o OR está disposto a conectar, o que é essencial para a seleção de nós. Todos os ORs conhecidos são categorizados em três níveis [Conrad and Shirazi 2014]:

1. Roteador de guarda de entrada: Estável, rápido e bem conhecido;
2. Roteador intermediário: todos os ORs conhecidos;
3. Roteador de saída: OR com políticas de saída coincidentes.

A seleção de ORs para garantir uma boa performance e prevenir a escolha de ORs corrompidos é feita por meio de um algoritmo de seleção de percurso. O OP busca as informações sobre a rede e os ORs nos servidores diretórios. As medidas disponíveis de largura de banda dos ORs nos diretórios serão utilizadas para a seleção, e aqueles que possuírem maior largura de banda terão maior probabilidade de serem selecionados como roteadores intermediários ou de saída. Já o roteador de guarda de entrada é escolhido aleatoriamente a partir de uma lista mantida no OP com três possíveis candidatos, selecionados de um conjunto de ORs com tempo de vida longo e conhecidamente rápidos e estáveis. Esse roteador escolhido é mantido como roteador de guarda de entrada por 30 dias, e então a lista é refeita e outro roteador será selecionado [Conrad and Shirazi 2014].

Após a seleção dos OR, o OP constroi um circuito até o primeiro OR, chamado de guarda de entrada (*entry guard*). Então esse circuito é utilizado para estender o circuito até o próximo OR, e assim iterativamente até que todos os OR selecionados façam parte do circuito. O circuito servirá, por fim, para encaminhar mensagens entre usuários. Utiliza-se criptografia cebola (*onion encryption*) para que as mensagens sejam encaminhadas anonimamente

pelo circuito até chegar ao último OR, conhecido como roteador de saída (*exit router*), de maneira que apenas o roteador de saída seja capaz de acessar e encaminhar a mensagem ao destino final [Conrad and Shirazi 2014].

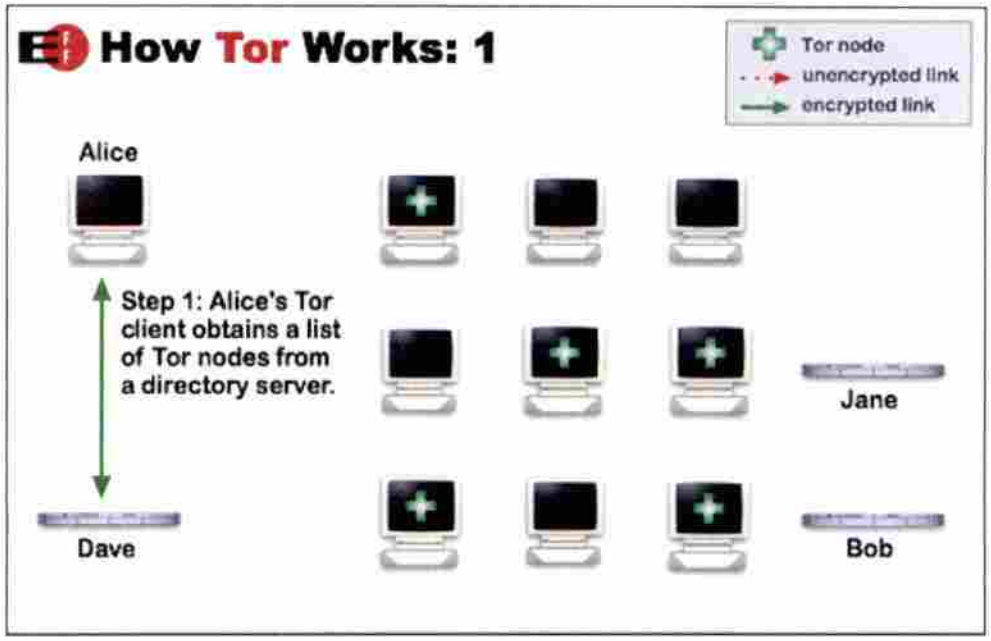


Figura 3.1: Funcionamento da rede TOR (1)

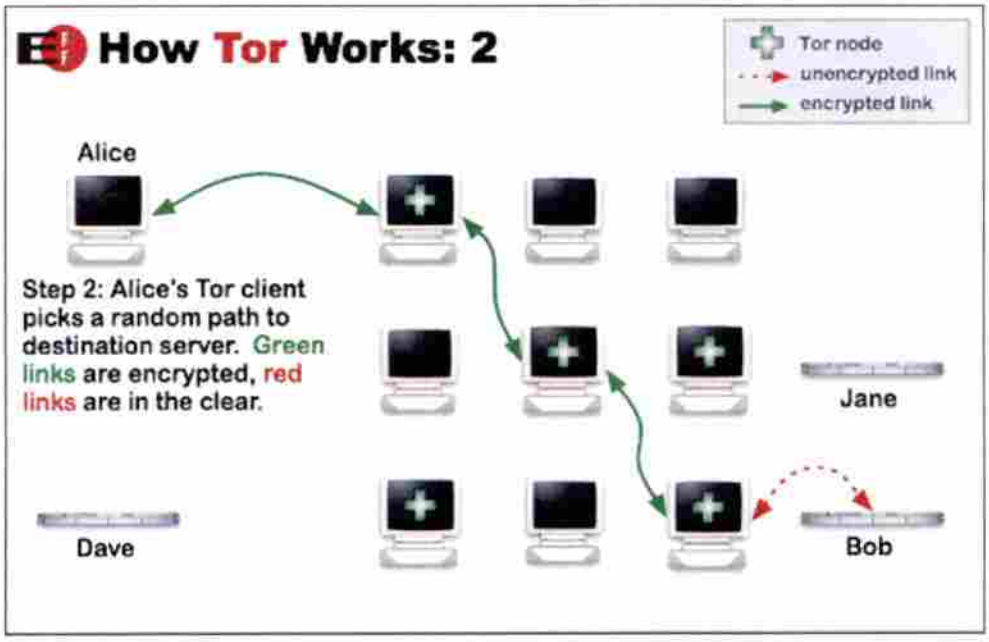


Figura 3.2: Funcionamento da rede TOR (2)

Os circuitos TOR construídos são bidirecionais, utilizados para aplicações baseadas em TCP, com utilização de criptografia em camadas. Cada camada é criptografada com criptografia de chave simétrica e as mensagens são criptografadas com a chave simétrica de cada OR. Com essa implementação, apenas o OR guarda de entrada conhece o endereço IP do nó

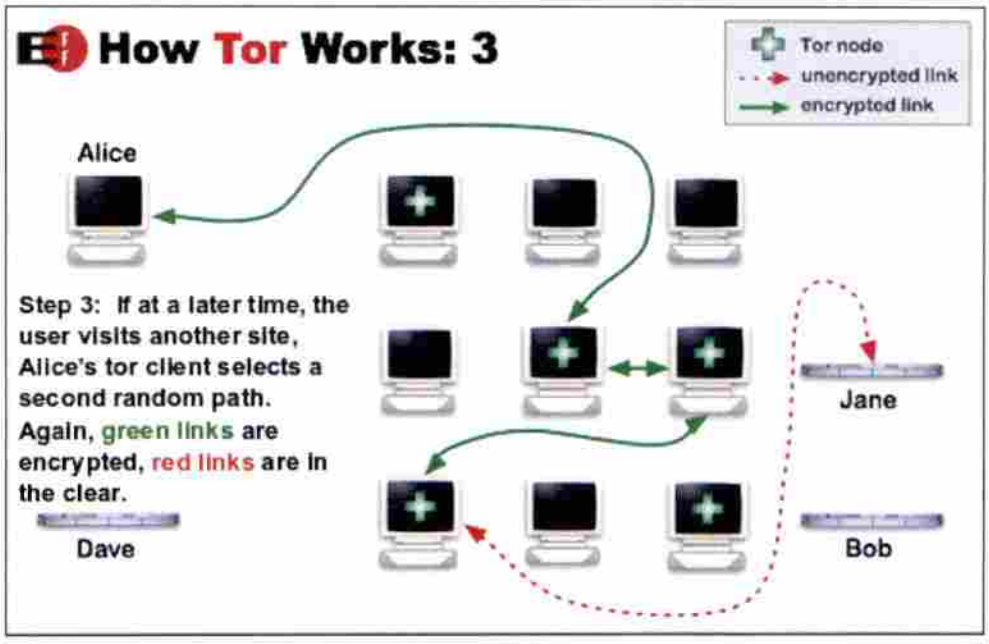


Figura 3.3: Funcionamento da rede TOR (3)

de origem e apenas o OR de saída conhece o endereço IP do destino. Os OR intermediários conhecem apenas seu predecessor e seu sucessor.

O circuito entre ORs se dá por meio de conexões TLS (*Transport Layer Security*), que previne que atacantes se passem por OR da rede ou alterem dados. Cada OR possui dois tipos de chaves: uma de longo prazo e uma de curto prazo. A chave de longo prazo é utilizada para assinatura de certificados TLS, descrição de roteadores (identifica cada OR, apresentando sua chave pública, endereço IP, largura de banda, etc)e diretórios. Já as chaves de curto prazo são utilizadas para a construção de circuitos [Conrad and Shirazi 2014].

Após estabelecer o circuito, o OP pode começar a enviar mensagens pelas células retransmissoras. Cada célula tem seu cabeçalho e carga útil criptografados iterativamente, utilizado a chave simétrica de cada OR participante do circuito, começando pelo roteador de saída, e voltando pelos nós intermediários até chegar no roteador de guarda de entrada [Conrad and Shirazi 2014].

Considere k_1 a chave simétrica trocada entre o OP e o roteador de guarda de entrada, k_2 a chave simétrica trocada entre o OP e o roteador intermediário e k_3 a chave simétrica trocada entre o OP e o roteador de saída. Assuma que $E_k(celula)$ é a função de criptografia utilizando a chave k. Então, a função de criptografia cebola que cifra as mensagens enviadas na rede TOR pode ser representada por [Conrad and Shirazi 2014]:

$$E_{k_1}(E_{k_2}(E_{k_3}(celula))), \tag{3.1}$$

À medida que as células percorrem o circuito as camadas de criptografia são decriptografadas por cada OR no caminho, uma a uma. Apenas o roteador de saída será capaz de extrair o en-

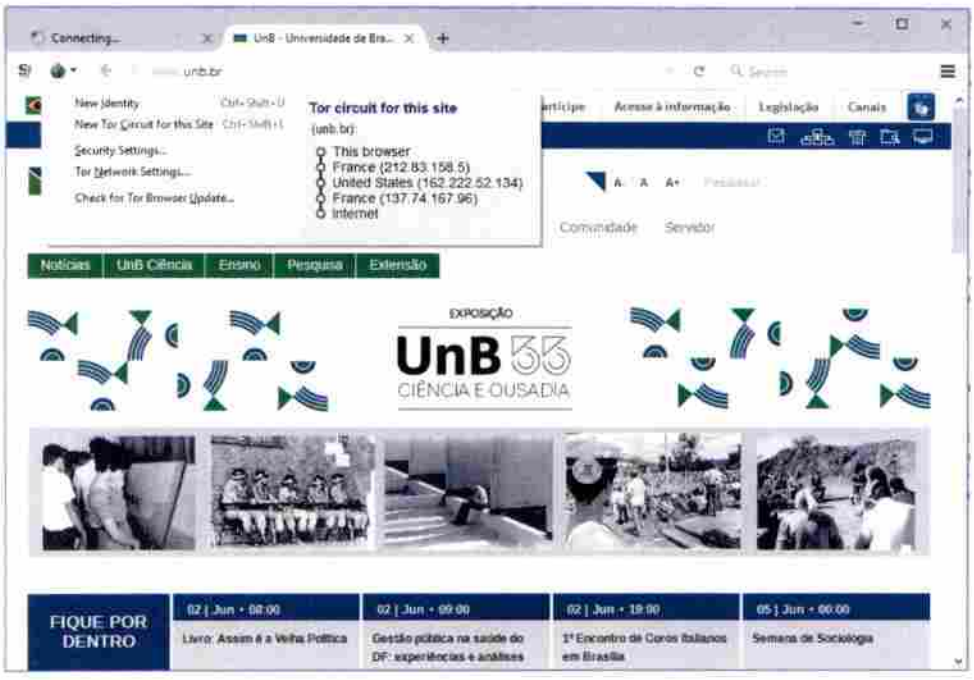


Figura 3.4: Navegador TOR - *surface web*

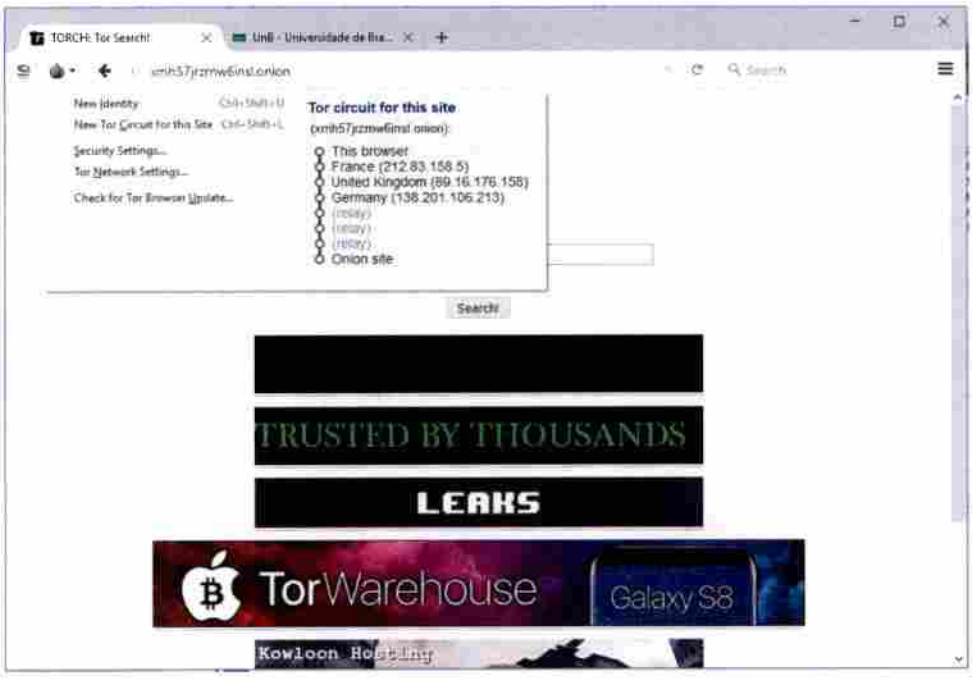


Figura 3.5: Navegador TOR - *deep web*

dereço IP do destino final e a carga útil (mensagem em claro ou uma mensagem criptografada fim-a-fim), e enviá-la para seu destino. A resposta será enviada pelo mesmo circuito, em que cada OR criptografa a célula com sua chave simétrica antes de encaminhá-la de volta ao OR predecessor, assim, apenas o OP é capaz de decifrar a resposta criptografada, pois só ele conhece as chaves simétricas negociadas com cada OR no circuito [Conrad and Shirazi 2014].

Assim, considerando a função de decriptografia D_k utilizada com a chave k, tem-se:

$$D_{k_3}(D_{k_2}(D_{k_1}(celula))). \tag{3.2}$$

A rede TOR pode ser utilizada para navegar em *sites* da *web* sem dar a adversários que fazem análise de tráfego na rede informações de localização, como apresentado nesta seção. O navegador TOR é um dos aplicativos do projeto TOR e é utilizado para navegação na Internet por usuários que buscam mais privacidade. O navegador TOR pode também ser utilizado para acessar sites na *Deep Web* e *Dark Web*, com extensão '.onion', bem como na *Surface Web*. As Figuras 3.4 e 3.5 mostram como é a aparência do navegador TOR quando acessa um site da *surface web* e quando acessa um site da *deep web*, respecitvamente, bem como mostram os circuitos por ele criados para a navegação na Internet de maneira anônima.

3.2 I2P

Também conhecida como *Invisible Internet Project*, a rede **I2P**, que vem sendo desenvolvida desde 2003, se define como sendo uma rede anônima em camadas, comutada por pacotes, que é resiliente, auto organizável e escalável, sobre a qual qualquer número de diferentes aplicações conscientes de segurança ou anonimato podem operar [Project 2017]. Seu objetivo, semelhante ao de várias outras ferramentas de anonimato, é proteger o usuário contra a vigilância e monitoramento por parte de terceiros, fornecendo um meio de comunicação que disponibilize um certo nível de anonimato, tornando mais difícil para os vigilantes, ou adversários, a aquisição de dados e informações que possam ser conseguidas por meio de ataques à rede.

A I2P provê vários tipos de aplicações próprias para variadas finalidades na Internet, como navegador web anônimo, *blogging*, correio eletrônico (susimail, I2P-Bote), servidor web, bate-papo (IRC, Jaber, I2P-Messenger), compartilhamento de arquivos (I2PSnark, iMule, etc), grupo de notícias, distribuição de conteúdo em nuvem (Tahoe-LAFS sobre I2P), entre outras aplicações que proveem anonimato.

Para entender como funciona a I2P, é importante conhecer alguns conceitos sobre a rede. A I2P não relaciona os nós finais, os usuários finais, com os roteadores e aplicações individuais, de maneira que cada roteador utilizado por um nó possui vários destinos locais associados, cada um, a um proxy para uma aplicação. Para que seja possível a comunicação entre os nós na rede, são construídos túneis unidirecionais, sendo eles *outbound*, túnel criado pelo usuário para enviar uma mensagem, ou *inbound*, túnel criado pelo usuário para receber uma mensagem, podendo ser de dois tipos: túneis exploratórios e túneis clientes. O terceiro conceito é o do banco de dados da rede I2P, conhecida como netDb, um par de algoritmos utilizados para o compartilhamento de metadados na rede, os do tipo *RouterInfo*, metadados com informações de roteamento, e *LeaseSets*, metadados com informações do usuário de

destino.

Os túneis são percursos diretos entre uma lista de roteadores selecionados. Um túnel possui um *gateway*, que é o primeiro roteador no túnel, que é o ponto de início, e um ponto final. O *gateway* de um túnel *inbound* recebe as mensagens de outros usuários e as direciona ao seu destino. Para que o usuário destinatário de uma mensagem possa responder ao usuário remetente, ele precisa decifrar as instruções adicionadas à mensagem pelo remetente. Um túnel exploratório possui banda estreita e são utilizados para contactar nós *floodfill*, isto é, nós que constroem e gerenciam o banco de dados da I2P, e adquirir informações da netDb. Já os túneis clientes são de banda larga e utilizados para encaminhar mensagens de aplicações e reter *LeaseSets*. Cada túnel tem duração máxima de 10 (dez) minutos, podendo se desfazer antes em caso de falha de um dos nós ou mesmo de saída do nó (em caso de ficar *offline*). Essa limitação na duração de tempo de um túnel busca prevenir ataques de análise de tráfego.

Com o intuito de que os nós permaneçam anônimos na rede ao se comunicarem por meio desses túneis, implementa-se criptografia em camadas (*garlic encryption*) e o roteamento alho (*garlic routing*). Como é uma rede ponto-a-ponto, os nós escolhidos como roteadores enviam seus metadados com informações de roteamento, *RouterInfo*, diretamente para a netDb, mas os metadados com informações dos demais usuários, os *LeaseSets*, são enviados através dos túneis *outbound*, anonimamente, a fim de evitar correlação entre o roteador e os metadados dos usuários a ele relacionados. Cada túnel possui uma identificação única, *tunnelID*, e o elemento mais externo da cadeia de nós do túnel que servem para sua identificação [Egger et al. 2013].

A maneira como os nós são selecionados para serem parte de um túnel e quais desses nós serão roteadores é de extrema relevância para a criação desse túnel, tanto em relação a desempenho quanto em relação ao anonimato que a rede poderá oferecer baseado nisso. Para a selecionar os nós que serão os roteadores, a I2P categoriza os perfis dos nós membros da rede a partir de medidas indiretas de seu comportamento. Enquanto os perfis são categorizados, a rede efetua cálculos de performance sobre cada nó a fim de resumir nesse valor calculado sua performance, permitindo que sejam comparados com outros nós e então classificados em quatro níveis: rápido e de alta capacidade, alta capacidade, não falhando e falhando. Os limiares de separação entre níveis são determinados dinamicamente e o nó escolhido para participar de um túnel pode aceitar ou não participar de um túnel.

Como pode-se observar, as Figuras 3.6 e 3.7 ilustram a criação dos túneis *outbound* e *inbound*. O banco de dados da rede, netDb, fornece aos gateways e aos nós finais informações necessárias para a construção dos túneis. Os nós na rede que armazenam e fornecem essas informações são nomeados de nós *floodfill*, normalmente selecionados entre os roteadores de nível rápido e com alta capacidade. Os comandos de consulta utilizados na rede são o *store* e *lookup*. Se uma consulta *store* for feita, o nó *floodfill* distribuirá a informação para todos os outros nós *floodfill*, por meio do algoritmo Kademlia. Quando um *floodfill* recebe uma consulta do tipo *lookup*, por questões de segurança ele não encaminhará a consulta para os outros nós *floodfill*, ele apenas responderá a consulta com a informação que possuir no

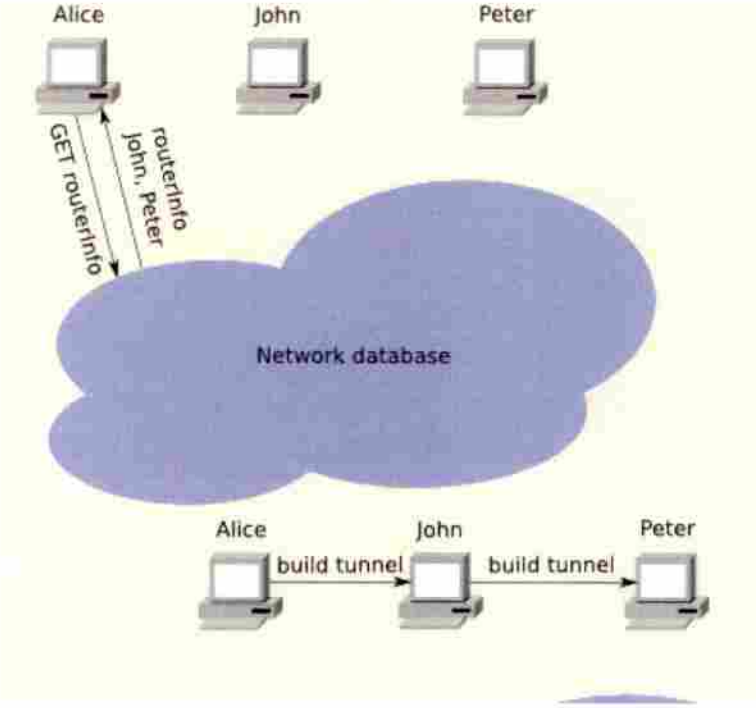


Figura 3.6: Criação de túnel *outbound*

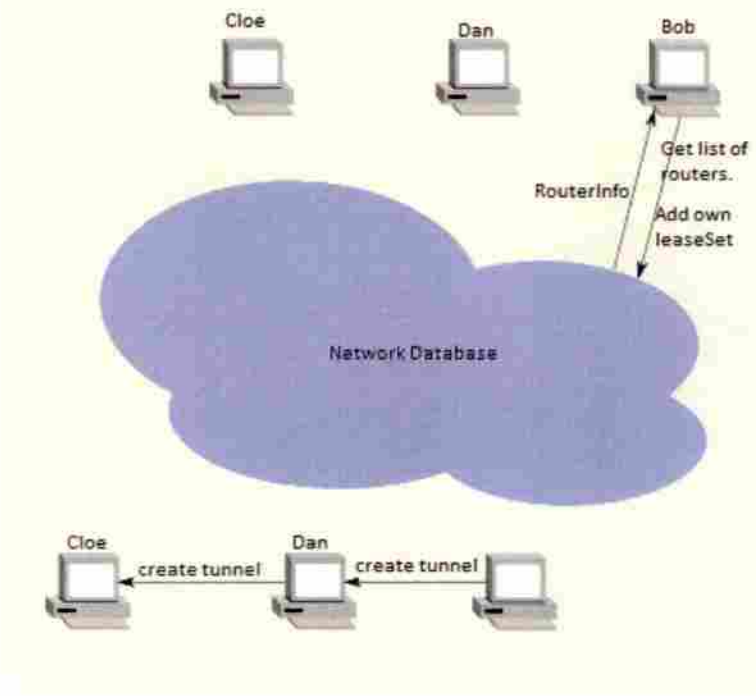


Figura 3.7: Criação de túnel *inbound*

momento.

A fim de manter o banco de dados com informações atualizadas, os dados contêm informação de tempo. Além disso, o dado possui a assinatura da fonte que o publicou e é verificado pelo roteador I2P que o armazena. Por isso há o empacotamento do código ne-

cessário para a manutenção correta de validade do dado, com consultas ocasionais feitas pelos roteadores a servidores SNTP, que permitem a detecção de possíveis distorções entre os roteadores e a camada de transporte. Cada *LeaseSet* possui informações que facilitam essa verificação de dados por parte dos roteadores, como a ID do túnel, horário de expiração, chave pública do destino, assinatura de todos os dados e identidade do gateway do túnel, além de ser armazenado na netDb sob uma chave do destino derivada do SHA256.

Para que a cadeia de comunicação seja atendida e mantido o anonimato, a I2P utiliza protocolos de transporte adaptados para que a confidencialidade e integridade das informações trasmitidas entre um roteador e outro se mantenha confiável enquanto um roteador autentica o outro. O primeiro protocolo de transporte utilizado pela I2P era baseado em TCP, mas com o crescimento da rede a I2P adotou um novo protocolo, este baseado em UDP, o SSU (*Secure Semireliable UDP*), para prover entrega segura, semiconfiável, autenticada e ordenada. O SSU deveria fazer controle de congestionamento, entretanto, após alguns problemas, um novo protocolo baseado em NIO-TCP, o NTCP, que é habilitado apenas para conexões *outbound*, podendo ser utilizado para conexões *inbound* se for feita uma configuração específica em seu NAT/Firewall e no arquivo "/config.jsp". Atualmente, a I2P suporta esses diferentes protocolos simultanemamente, dando alta prioridade para uso do NTCP em conexões *outbound*, enquanto o SSU é habilitado para conexões *outbound* e *inbound*.

Como mencionado anteriormente, a rede I2P utiliza a criptografia em camadas, que pode ser também chamada de criptografia alho, muito similar ao que é aplicado no roteamento cebola (*onion routing*) utilizado no TOR, no que se refere à criptografar as mensagens várias vezes, ou seja, várias camadas de criptografia feitas na fonte de transmissão da mensagem e decifradas, camada por camada, por cada nó pelo qual passa até chegar, com sua última camada criptografada, ao seu destino final, único com a chave para decifrar a mensagem. Mas o que diferencia o roteamento alho do roteamento cebola, de acordo com a equipe de colaboradores da I2P, é que no roteamento alho há o empacotamento de múltiplas mensagens juntas. Os túneis construídos são utilizados com criptografia em camadas, com **ElGamal/AES+SessionTag**. Uma camada acima dos túneis, a I2P envia mensagens fim-a-fim entre os usuários, com a criptografia alho.

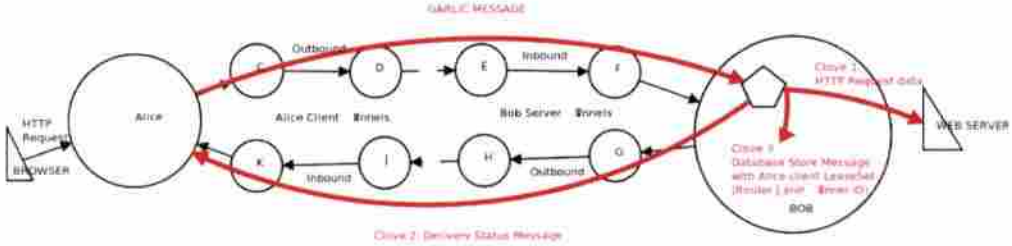


Figura 3.8: Mensagem Garlic

Cada mensagem alho, isto é, cada mensagem criptografada com **ElGamal/AES+SessionTag**, torna-se dentes de alho (*garlic cloves*. Normalmente cada mensagem possui apenas um dente de alho, enviando de tempos em tempos uma mensagem

com mais dois dentes adicionais, um que funciona como um ACK (*Delivery Status Message*) para apresentar ao usuário originário da mensagem uma resposta com a situação da entrega da mensagem, e outro que possui instruções para se contactar o usuário de origem (*Database Store Message*), com dados para comunicação com o usuário originário da mensagem, os *LeaseSet*. Para se manter o padrão de anonimato, os dentes com mensagens de ACK são embrulhados em outras mensagens alho pelo usuário origem, mantendo-os criptografados para os nós que estiverem no caminho do túnel de volta. A Figura 3.8 ilustra a transmissão de uma mensagem alho ou mensagem Garlic.

A I2P utiliza combinações de criptografia simétrica, criptografia assimétrica, assinaturas e *hashes*. Como apresentado anteriormente, a criptografia utilizada na I2P é feita em camadas. A primeira camada a utilizar criptografia é a camada de transporte, utilizando o TLS (*transport layer security*). Além da camada de transporte criptografada, os túneis também são criptografados. Além disso, as mensagens são criptografadas como mensagens Garlic. Contanto que a interação entre os usuários seja dentro da rede I2P, as mensagens na rede são criptografadas fim-a-fim [Conrad and Shirazi 2014].

Os túneis I2P utilizam criptografia simétrica AES256 com CBC, a fim de que cada roteador no túnel veja apenas as instruções de entrega (*Delivery Instructions* o *gateway outbound* primeiramente estabelece uma chave única de sessão com cada roteador do túnel utilizando Diffie-Hellmann, criptografando a mensagem para cada salto, e cada roteador decriptografa a mensagem utilizando a chave única de sessão que foi trocada entre ele e o *gateway outbound*. Já a criptografia assimétrica é utilizada para criptografar as mensagens, com ElGamal, ou seja, cada mensagem é criptografada com a chave pública do destinatário, inclusive mensagens de gerenciamento trocadas entre os roteadores [Müller 2016]. Por fim, cada mensagem é criptografada com uma combinação de Garlic e chave pública 20148 bit ElGamal para que não passe em claro entre o roteador final do túnel *outbound* e o *gateway* do túnel *inbound*. As etapas da criptografia em camadas podem ser vistas na Figura 3.9.

3.3 Conclusão

Como apresentado nas seções anteriores, são variadas as técnicas e ferramentas de anonimato existentes. Cada uma voltada para um propósito, mas muito similares em relação ao objetivo principal, que é manter o anonimato do usuário na rede Internet. Cada ferramenta tem destaque em um tipo de uso, como navegação e compartilhamento de sites de maneira anônima, troca de mensagens eletrônicas ou o armazenamento distribuído de conteúdo de maneira anônima. Todas oferecem benefícios para quem busca uma forma de comunicação que possa oferecer o máximo de privacidade e anonimato possível, mas com custos embutidos, como o custo computacional para a implementação e utilização das técnicas e ferramentas, que podem influenciar na performance do sistema, bem como o custo social de se tornar possível que criminosos utilizem essas ferramentas para cometerem crimes de

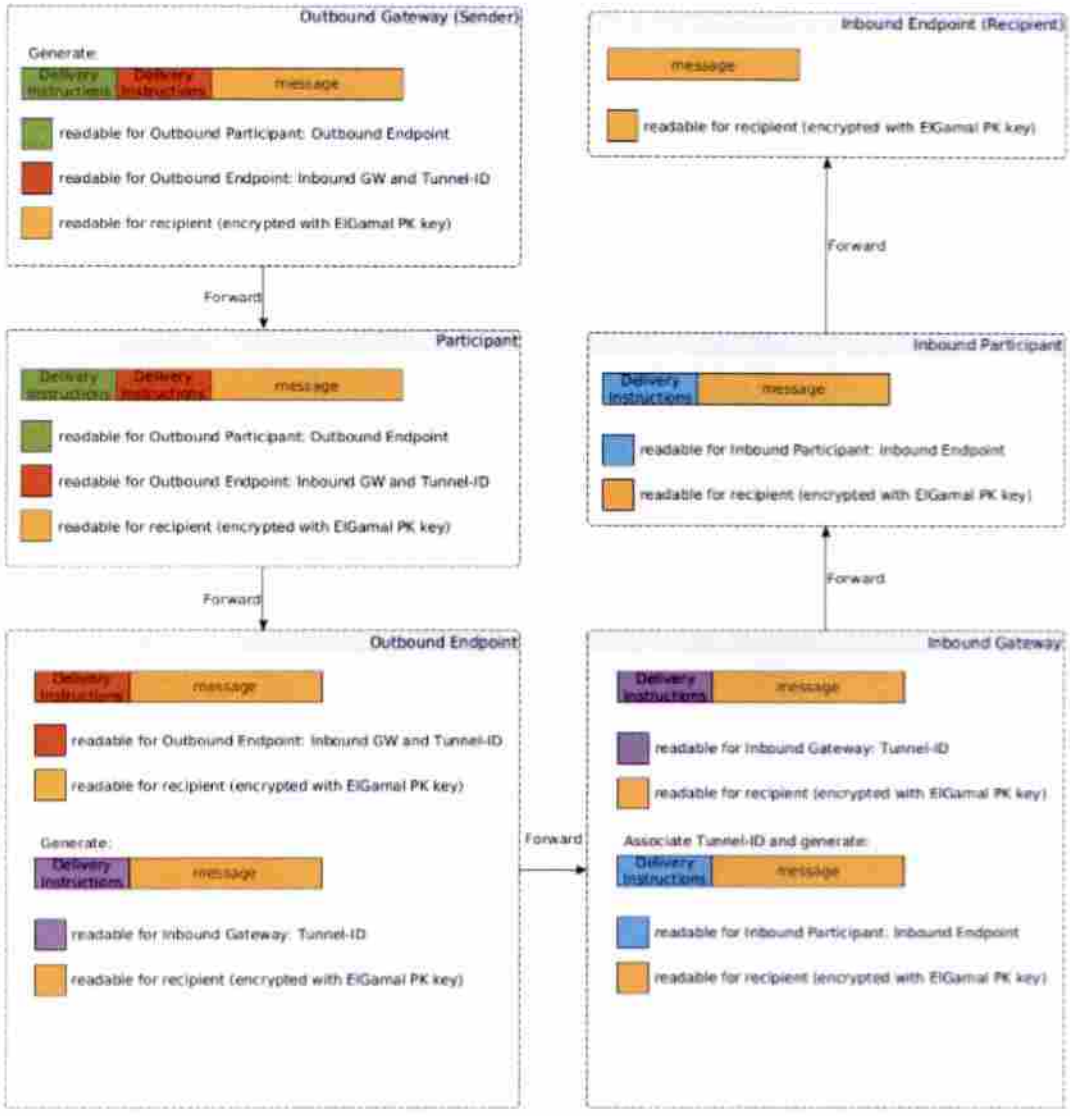


Figura 3.9: Criptografia em Camadas na rede I2P

maneira anônima, tornando mais difícil que sejam descobertos.

Capítulo 4

Análise de Redes de Anonimato

Este Capítulo apresenta uma análise quantitativa dos métodos de *padding* e de rerroteamento, técnicas de anonimato apresentadas no Capítulo 2 e uma análise comparativa entre as redes de anonimato TOR e I2P por meio de classificação quantitativa e qualitativa do nível de anonimato que essas redes podem oferecer.

4.1 Análise Comparativa TOR x I2P

Do Capítulo 3 é possível notar que tanto as redes TOR e I2P são redes de anonimato baseadas em proxy, que permitem que usuários enviem e recebam informações por meio de túneis de anonimato. Ambas podem ser vulneráveis a ataques de adversários passivos globais, que conseguem analisar o tráfego da comunicação a partir do proxy de saída, que são os nós que têm acesso ao texto claro das mensagens que entram e saem dos túneis. Uma das principais diferenças entre essas duas redes é que a TOR possui gerenciamento centralizado de sua base de dados, enquanto a I2P possui base de dados com gerenciamento distribuído. Em relação a linguagem de computação implementada, a rede TOR adota a linguagem C, e a I2P o Java.

A rede TOR é maior que a I2P, possui mais usuários e mais desenvolvedores como colaboradores. Além disso, tem maior destaque no meio acadêmico e mais financiadores. Tecnicamente, resolveu alguns problemas relacionados a escalonamento que a I2P ainda está trabalhando para resolver, mas devido a quantidade crescente de usuários, vem se adaptando em relação a tentativas de bloqueio ou DOS (*Denial of Service*), negação de serviço. Devido ao TLS e às pontes, o TOR se mostra mais resistente a bloqueios de nível de estado do que a I2P. TOR possui um núcleo centralizado com alta capacidade, propiciando maior vazão e menor latência que a I2P, e uso de memória mais eficiente. O núcleo centralizado da TOR também reduz a complexidade em cada nó, o que possibilita melhor reação a ataques Sybil ??. Para muitos, o fato de a rede TOR ser implementada em linguagem C também pode ser considerado uma vantagem em relação a I2P, que utiliza Java, mas cabe ao leitor fazer juízo

de valor em relação a este aspecto.

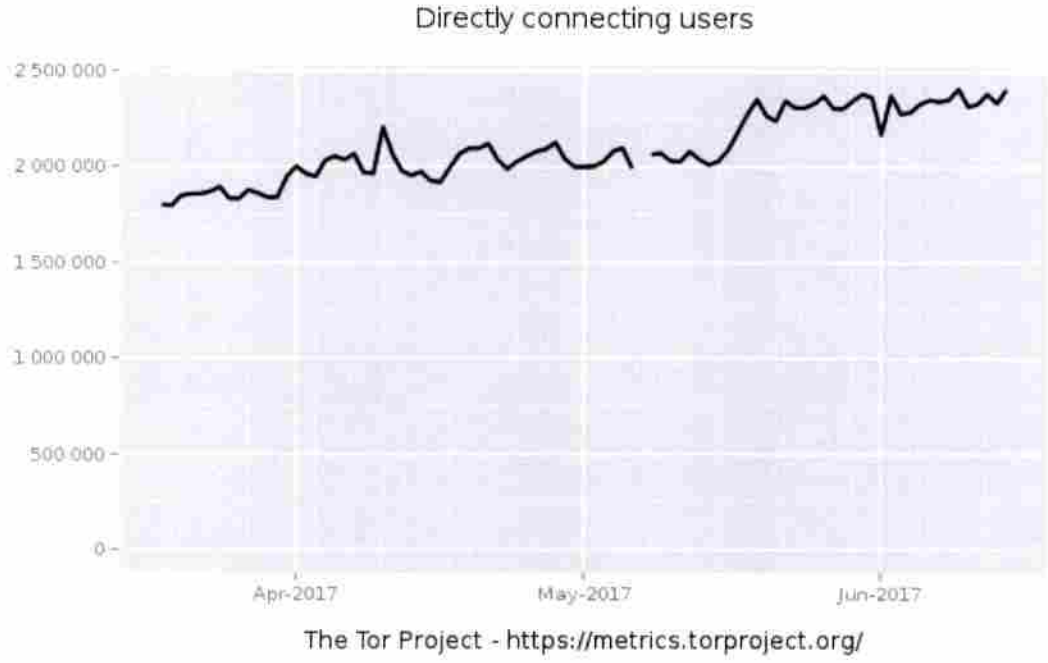


Figura 4.1: Usuários Conectados na Rede TOR

A rede I2P é descentralizada, distribuída e auto-organizada, pensada para serviços anônimos que são mais rápidos na I2P que na rede TOR. A seleção de quais nós serão roteadores é feita por meio de classificação de perfil, o que não impede que qualquer nó na rede possa se tornar um roteador ou um nó da netDb. Diferentemente da TOR, a I2P é comutada por pacotes, e não por circuitos, sendo uma rede amigável para aplicações ponto-a-ponto, o que permite maior equilíbrio na distribuição de carga na rede, entre os nós, sem se prender a um circuito único. Este último aspecto pode ser muito benéfico em se tratando de proteção contra análise de tráfego, já que o tráfego da carga fica distribuído entre múltiplos nós e cada usuário possui pontos de entrada de rede (túneis inbound) comuns para remetentes ou destinos distintos, enquanto na TOR existe um circuito para cada comunicação entre um usuário e outro. A despeito dos circuitos na rede TOR, que duram por um bom tempo, os túneis da I2P tem vida de curta duração, o que pode oferecer maior dificuldade a um adversário ativo já que fornece menos amostras das características de seu tráfego. Também, diferente da rede TOR, a rede I2P trabalha com protocolos TCP e UDP na camada de transporte **??**.

Os dados mais recentes encontrados apresentam uma estimativa de que a rede TOR possuía, em Junho de 2017, aproximadamente dois milhões de usuários e aproximadamente sete mil roteadores, como mostram os gráficos 4.1 e 4.2. Já as informações de quantitativo de usuários da rede I2P são datados do ano de 2013, e estimavam uma quantidade aproximada de vinte mil usuários na rede [Egger et al.]. As duas redes possuem vantagens e desvantagens em relação a aspectos que devem ser levados em consideração pelo usuário, para que possa escolher aquela que, para o seu perfil e necessidade, melhor se adeque.

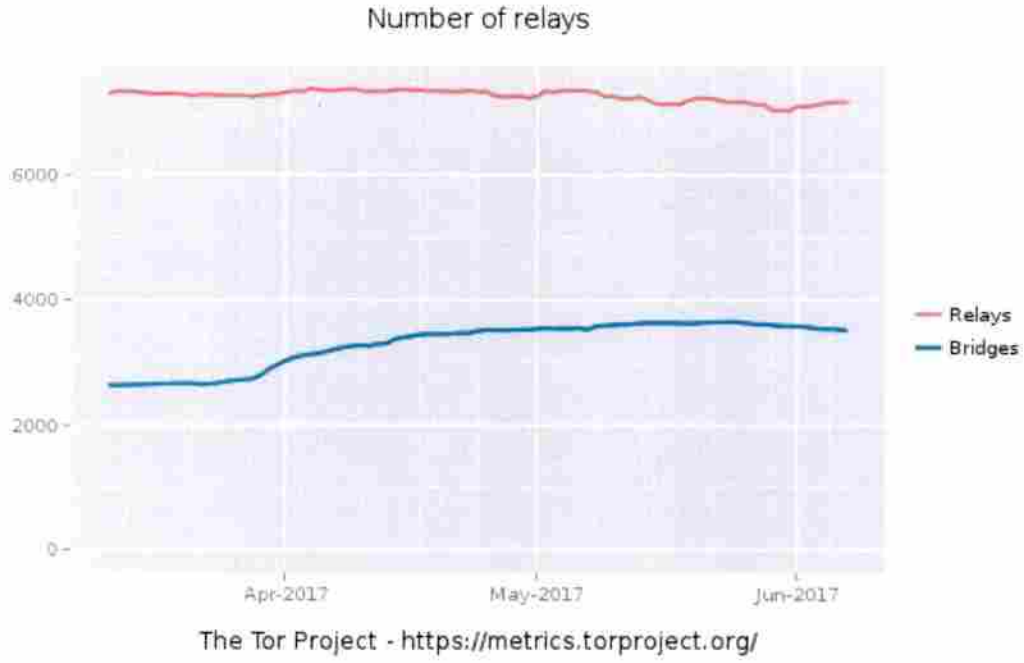


Figura 4.2: Roteadores na Rede TOR

4.2 Análise Quantitativa

4.2.1 Padding

Considere que foi adotada a técnica de Padding no sistema apresentado na seção 2.3.1 pela Equação 2.8 e que permite-se que apenas uma mensagem seja adicionada por enlace para cada *pad*. Isso poderia resultar em 64 possíveis matrizes de tráfego reais, visto que ao todo são seis elementos na matriz que poderiam sofrer alteração em seu valor por subtração de uma unidade, ou seja, cada elemento não pertencente à diagonal da matriz poderia ser representado por dois valores distintos, tal que isso daria 2^6 combinações possíveis. Como a matriz observada também pode ser considerada uma possível matriz de tráfego real, então teríamos mais 63 matrizes de tráfego possíveis, por exemplo:

$$MT_1 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 3 \\ 1 & 1 & 0 \end{bmatrix}, MT_2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 3 \\ 1 & 1 & 0 \end{bmatrix}, MT_3 = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 2 \\ 1 & 1 & 0 \end{bmatrix}, \ldots \tag{4.1}$$

De maneira simples, se considerarmos o caso mais simples, tal que a matriz observada passou por apenas um *pad*, teríamos, para o contexto representado acima, uma probabilidade de encontrar a matriz de tráfego real igual a $1/64$, ou seja, aproximadamente $1,56\%$ de que o tráfego real seja descoberto pelo adversário.

Em um cenário mais realístico, com mais nós envolvidos, a probabilidade de um adversá-

rio encontrar a matriz de tráfego real diminui ainda mais considerando as mesmas condições acima apresentadas, visto que para cada nó adicionado à rede, a quantidade de possíveis matrizes é multiplicada por uma potência de 2, pois a quantidade de elementos que podem sofrer permutações aumentará para $N^2 - N$, em que N representa a quantidade de nós na rede. Assim, a quantidade de matrizes possíveis será igual a 2^{N^2-N} considerando apenas um *pad* por período de observação. Se forem consideradas duas rodadas de *padding*, então seriam três possíveis valores para quantidades de mensagens enviadas em cada enlace entre os nós e, portanto, a quantidade de matrizes possíveis poderia ser encontrada pela fórmula 3^{N^2-N} ou, de maneira mais geral, para cada x rodadas de *padding* a quantidade de matrizes possíveis será $(x+1)^{N^2-N}$.

Não é difícil notar que quanto mais rodadas de *padding* são realizadas pelo sistema, maior o custo computacional e financeiro, já que envolve diretamente a capacidade de transmissão e recepção de cada nó, bem como a capacidade de tráfego dos enlaces na rede. A adição de mensagens tem como consequência a utilização da capacidade da rede, um maior processamento por parte dos equipamentos que transmitem e recebem as mensagens e maior utilização de memória, além de outros fatores que agregam custo ao método. Por isso, quando se busca um maior graud e anonimato, também pode ser possível a combinação de diferentes métodos de prevenção de análise de tráfego, o que pode, em alguns casos, equilibrar o custo com o benefício, como é o caso da combinação do método de *padding* com o rerroteamento, que será apresentado na próxima seção [Newman et al. 2003].

4.2.2 Rerroteamento

Como exemplo voltemos à matriz de tráfego observada dada pela Equação 2.8. Considere que apenas uma unidade de rerroteamento foi efetuada. Então, as possíveis matrizes de tráfego real resultantes da matriz observada $MT_{observada}$ poderiam ser:

$$MT_1 = \begin{bmatrix} 0 & 2 & 1 \\ 1 & 0 & 3 \\ 2 & 0 & 0 \end{bmatrix}, MT_2 = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 1 & 1 & 0 \end{bmatrix}, MT_3 = \begin{bmatrix} 0 & 2 & 1 \\ 2 & 0 & 2 \\ 1 & 1 & 0 \end{bmatrix}, \tag{4.2}$$

$$MT_4 = \begin{bmatrix} 0 & 2 & 1 \\ 1 & 0 & 3 \\ 0 & 2 & 0 \end{bmatrix}, MT_5 = \begin{bmatrix} 0 & 2 & 1 \\ 0 & 0 & 4 \\ 1 & 1 & 0 \end{bmatrix}, MT_6 = \begin{bmatrix} 0 & 3 & 0 \\ 1 & 0 & 3 \\ 1 & 1 & 0 \end{bmatrix}. \tag{4.3}$$

Agora, se considerarmos que foram realizadas duas unidades de rerroteamento, então teremos as seguintes possíveis matrizes de tráfego real a partir da matriz observada 2.8:

$$MT_7 = \begin{bmatrix} 0 & 3 & 0 \\ 1 & 0 & 3 \\ 1 & 1 & 0 \end{bmatrix}, MT_8 = \begin{bmatrix} 0 & 0 & 3 \\ 1 & 0 & 3 \\ 1 & 1 & 0 \end{bmatrix}, MT_9 = \begin{bmatrix} 0 & 2 & 1 \\ 3 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}, MT_{10} = \begin{bmatrix} 0 & 2 & 1 \\ 0 & 0 & 4 \\ 2 & 0 & 0 \end{bmatrix} \tag{4.4}$$

$$MT_{11} = \begin{bmatrix} 0 & 2 & 1 \\ 0 & 0 & 4 \\ 0 & 2 & 0 \end{bmatrix}, MT_{12} = \begin{bmatrix} 0 & 1 & 2 \\ 0 & 0 & 4 \\ 1 & 1 & 0 \end{bmatrix}, MT_{13} = \begin{bmatrix} 0 & 3 & 0 \\ 0 & 0 & 4 \\ 1 & 1 & 0 \end{bmatrix}, MT_{14} = \begin{bmatrix} 0 & 2 & 1 \\ 2 & 0 & 2 \\ 2 & 0 & 0 \end{bmatrix} \tag{4.5}$$

$$MT_{15} = \begin{bmatrix} 0 & 2 & 1 \\ 2 & 0 & 2 \\ 0 & 2 & 0 \end{bmatrix}, MT_{16} = \begin{bmatrix} 0 & 0 & 3 \\ 2 & 0 & 2 \\ 1 & 1 & 0 \end{bmatrix}, MT_{17} = \begin{bmatrix} 0 & 3 & 0 \\ 2 & 0 & 2 \\ 1 & 1 & 0 \end{bmatrix}. \tag{4.6}$$

A partir da matriz de tráfego observada, assumindo-se terem sido realizadas três unidades de rerroteamento, o resultado seriam mais algumas possíveis matrizes de tráfego real, o que diminuiria a probabilidade de o adversário obter a matriz de tráfego real e, consequentemente, aumentaria o grau de anonimato. Outro fator que é de extrema relevância para a análise do grau de anonimato é a quantidade nós na rede, quanto mais nós, mais combinações de matrizes de tráfego real são possíveis.

A probabilidade de o adversário descobrir a matriz de tráfego real MT_{real} é inversamente proporcional à quantidade de possíveis matrizes de tráfego real resultantes de transformações na matriz de tráfego observada, ou seja, quanto maior a quantidade de possíveis matrizes a partir de uma observada, menor a probabilidade de o adversário a descobrir e, portanto, maior o grau de anonimato que o sistema provê.

4.2.3 TOR x I2P

Como apresentado no Capítulo 2, é possível se obter uma medida quantitativa do nível de anonimato de um sistema a partir do cálculo de entropia que esse sistema possui. Quanto maior a entropia, maior o nível de anonimato [Serjantov and Danezis]. Para isso, deve-se levar em consideração variáveis como a quantidade de nós na rede analisada e os modelos de ameaça à rede.

Para uma análise quantitativa do nível de anonimato que as redes TOR e I2P podem oferecer, foram considerados dois cenários. No Cenário 1, assume-se um modelo de ameaça de adversário passivo global, ou seja, um modelo de ameaça em que o adversário não interfere na rede de maneira ativa, apenas monitora e analisa o tráfego. No Cenário 2, considera-se o modelo de ameaça em que o atacante é ativo e, por meio de nós maliciosos que ele controla na rede, pode ter mais informações sobre a rede e os usuários a ela conectados por meio de análise de tráfego.

O Cenário 1 considera que não há adversários ativos, apenas passivos, tendo como variável principal a quantidade de nós na rede. Assim, considera-se que não há nós maliciosos na rede e, portanto, é igualmente provável para um observador que uma mensagem específica tenha sido enviada por qualquer um dos usuários da rede. O nível de anonimato é calcu-

lado pela Equação 2.1. Considere ainda que os adversários conseguem monitorar apenas um conjunto de nós na rede, em que a quantidade de nós do conjunto é N. Assim, quando N é máximo, o adversário poderá ser considerado passivo global.

O Gráfico 4.3 mostra o valor de entropia dado que há N nós clientes na rede TOR, com valores de referência encontrados nos Gráficos 4.1 e 4.2, e mostra que o nível de anonimato é maior quanto maior for a quantidade de usuários na rede. Os resultados obtidos mostram valores de entropia calculados para a quantidade de usuários N, que começa com o valor de $N = 1000$ usuários e aumenta com razão igual a 1000 até chegar no valor máximo de usuários da rede TOR encontrados, $N = 2$ milhões de usuários.

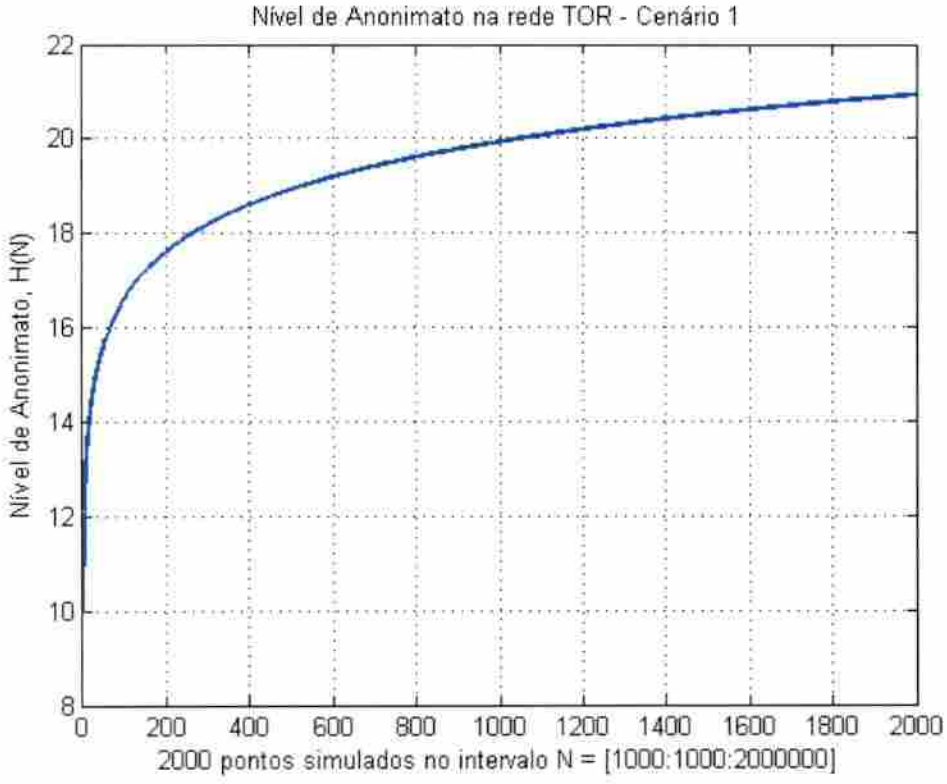


Figura 4.3: Nível de Anonimato na Rede TOR para N clientes - Cenário 1

Observa-se que o nível de anonimato aumenta à medida que a quantidade de nós na rede aumenta, como já mencionado anteriormente [Pfitzmann and Kohntopp 2001]. Ainda, o valor normalizado de entropia $D(S)$ dado pela Equação 2.2 quando todos os usuários são igualmente prováveis remetentes de uma dada mensagem será máximo e igual a 1, como mostra o gráfico 4.4.

Para o mesmo Cenário 1 apresentado anteriormente, o Gráfico 4.5 apresenta os valores de entropia $H(N)$ para a quantidade N de nós na rede I2P (dados do ano de 2013 [Egger et al.]). Deste Gráfico pode-se inferir que, de fato, quanto maior a quantidade de nós, maior o nível de anonimato que a rede pode oferecer quando todos os nós possuírem a mesma probabilidade de terem enviado uma dada mensagem.

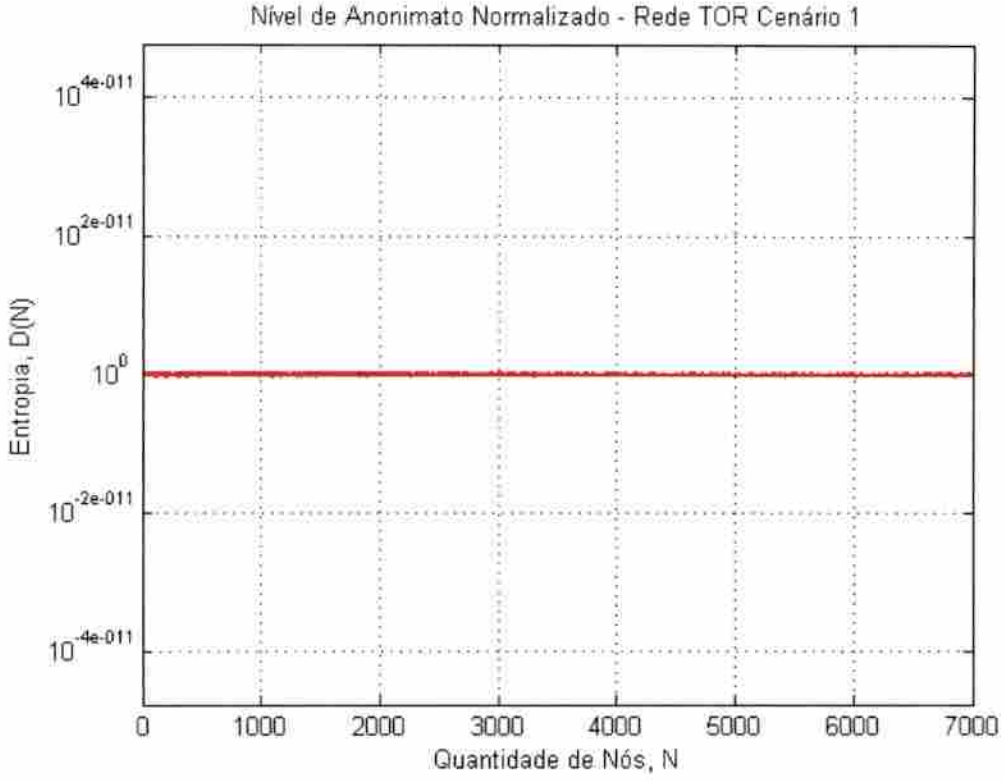


Figura 4.4: Nível de Anonimato Normalizado - Rede TOR Cenário 1

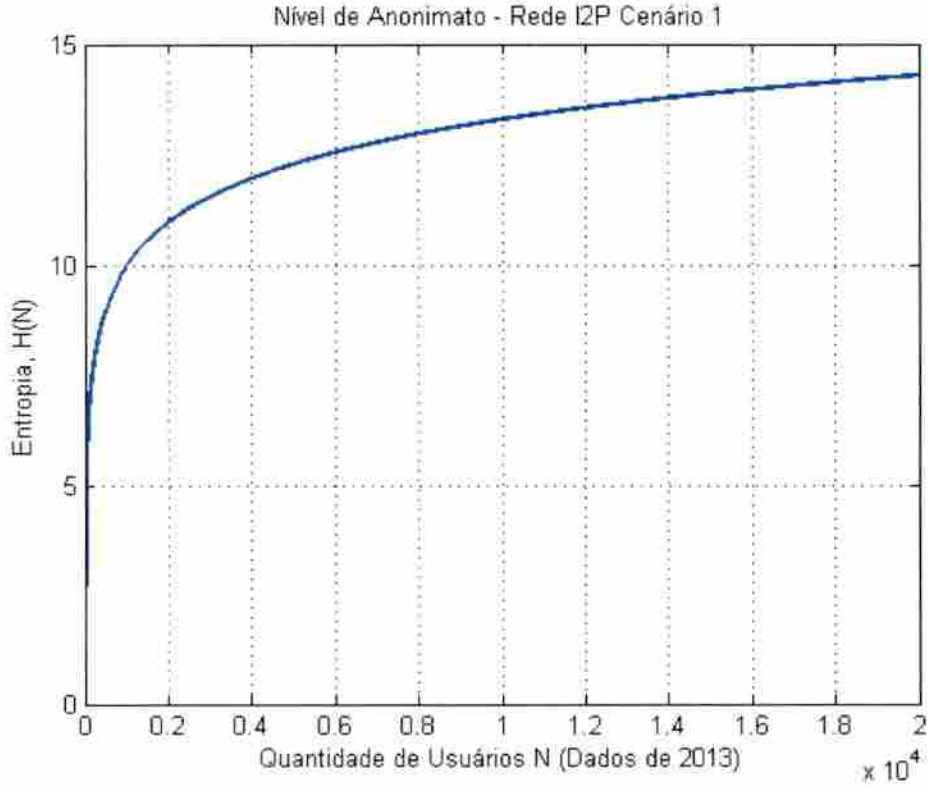


Figura 4.5: Nível de Anonimato na Rede I2P - Cenário 1

Como esperado, sem interferência externa de um adversário ou nós maliciosos na rede, se a única variável no sistema é a quantidade de nós que ele possui, então, considerando a Equação 2.2, o anonimato nas redes TOR e I2P pode ser considerado perfeito nessas condições. O Gráfico 4.6 apresenta o nível de anonimato normalizado $D(N)$ para a rede I2P.

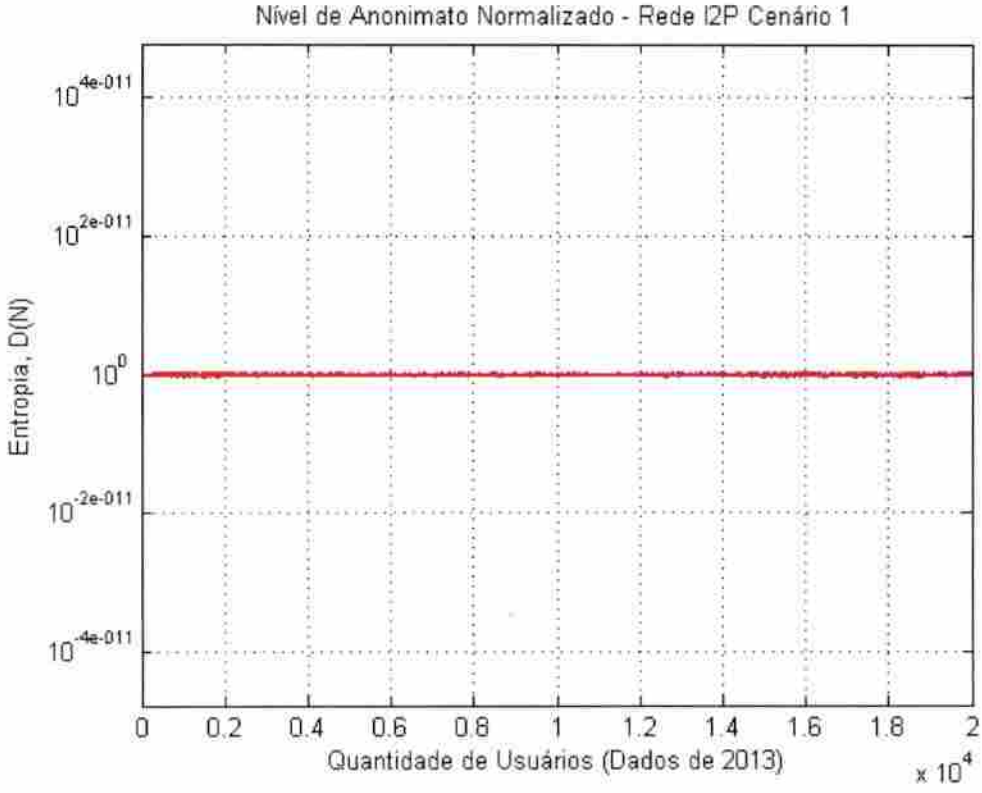


Figura 4.6: Nível de Anonimato Normalizado - Rede I2P Cenário 1

Para o Cenário 1, tanto na rede TOR quanto na rede I2P, quanto maior a quantidade de nós na rede, maior o nível de anonimato que ela oferece. Dados de 2017 estimam que a rede TOR possui aproximadamente 2 milhões de usuários e aproximadamente 7 mil roteadores. Os dados mais recentes encontrados sobre a quantidade de usuários da rede I2P são de 2013, com estimativa de aproximadamente 20 mil usuários. Se forem consideradas apenas as variáveis indicadas no Cenário 1, então, de acordo com a classificação quantitativa de anonimato apresentada no Capítulo 2 dada pela Equação 2.1, a rede com mais usuários pode oferecer, nessas condições, um maior nível de anonimato.

Considere agora o Cenário 2, em que o modelo de ameaça considera adversários ativos. Assuma que o adversário tem controle sobre alguns nós da rede e possui informações sobre um ou mais usuários num conjunto de nós na rede, isto é, o adversário conhece o tráfego de mensagens entre o usuário s_i e o usuário s_j e sabe com qual probabilidade uma dada mensagem na rede foi enviada pelo usuário s_i num dado momento. Suponha que o usuário s_i possui a maior probabilidade p_i de ser o remetente de uma mensagem, e os demais $(N - 1)$ usuários sejam igualmente prováveis remetentes, o que resulta em uma distribuição de probabilidade assimétrica.

O nível de anonimato calculado para o Cenário 2 utiliza a Equação 2.1 e considera que o usuário i possui a maior probabilidade p_i de ser o remetente, tal que:

$$p_i = \frac{a}{N}, \tag{4.7}$$

em que $1 \leq a \leq N$, e a probabilidade dos demais usuários é dada por q:

$$q_j = \frac{1 - p_i}{N - 1}. \tag{4.8}$$

O Gráfico 4.7 apresenta o resultado calculado para o nível de anonimato na rede TOR considerando como variáveis a quantidade de nós na rede e a probabilidade de um destes ser o remetente da mensagem. Para a simulação, foram considerados dois (2) milhões de nós e distribuição de probabilidade assimétrica, em que um dos nós possui probabilidade de ser o remetente igual a $0,1$, para N nós na rede, e os demais usuários possuem a mesma probabilidade $(1 - 0,1)/N$ de serem os remetentes da mensagem. A simulação mostra valores de entropia dados N usuários na rede, em que N começa com valor igual a 1000 e cresce com razão 1000 até alcançar o total de 2 milhões de usuários. Os resultados são semelhantes aos encontrados para outras quantidades de nós, um maior nível de anonimato quanto maior for a quantidade de usuários na rede, mesmo que o adversário seja ativo e saiba qual a probabilidade de que um determinado usuário seja o remetente de uma mensagem trafegada na rede.

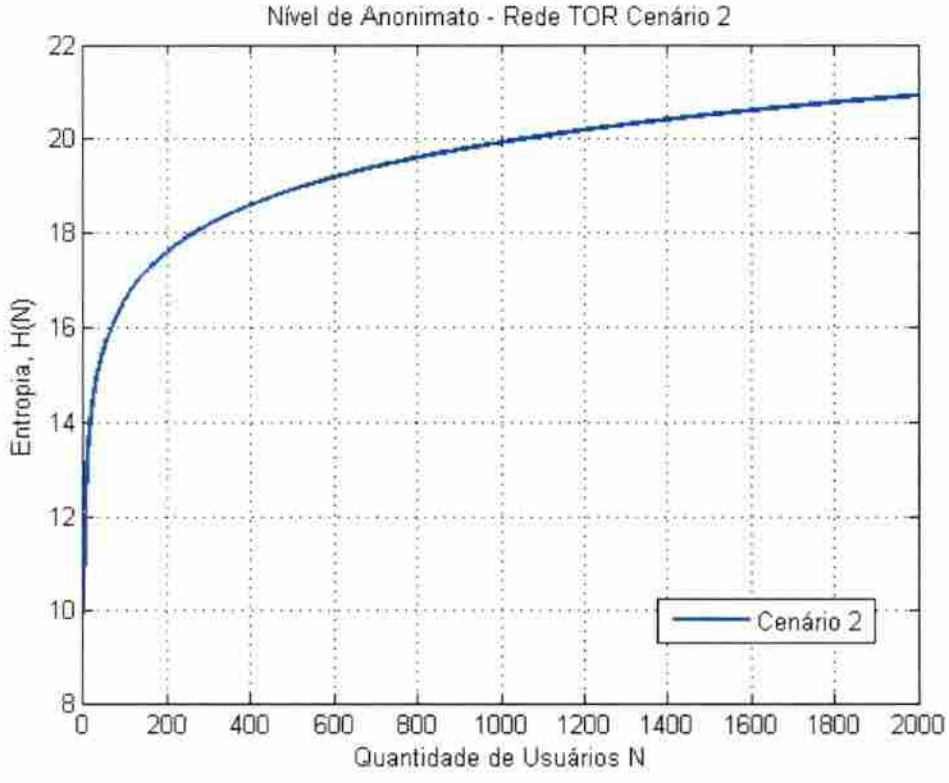


Figura 4.7: Nível de Anonimato rede TOR para N usuários - Cenário 2

Na Figura 4.8 pode-se observar o mesmo comportamento obtido nos gráficos anteriores, em que a entropia aumenta à medida que a quantidade de usuários na rede aumenta. Ainda, para calcular o nível de anonimato da rede I2P no Cenário 2, foi considerada a mesma distribuição de probabilidade assimétrica utilizada para a simulação do nível de anonimato na rede TOR, e foram considerados até vinte (20) mil usuários na rede I2P.

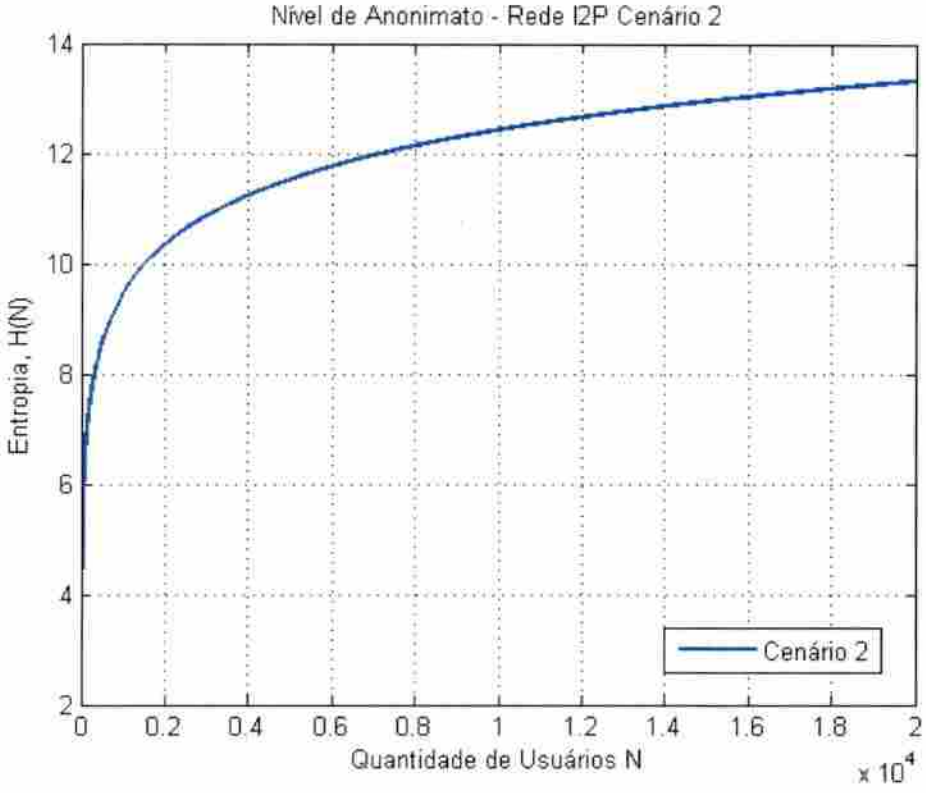


Figura 4.8: Nível de Anonimato rede I2P para N usuários - Cenário 2

Percebe-se, pela análise dos gráficos apresentados, que quanto maior o número de usuários numa rede, maior o nível de anonimato que essa rede poderá oferecer. Porém, para modelos de ameaça distintos, percebe-se que, para um mesmo número de usuários, o Cenário 1 devolve um maior valor de entropia e, portanto, um maior nível de anonimato. Tem-se a comparação entre o Cenário 1 e o Cenário 2 no Gráfico 4.9, considerando setenta (70) usuários, $N = 70$.

Do Gráfico comparativo 4.9 pode-se inferir que, apesar de diferentes modelos de ameaça, para uma mesma quantidade de usuário o nível de anonimato calculado pela entropia de Shannon é muito próximo. No Cenário 2 o adversário é capaz de investigar um usuário, e, apesar de a entropia no Cenário 1 ser maior, o Cenário

Ainda no Cenário 2, considere agora uma distribuição de probabilidade em que o adversário é capaz de investigar não um usuário, mas seis deles, tal que a probabilidade de que estes seis usuários sejam remetentes de uma mensagem seja a mesma e igual a 10

Dos Gráficos do Cenário 2.1 pode-se inferir que, assim como nos Cenários 1 e 2, a en-

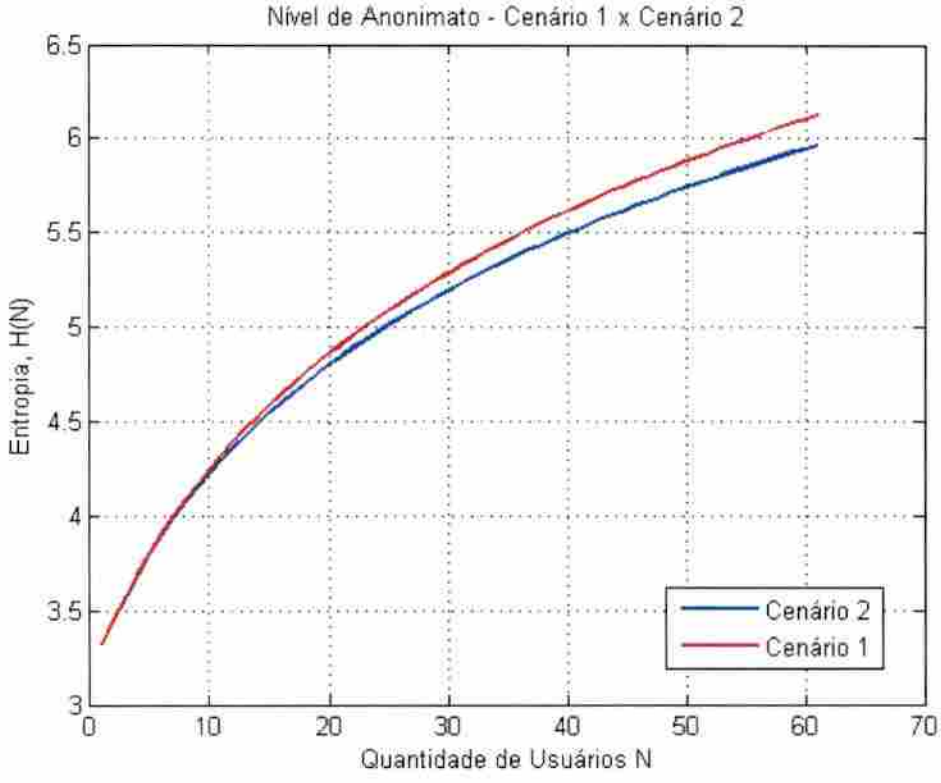


Figura 4.9: Nível de Anonimato - Cenário 1 x Cenário 2

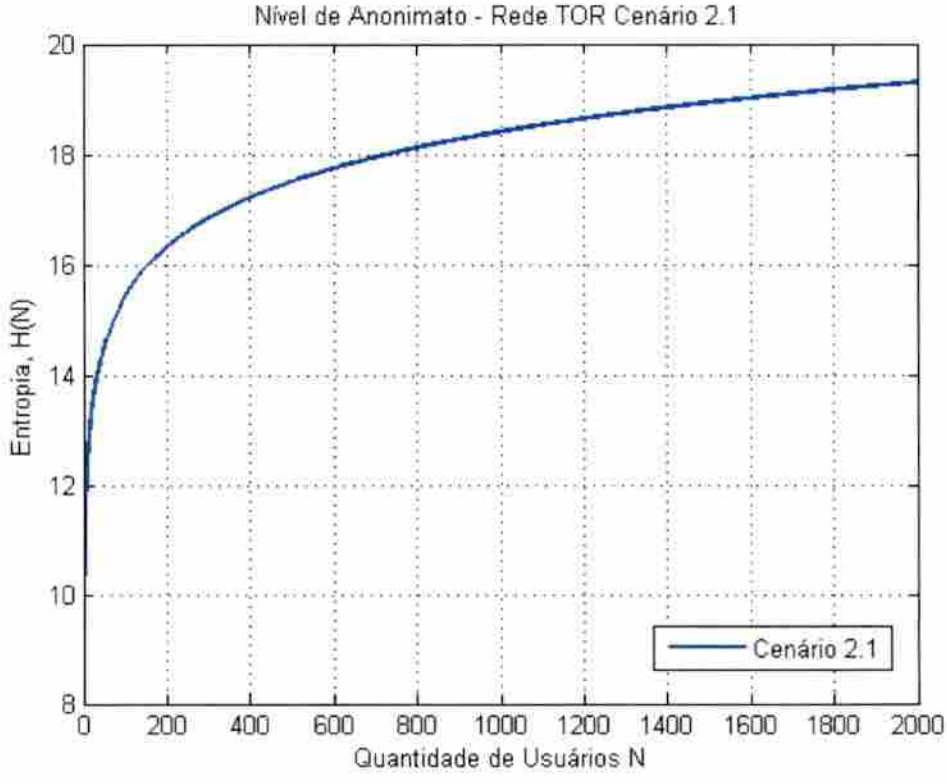


Figura 4.10: Nível de Anonimato - TOR Cenário 2.1

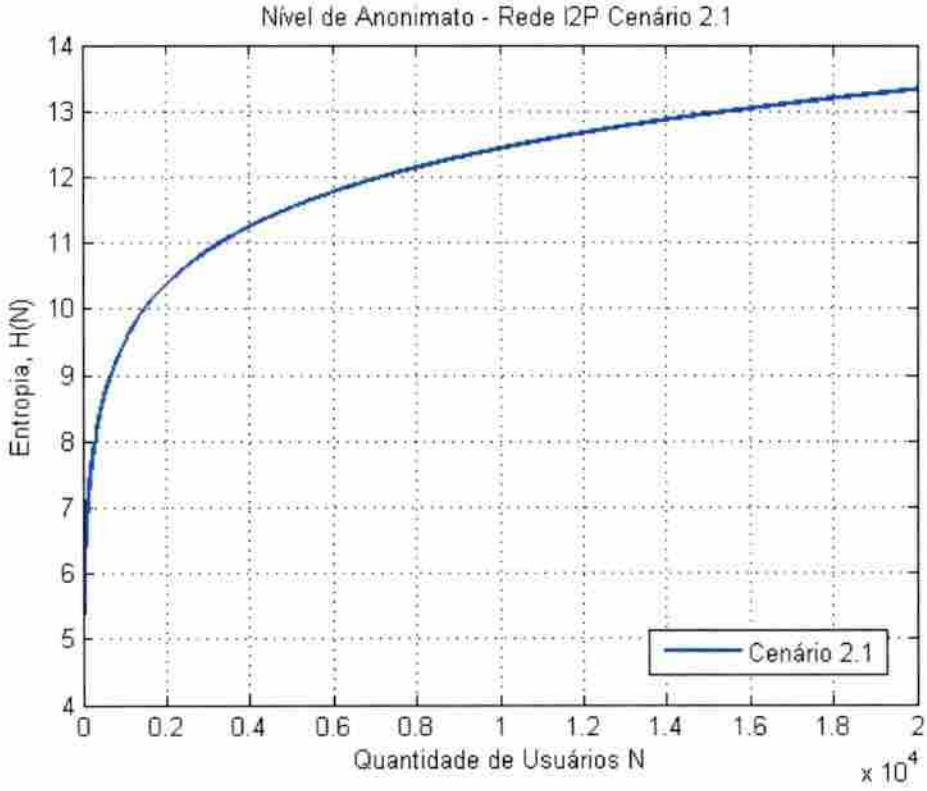


Figura 4.11: Nível de Anonimato - I2P Cenário 2.1

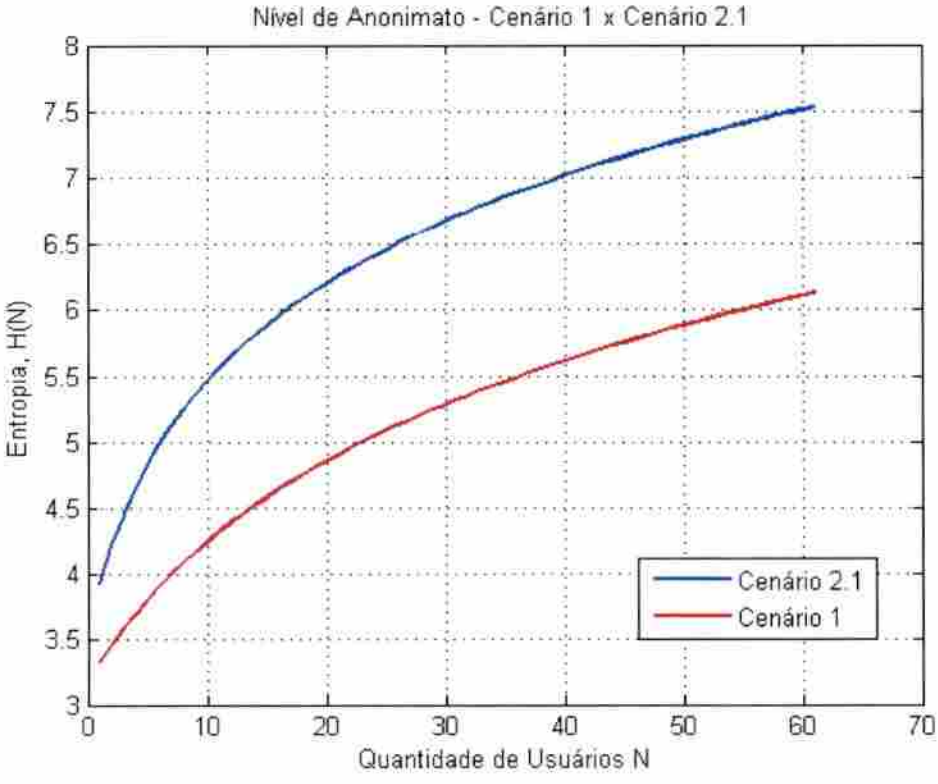
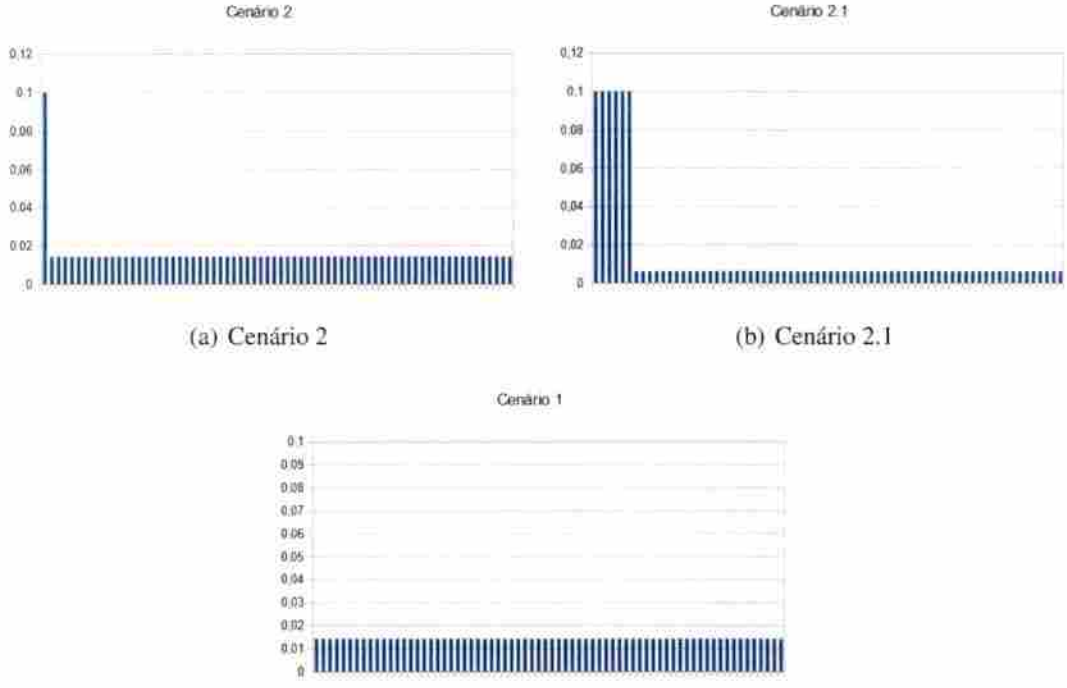


Figura 4.12: Nível de Anonimato - Cenário 1 x Cenário 2.1

tropia aumenta à medida que a quantidade de usuários na rede aumenta. Tanto o Gráfico 4.9 quanto o 4.12 mostram que tanto a quantidade de usuários quanto a probabilidade associada a cada um deles influencia no nível de anonimato que a rede oferece, segundo a definição de nível de anonimato dada pela entropia de Shannon. Estes Gráficos também mostram que a distribuição de probabilidade para casos distintos, isto é, o modelo de ameaça e o conhecimento que um adversário possui, influencia também no nível de anonimato.

Comparando-se os Cenários 2 e 2.1, no primeiro o adversário ativo é capaz de investigar apenas um usuário, ou seja, sabe que esse usuário possui uma probabilidade p_i de ser remetente de uma dada mensagem na rede, enquanto no segundo o adversário ativo é capaz de investigar seis usuários da rede, e sabe com que probabilidade cada um dos seis usuários pode ser o remetente de uma dada mensagem. No Cenário 2.1 analisado, a distribuição de probabilidade é tal que os seis usuários que podem ser investigados pelo adversário possuem a mesma probabilidade p_i de serem os remetentes de uma mensagem observada, e o restante dos usuários possuem a mesma probabilidade q_i de serem os remetentes dessa mensagem. As Figura 4.13(a) e 4.13(b) mostram as distribuições de probabilidade que cada usuário na rede tem de ser o remetente de uma mensagem observada para os Cenários 1, 2 e 2.1 considerando $N = 70$ usuários.



(a) Cenário 2

(b) Cenário 2.1

(c) Cenário 1

Figura 4.13: Distribuição de Probabilidade

Analisando apenas o valor final de entropia de cada cenário, percebe-se que o Cenário 1 apresenta o maior nível de anonimato se comparado ao Cenário 2, o que mostra o Gráfico 4.9, mas apresenta menor nível de anonimato em relação ao Cenário 2.1, como pode-se inferir do Gráfico 4.12. Se for levado em consideração apenas o valor da entropia de shannon, poderia-

se chegar a conclusão de que o Cenário 2.1 é o que fornece maior nível de anonimato em relação aos outros dois. Entretanto, a probabilidade de que um adversário tenha sucesso ao identificar um usuário como sendo o remetente de uma mensagem é maior no Cenário 2.1, com 10

De maneira simplista, se for considerada apenas a análise qualitativa feita por meio do cálculo do nível de anonimato dado pela entropia de Shannon, a rede TOR pode oferecer maior nível de anonimato para seus usuários, por possuir uma maior quantidade de usuários do que a quantidade estimada na rede I2P.

4.3 Análise Qualitativa TOR x I2P

Baseado nos resultados obtidos na Seção 4.2.3, pode-se fazer uma classificação qualitativa das redes TOR e I2P de acordo com as definições de graus de anonimato apresentadas na Seção 2.1.1.

Dos dados observados nos gráficos apresentados na seção 4.2.3, considerando-se o nível de anonimato dado pelo valor de entropia de Shannon, a quantidade de usuários na rede, o modelo de ameaça e a distribuição de probabilidade dos usuários em relação a uma mensagem, baseado na escala de graduação de anonimato apresentada na seção 2.1.1 pode-se classificar qualitativamente o grau de anonimato de um usuário nas redes analisadas. Um usuário no Cenário 1 pode ser considerado acima de suspeita, tanto na rede TOR quanto na rede I2P. No Cenário 2, um usuário é possível inocente, assim como no Cenário 2.1.

Apesar de muito semelhantes em alguns aspectos técnicos, as redes TOR e I2P possuem algumas diferenças principais que são diretamente relacioandas ao tipo de anonimato que o usuário pode esperar do uso de cada uma delas contra determinados tipos de ataques. O ponto principal, apresentado na análise comparativa anterior, é que a rede TOR é centralizada e seus roteadores são conhecidos, diferentemente da I2P, que é uma rede ponto a ponto em que os nós roteadores são usuários comuns, que podem ser facilmente manipulados para ataques em que os nós maliciosos conseguem obter várias informações sobre o tráfego e usuários na rede. Para compartilhamento de arquivos, a rede I2P pode se mostrar melhor, mas para a navegação com mais desempenho, a rede TOR proporciona melhor experiência.

Capítulo 5

Conclusão

Este trabalho tinha como objetivo apresentar uma análise quantitativa e qualitativa sobre o nível de anonimato das redes TOR e I2P. Para isso foram realizadas simulações por meio de ferramenta matemática que devolve em forma de gráficos os resultados quantitativos do nível de anonimato calculado pela entropia de Shannon de cada rede dadas as variáveis de quantidade de usuários e modelo de ameaça, representado pela probabilidade de cada usuário ser o remetente real de uma mensagem trafegada na rede. A análise qualitativa foi feita após a quantitativa, utilizando a classificação de grau de anonimato apresentada no Capítulo 2.

Como apresentado nos capítulos anteriores, pode-se concluir que tanto a rede TOR quanto a rede I2P proveem um bom nível de anonimato, e a escolha entre uma e outra depende da finalidade do uso. Se o usuário busca melhor performance na navegação e pesquisa na Internet, a rede TOR pode ser mais recomendada. Já se o objetivo é o compartilhamento de arquivos, a rede I2P pode se mostrar mais interessante. Deve-se lembrar, como explicado no Capítulo 3, que o usuário deve saber configurar sua conexão para aumentar o nível de anonimato e segurança ao utilizar essas redes.

Os resultados das simulações mostram que, quanto maior o número de usuários na rede, maior o nível de anonimato que ela pode oferecer. Não apenas isso, o modelo de ameaça também tem grande influência no nível de anonimato oferecido. Do capítulo 4 pode-se concluir que o modelo de ameaça, se o adversário é passivo ou ativo, e o tipo de distribuição de probabilidade influenciam diretamente no nível de anonimato e que, dependendo da distribuição, mesmo que o valor de entropia para a rede seja alto, talvez o modelo seja pior para a privacidade do que um cenário com menor valor de entropia. Ainda, pode-se notar, pelas equações apresentadas, que é possível se obter o mesmo nível de anonimato em redes com quantidade de usuários diferentes, desde que o modelo de ameaça de cada uma também seja distinto. Ainda, para cada distribuição de probabilidade, isto é, para cada tipo de adversário, a análise retorna um resultado diferente, sendo necessária uma análise caso a caso.

Em geral conlcui-se que, como esperado, quanto maior a quantidade de usuário, maior o nível de anonimato e quanto mais uniforme a distribuição de probabilidade entre os usuários,

melhor o nível de anonimato e mais difícil é, para um atacante, descobrir a identidade de um usuário no conjunto de nós observado.

Para continuidade dos estudos relacionados ao anonimato e nível de anonimato de redes como o TOR e I2P, sugere-se a implementação de um ambiente de teste em laboratório com o mínimo de poder computacional em que os diferentes modelos de ameaça possam ser aplicados: adversário passivo e adversário ativo. Pode-se coletar dados a partir de análise de tráfego em um conjunto de nós em cada rede, seja apenas observando ou seja incluindo nós maliciosos na rede, de forma prática, implementando ataques conhecidos, como por exemplo ataques Sybil, dentre outros.

Referências Bibliográficas

[Conrad and Shirazi 2014] Conrad, B. and Shirazi, F. (2014). A survey on tor and i2p. In *Proceedings of the 9th International Conference on Internet Monitoring and Protection (ICIMP 2014).*

[Danezis 2004] Danezis, G. (2004). *Designing and attacking anonymous communication systems.* Technical Report Number 594, UCAM-CL-TR-594 ISSN 1476-2986, July, 2004, 15 JJ Thomson Avenue, Cambridge CB3 0FD, United Kingdom.

[Dingledine et al. 2004] Dingledine, R., Mathewson, N., and Syverson, P. (2004). Tor: The second-generation onion router.

[Egger et al.] Egger, C., Schlumberger, J., Kruegel, C., and Vigna, G. Practical attacks against the i2p network. Technical report, University of California, Santa Barbara.

[Egger et al. 2013] Egger, C., Schlumberger, J., Kruegel, C., and Vigna, G. (2013). Practical attacks against the i2p network. In *Proceedings of the 16th International Symposium on Research in Attacks, Intrusions and Defenses (RAID 2013).*

[Müller 2016] Müller, J. (2016). Analysis of the i2p network - information gathering and attack evaluations. Bachelors thesis, Bern University of Applied Sciences. Title : Analysis of the I2P Network. School: Bern University of Applied Sciences - Department of Computer Science.

[Murdoch] Murdoch, S. J. Quantifying and measuring anonymity. Technical report, University of Cambridge Computer Laboratory.

[Newman et al. 2003] Newman, R. E., Moskowitz, I. S., Syverson, P., and Serjantov, A. (2003). Metrics for traffic analysis prevention. pages 1–18.

[Pfitzmann and Kohntopp 2001] Pfitzmann, A. and Kohntopp, M. (2001). Anonymity, unobservability, and pseudonymity - a proposal for terminology.

[Project 2017] Project, I. I. (2017). *I2P: A Scalable Framework For Anonymous Communication, https://geti2p.net/pt-br/docs/how/tech-intro.*

[Reiter and Rubin 1998] Reiter, M. and Rubin, A. (1998). Crowds: Anonymity for web transactions. In *ACM Transactions on Information and System Security (TISSEC).*

[Serjantov and Danezis] Serjantov, A. and Danezis, G. Towards an information theoretic metric for anonymity.