Dois Pesos, Duas Medidas: Gerenciamento de Identidades Orientado a Desafios Adaptativos para Contenção de *Sybils*

Gustavo Huff Mauch, Flávio Roberto Santos, Weverton Luis da Costa Cordeiro, Marinho Pilla Barcellos, Luciano Paschoal Gaspary

Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Porto Alegre – RS – Brasil

{ghmauch, frsantos, wlccordeiro, marinho, paschoal}@inf.ufrgs.br

Abstract. *The Sybil attack consists on the indiscriminate creation of counterfeit identities by a malicious user (attacker). An effective approach to tackle such attack consists of establishing computational puzzles to be solved prior to granting new identities. Despite its potentialities, solutions based on such approach do not distinguish between identity requests from correct users and attackers, and thus require both to afford the same cost per identity requested. To tackle this problem, in this paper we propose the use of adaptive computational puzzles to limit the spread of Sybils. We estimate a trust score of the source of identity requests in regard to the behavior of others. The higher the frequency a source requests identities, the lower its trust score and, consequently, the higher the complexity of the puzzle to be solved by the user(s) associated to that source. Results achieved by means of an experimental evaluation evidence our solution's ability to establish more complex puzzles to potential attackers, while minimally penalizing legitimate users.*

Resumo. *O ataque Sybil consiste na criação indiscriminada de identidades forjadas por um usuário malicioso (atacante). Uma abordagem promissora para mitigar esse ataque consiste em conceder novas identidades mediante a resolução de desafios computacionais. Apesar de suas potencialidades, as soluções baseadas em tal abordagem não distinguem solicitações de usuários corretos das de atacantes, fazendo com que ambos paguem o mesmo preço por identidade solicitada. Para lidar com esse problema, neste artigo propõe-se o uso de desafios adaptativos como limitante à disseminação de Sybils. Estima-se um grau de confiança da fonte de onde partem as solicitações de identidade em relação às demais. Quanto maior a frequência de solicitação de identidades, menor o grau de confiança e, conseqüentemente, maior a complexidade do desafio a ser resolvido pelo(s) usuário(s) associado(s) àquela fonte. Resultados obtidos por meio de experimentação mostram a capacidade da solução de atribuir desafios mais complexos a potenciais atacantes, penalizando minimamente usuários legítimos.*

1. Introdução

O ataque *Sybil* [Douceur 2002] representa um dos mais elementares ataques de autenticidade em redes P2P, e consiste na criação de múltiplas identidades falsas, denominadas identidades (ou pares) *Sybil*. A idéia motivadora desse ataque é que um atacante possa controlar a maioria, ou pelo menos uma parte significativa, das identidades presentes na rede. Deste modo, toda interação entre pares terá grande chance de ser mediada por uma das identidades controladas e alterada da forma que mais aprouver ao seu controlador [Barcellos and Gaspary 2006]. Um atacante com várias identidades falsas pode

também subverter algoritmos baseados em votação, manipulando com isso a reputação de pares ou conteúdos compartilhados na rede. Mais ainda, o ataque *Sybil* serve como base para o lançamento de outros ataques em redes P2P, tais como Eclipse [Singh et al. 2006] e *Free-riding* [Feldman et al. 2006].

Uma abordagem bastante promissora para mitigar ataques *Sybil* consiste em atribuir ou renovar a concessão de identidades aos usuários solicitantes mediante a resolução de desafios computacionais [Borisov 2006]. A idéia por trás da exigência da resolução de desafios é que pares legítimos provem suas boas intenções com a rede, comprometendo uma parte de seus recursos. Ao mesmo tempo, pares maliciosos interessados em criar múltiplas identidades serão obrigados a passar grande parte de seu tempo processando desafios e, portanto, consumindo recursos, o que reduz seu poder de assumir um número elevado de identidades.

Diversos trabalhos foram publicados propondo o emprego de desafios computacionais para o gerenciamento de identidades em redes P2P [Castro et al. 2002, Borisov 2006, Rowaihy et al. 2007]. Apesar de suas potencialidades, as propostas que adotam tal abordagem não fazem distinção entre solicitações de identidades oriundas de usuários corretos e de atacantes. A medida que ambos estão sujeitos ao pagamento do mesmo preço (computacional) por cada identidade solicitada, essas propostas podem não ser efetivas quando os recursos computacionais dos atacantes são muito superiores aos que os usuários legítimos dispõem. Assumindo desafios de uma determinada dificuldade, atacantes com *hardware* de maior capacidade conseguiriam resolver um conjunto muito superior de desafios e, com isso, obter um número elevado de identidades. Aumentar uniformemente a dificuldade dos desafios poderia, no outro extremo, tornar proibitivo o ingresso de pares legítimos à rede.

Para lidar com essa limitação, neste artigo propõe-se o uso de desafios adaptativos como estratégia de contenção contra a disseminação de *Sybils*. Em contraste com as propostas existentes na literatura, nossa solução estima um grau de confiança da *fonte* de onde parte a solicitação de identidade em relação ao comportamento das demais fontes. No contexto deste trabalho, *fonte* pode referir-se à estação de um usuário (identificada pelo seu endereço IP), à rede local a qual a estação pertença, a um sistema autônomo (*Autonomous System*, AS), etc. Essa decisão depende essencialmente da granularidade que se deseje ou seja possível adotar para a fonte (por exemplo, no caso de usuários posicionados atrás de redes usando NAT, a granularidade a ser considerada é associar todos os usuários daquela rede a uma única fonte). À medida que aumenta a frequência com que novas solicitações por identidades partem de uma dada fonte, diminui a confiabilidade da mesma. Consequentemente, maior será a complexidade do desafio computacional a ser resolvido antes que a identidade solicitada seja obtida pelo(s) usuário(s) associado(s) àquela fonte. Para avaliar a eficácia da solução proposta na contenção de ataques *Sybil*, foram realizadas simulações considerando traços históricos de solicitação de identidades em uma comunidade P2P. Os resultados obtidos mostram a capacidade da solução em atribuir desafios computacionais mais complexos a potenciais atacantes, ao passo que usuários legítimos são minimamente penalizados.

O restante do artigo está organizado como segue. A Seção 2 discute alguns dos principais trabalhos relacionados ao gerenciamento de identidades em redes P2P. A Seção 3 apresenta o mecanismo proposto para o uso de desafios adaptativos como uma proteção ao ataque *Sybil*, enquanto que a Seção 4 descreve a avaliação conduzida para avaliar a eficácia da mesma. A Seção 5 discute questões relacionadas ao emprego da solução proposta em arcabouços P2P. Por fim, a Seção 6 conclui o artigo com as considerações finais e possíveis desdobramentos para pesquisas futuras.

2. Trabalhos Relacionados

As investigações conduzidas para lidar com ataques *Sybil* podem ser classificadas de acordo com o mecanismo utilizado para garantir a autenticidade dos pares. As principais classes são *(i)* soluções baseadas em *identidades fracas*, e *(ii)* soluções baseadas em *identidades fortes*. A primeira reúne soluções em que cada par possui ampla autonomia para criar sua própria identidade. Nesse caso, são estabelecidas estimativas para número de identidades *Sybil* que são *aceitas* pelos demais pares. Tais estimativas podem ser úteis, por exemplo, para aplicações que tolerem uma fração previsível de pares *Sybil*. As soluções propostas por Yu Haifeng *et al.* [Yu et al. 2006, Yu et al. 2008] e George Danezis *et al.* [Danezis et al. 2005] são exemplos. Ambas exploram redes sociais para estimar o limite máximo de identidades *Sybil* presentes na rede em um dado momento. No entanto, elas apresentam problemas de violação de anonimidade e incapacidade de garantir autenticidade dos pares.

A segunda classe, por sua vez, reúne soluções nas quais os pares apenas podem obter identidades junto a entidades certificadoras. A principal vantagem é a dificuldade de um par criar e controlar várias identidades. No entanto, tais soluções podem diminuir a escalabilidade da rede P2P e inserir um ponto único de falha. Mais ainda, podem exigir que usuários confiem em entidades certificadoras desconhecidas, além de tornar inviável o acesso de potenciais usuários (por exemplo, quando for necessário informar dados pessoais ou pagar taxas para obter uma identidade). As propostas que se enquadram nessa classe buscam minimizar algumas das conseqüências danosas da introdução de uma entidade central. Por exemplo, em [Morselli et al. 2006] e [Aberer et al. 2005] os autores focaram na descentralização da infra-estrutura de distribuição de chaves públicas (*Public-Key Infrastructure*, *PKI*). No entanto, essas propostas requerem a troca de uma grande quantidade de mensagens entre os pares, além de dependerem da colaboração de um número mínimo de pares para funcionarem como esperado.

No artigo que descreve o ataque *Sybil* [Douceur 2002], Douceur argumenta que não é possível resolver o problema de autenticação sem fazer uso de algum grau de centralização. Com base nessa afirmativa, e considerando as severas limitações decorrentes do uso de entidades certificadoras, tem ganhado força a corrente de soluções em que a concessão de identidades é feita mediante a resolução de desafios computacionais. O principal objetivo é causar a diminuição da capacidade que usuários maliciosos tem de criar identidades falsas, sem abrir mão das características “natas” de redes P2P (por exemplo, escalabilidade, descentralização e autonomia dos pares).

As soluções baseadas em desafios computacionais tem apresentado bons resultados ao utilizarem desafios que são criados ou verificados de forma distribuída. Nikita Borisov [Borisov 2006], por exemplo, mostrou a viabilidade da utilização de desafios gerados de forma distribuída e periódica, propondo uma solução na qual os pares que participaram da geração do desafio são capazes de validar a resolução dos mesmos. Em [Rowaihy et al. 2007], por sua vez, propõe-se um esquema com múltiplas entidades geradoras de desafios. Esse esquema requer que, para a obtenção de identidades, usuários contatem uma dessas entidades e resolvam uma seqüência de desafios propostos.

Apesar de promissoras, as soluções existentes não abordam a questão do dimensionamento da complexidade dos desafios. Ao utilizarem desafios com mesma complexidade computacional para todos os usuários, o problema que surge é buscar o melhor ponto de equilíbrio entre usar desafios mais complexos para coibir atacantes com alto poder computacional, e não tão complexos, para não penalizar usuários legítimos com *hardware* menos capacitado. Nesse contexto, a utilização de um peso e uma medida na atribuição dos desafios tenderá a favorecer os atacantes em detrimento dos usuários

legítimos. O diferencial da proposta deste artigo é parametrizar a dificuldade dos desafios de acordo com o comportamento que cada fonte apresentar na rede. Os usuários associados a fontes cujo comportamento seja mais similar ao comportamento médio das demais fontes serão beneficiados com desafios menos complexos. Por outro lado, usuários associados a fontes com comportamento atípico deverão arcar com desafios mais custosos para a obtenção de identidades.

3. Proposta de Solução para Combater Ataques *Sybils*

A solução proposta neste artigo visa estabelecer o uso de desafios computacionais adaptativos para o gerenciamento de identidades em redes P2P. De uma forma geral, há três questões chave associadas à adoção de desafios adaptativos: (*i*) como caracterizar o comportamento das fontes (ou dos usuários associados às mesmas), (*ii*) como calcular o custo de um desafio a partir dos comportamentos observados, e (*iii*) como adaptar os desafios considerando a dinâmica do comportamento dos usuários da rede, observados pelo sistema como fontes de requisições de identidades. Cada uma dessas questões é abordada nas subseções a seguir.

3.1. Empregando *Taxas de Recorrências* para Caracterizar Comportamentos

Para permitir a caracterização do comportamento das diversas fontes de solicitação de identidades, duas métricas são introduzidas no contexto deste artigo: *taxa de recorrência da fonte* (ϕ) e *taxa de recorrência da rede* (Φ). A primeira reflete a frequência com que os usuários associados a uma determinada fonte solicitam novas identidades ao serviço de *bootstrap* da rede P2P em um intervalo de tempo t_w (com $t_w > 0$). A segunda, por sua vez, reflete a frequência média com que as fontes recorrem ao serviço de *bootstrap* para solicitar novas identidades.

O valor da taxa de recorrência da rede é calculado segundo a Equação 1, a qual utiliza a média harmônica das taxas de recorrências das fontes. Nessa equação, ϕ_i representa a taxa de recorrência da *i-ésima* fonte da rede P2P. Note que o caso em que $\phi_i = 0$ equivale à situação em que nenhum usuário da fonte i solicitou alguma identidade; nesse caso, tal fonte não é conhecida e, portanto, não é considerada para o cálculo da taxa de recorrência da rede.

$$\Phi = \frac{n}{\sum_{i=0}^{n} \frac{1}{\phi_i}} \tag{1}$$

É importante frisar que a média harmônica foi escolhida em detrimento de outras medidas estatísticas (média simples, média geométrica e a mediana) por ser especialmente resistente a alterações causadas por recorrências muito discrepantes do padrão observado (*outliers*). Essa característica é desejável e extremamente importante, uma vez que torna mais difícil para atacantes manipularem o comportamento da rede (por exemplo através de um ataque em conluio) para tornarem-se menos suspeitos. A resistência da taxa de recorrência da rede a ataques em conluio é avaliada em maior profundidade na Seção 4.

3.2. Calculando o Grau de Confiança a partir dos Comportamentos Observados

Considerando que o objetivo de um ataque *Sybil* é controlar uma fração significativa de identidades na rede P2P, para executá-lo o atacante deverá solicitar um grande número de identidades ao serviço de *bootstrap*. A consequência direta desse comportamento é um aumento da taxa de recorrência da fonte associada ao atacante. Por outro lado, é esperado

que as fontes com usuários legítimos recorram minimamente para solicitar identidades (por exemplo, no momento que se registrarem na rede P2P). Logo, a idéia principal para conter ataques *Sybil* é atribuir desafios mais complexos ao(s) usuário(s) associado(s) às fontes cujas taxas de recorrência se tornarem superiores à taxa de recorrência da rede.

A partir da comparação entre o comportamento de cada fonte (inferido a partir de ϕ) e do comportamento considerado padrão para a rede (inferido a partir de Φ), é calculada a *relação entre as taxas de recorrências da fonte e da rede* (ρ). Obtida de acordo com a Equação 2, ela assume valores menores que zero para indicar quantas vezes a taxa de recorrência da fonte *i* é menor que a da rede, e maiores que zero para indicar quantas vezes a taxa de recorrência da fonte *i* é maior.

$$\rho = \begin{cases} -\frac{\Phi(t)}{\phi_i(t)} & \text{se } \phi_i(t) < \Phi(t) \\ \frac{\phi_i(t)}{\Phi(t)} & \text{se } \phi_i(t) \geq \Phi(t) \end{cases} \tag{2}$$

A relação entre as taxas de recorrências da fonte e da rede (ρ) serve como base para o cômputo do *grau de confiança da fonte* origem das solicitações de identidade (C). Esse grau, estimado para cada instante t de acordo com a Equação 3, assume valores no intervalo $[0, 1]$: em um extremo, valores mais próximos de 1 indicam uma maior confiança sobre a legitimidade do(s) usuário(s) associado(s) à fonte em questão; no outro extremo, valores mais próximos de 0 indicam maior desconfiança, isto é, uma maior probabilidade de que o(s) usuário(s) associados à fonte em questão está(ão) lançando um ataque *Sybil*. A complexidade do desafio computacional aplicado ao(s) usuário(s) será determinada pelo grau de confiança da fonte ao(s) qual(is) ele(s) está(ão) associado(s), no momento da solicitação de uma nova identidade. A Equação 3 é normalizada de modo que os extremos 0 e 1 representem total desconfiança e total confiança sobre uma determinada fonte, respectivamente.

$$C(t) = 0.5 - \frac{\arctan(a \times (\rho - c)^{(1+2\times b)})}{\pi} \tag{3}$$

A Figura 1 mostra quatro diferentes configurações que ilustram como varia o grau de confiança obtido para uma determinada fonte em função de ρ. Nessas configurações, os termos a, b e c da Equação 3 assumem valores arbitrários e desempenham um importante papel no controle da agressividade com que as configurações decrescem, da amplitude e da translação das mesmas, respectivamente.

A partir da Figura 1 é possível observar duas propriedades importantes que a Equação 3 apresenta. A primeira reside no fato do valor de confiança ter variações mínimas para valores de ρ mais próximos de 0 (situação em que a fonte se comporta de forma semelhante ou igual à média da rede), proporcionando assim uma certa tolerância na avaliação dos comportamentos das fontes. Considerando por exemplo a configuração $(a = 0,1\ ,\ b = 2\ ,\ c = 0)$ na Figura 1, dentro do intervalo $-2 \leq \rho \leq 2$, variações são pouco consideradas, por serem ligeiramente semelhantes ao padrão observado na rede. Os comportamentos que desviam significativamente desse intervalo, no entanto, terão atribuídos menores (ou maiores) valores de confiança, como pode ser observado pelas súbitas variações da configuração $(a = 0,1\ ,\ b = 2\ ,\ c = 0)$ nos intervalos $-5 \leq \rho \leq -2$ e $2 \leq \rho \leq 5$. A segunda propriedade reside no fato de ser assintótica em 0 e 1. Desse modo, para $\rho \to -\infty$ ou $\rho \to +\infty$, sempre haverá um valor de confiança associado.

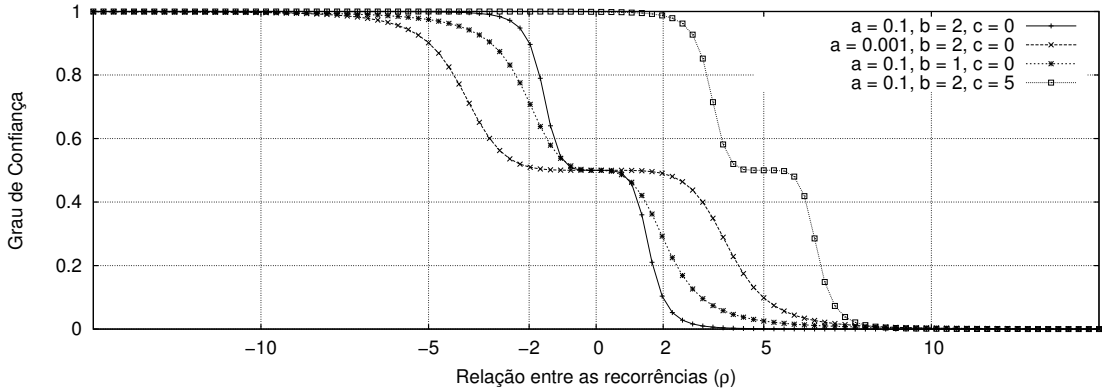


Figura 1. Exemplos de valores para os parâmetros a, b, e c da Equação 3 para cálculo do grau de confiança da fonte

3.3. Lidando com a Dinâmica do Comportamento dos Usuários da Rede

Uma característica importante de redes P2P é a ampla autonomia concedida aos pares. Desse modo, os pares podem entrar e sair da rede de acordo com seus interesses e disponibilidade, sem depender de entidades externas. Um dos possíveis desdobramentos dessa dinamicidade é a ocorrência de variações constantes (e eventualmente significativas) do padrão de comportamento tanto das fontes quanto da rede P2P como um todo. A seguir, é discutido como o mecanismo proposto lida com a dinâmica dos comportamentos observados.

A Equação 3, embora seja capaz de determinar a confiança de uma determinada fonte no instante t, não considera o histórico de comportamento da mesma. Com o objetivo de representar de modo apropriado o grau de confiança de uma determinada fonte, ao mesmo tempo considerando o histórico do comportamento da mesma, é inserido na solução um parâmetro β, o qual permite o cálculo da confiança suavizada, C_s, conforme apresentado na Equação 4. O parâmetro β é um fator de suavização que determina o peso do passado no cálculo do valor de confiança para o instante atual (t), e assume valores no intervalo (0,1]. Em um extremo, valores de β mais próximos de 0 conferem um peso maior ao comportamento histórico da fonte em questão. Em outro extremo, valores de β mais próximos de 1 dão um peso maior ao comportamento atual da fonte. No caso especial em que $\beta = 1$, o valor de confiança atual (tal como calculado pela Equação 3) é considerado integralmente, sendo o passado totalmente desconsiderado.

$$C_s(t) = \beta \times C(t) + (1 - \beta) \times C(t - 1) \tag{4}$$

A adição do parâmetro β ao cálculo da confiança é importante para tratar adequadamente as alterações no comportamento de cada fonte de solicitação de identidades. Em particular, as alterações intencionais e repentinas de comportamento, de usuários interessados em obter benefícios, como os *traidores*, são capturadas e refletidas no grau de confiança da fonte à qual o mesmo está associado. Um traidor é um atacante que busca angariar altos valores de confiança em sistemas de reputação e passa, então, a se aproveitar dela para prejudicar outros pares, ou obter vantagens indevidas. O correto dimensionamento do valor de β, nesse contexto, pode impedir que um traidor manipule a solução proposta de modo que a fonte em que se situa consiga (ou recupere) rapidamente uma alta confiança do sistema. Na medida em que o passado é considerado para determinar o presente, somente aquelas fontes cujos usuários apresentem bom comportamento

histórico serão considerados confiáveis.

Outra questão importante, ainda em relação à dinâmica do comportamento da rede, reside no fato de que as taxas de recorrência podem variar em épocas diferentes. Por exemplo, é razoável esperar que em determinados períodos mais usuários estejam interessados em ingressar na rede P2P e, consequentemente, mais requisições por identidades sejam realizadas. Por outro lado, também é razoável esperar uma queda no número de usuários que ingressam na rede em outros períodos, o que se reflete em menos requisições. Sem considerar essa sazonalidade no comportamento dos usuários (e da rede como um todo), usuários legítimos podem ser considerados suspeitos pela solução proposta caso suas fontes solicitem identidades com maior frequência, mesmo que estejam acompanhando o comportamento dos demais usuários. Por outro lado, caso todas as solicitações desde o início ($t = 0$) fossem consideradas, seria mais fácil para um atacante lançar mão de ataques *Sybils*. Isso seria possível visto que a quantidade de requisições cresceria indefinidamente, consequentemente ofuscando as altas taxas de recorrência de fontes suspeitas.

Para acomodar questões de sazonalidade no comportamento dos pedidos de identidades, optou-se por utilizar uma *janela deslizante* – um intervalo de tempo t_w, que se inicia no passado e termina no momento presente – para restringir a quantidade de requisições a serem consideradas no cálculo das taxas de recorrência de cada fonte e da rede. Note que t_w corresponde ao tempo considerado para calcular a taxa de recorrência ϕ de cada fonte no sistema e, consequentemente, a taxa de recorrência da rede, Φ (tal como discutido na Subseção 3.1). A medida em que o tempo passa, a janela avança em passos com duração t_d (com $t_d \leq t_w$); com isso, as solicitações de identidade mais antigas vão sendo desconsideradas, dando lugar à solicitações mais recentes, as quais são mais representativas do estado atual da rede P2P.

4. Avaliação da Solução Proposta

Para avaliar a viabilidade técnica, a eficácia e a eficiência do uso de desafios adaptativos para combater *Sybil* em redes P2P, foi realizada a implementação prototípica de um serviço de *bootstrap*. Por meio desse protótipo, foram executados diversos experimentos, considerando solicitações sintéticas de identidades baseadas em traços históricos de uma comunidade P2P. Como resultados da avaliação conduzida, procurou-se observar que (*i*) os desafios computacionais propostos para usuários legítimos penalizam minimamente os mesmos, (*ii*) desafios atribuídos a potenciais atacantes possuem maiores complexidades computacionais, e (*iii*) a solução proposta é robusta e resiliente mesmo na presença de uma fração significativa de atacantes, bem como sob a ocorrência de ataques em conluio.

O restante desta seção está organizado como segue. A Subseção 4.1 descreve a configuração do ambiente considerado na análise (características dos traços históricos de solicitações de identidades usados, parâmetros da solução, etc.). A Subseção 4.2, por sua vez, apresenta os resultados obtidos pela solução proposta na contenção de ataques *Sybil*.

4.1. Configuração do Ambiente de Experimentação

A Tabela 1 apresenta um resumo das características do traço utilizado nos experimentos, dos valores para os parâmetros envolvidos na solução, e a característica dos ataques *Sybil* considerados na análise da eficácia e eficiência da proposta. Cada um destes é discutido em detalhes a seguir.

Os experimentos foram feitos com base em traços históricos de solicitações de identidades na comunidade P2P fechada Bitsoup [Bitsoup.org 2009]. Uma vez que a admissão na comunidade é feita mediante autenticação por usuário e senha, e a criação de

Tabela 1. Informações sobre o ambiente considerado na avaliação experimental.

Características do Traço Empregado	
Duração	15 dias
Quantidade de identidades solicitadas	625.079 identidades
Número de fontes distintas	44.315 fontes
Intervalo médio entre requisições	2,08 segundos
Quantidade média de requisições por fonte	14,21 identidades
Parâmetros da Solução	
a	0,1
b	2
c	5
β	0,125
Duração da Janela (t_w)	8 horas
Passo da Janela (t_d)	1 hora
Estratégias de Ataque	
Taxa de requisição por fonte atacante	1; 1,25; 1,5; 2; e 2,5 requisições/hora
Quantidade de fontes atacantes	1; 100; 500; 1.000; e 2.000

novas contas é moderada, assumiu-se para os fins da avaliação que o traço não contém registros de ataques *Sybils*. Essa premissa baseia-se na idéia de que o custo necessário para criar e manter diversas identidades falsas em uma comunidade fechada com moderação é de várias ordens de grandeza maior do que em uma comunidade aberta sem moderação.

Os traços considerados registram atividades de solicitação de identidades durante 15 dias consecutivos. Durante esse período, foram obtidas 625.079 identidades, solicitadas por 44.315 fontes distintas, perfazendo uma média de $14,21$ solicitações de identidades por fonte, e uma taxa global de 1 solicitação a cada $2,08$ segundos.

Os parâmetros envolvidos na experimentação foram definidos como segue. O valor de β foi definido como $0,125$, isto é, o comportamento histórico do grau de confiança possui um peso de 87,5% sobre o valor atual. Esses valores mostraram-se adequados, após sucessivas experimentações (omitidas neste artigo por questões de restrição de espaço), para impedir que fontes com histórico de *mal comportamento* alcançassem valores elevados de confiança ao tornarem-se repentinamente *bem comportadas*. Os valores de a, b e c, por sua vez, foram definidos como $0, 1$, 2 e 5, respectivamente, visando controlar a forma como a relação entre as recorrências se reflete no grau de confiança obtido pela fonte. Com esses valores, uma taxa de recorrência da fonte similar ou igual a da rede ($\phi \simeq \Phi$) fará com que a fonte alcance um grau de confiança de aproximadamente 1 (por exemplo, vide configuração $a = 0,1$, $b = 2$ e $c = 5$ na Figura 1). A janela deslizante tem duração de 8 horas ($t_w = 8 \times 60$ *min*) e desliza de hora em hora ($t_d = 1 \times 60$ *min*). A duração de 8 horas mostrou ser adequada considerando as características da comunidade Bitsoup.org (materalizadas nos traços históricos estudados), sendo capaz de capturar adequadamente o comportamento passado de cada usuário, e ao mesmo tempo desconsiderando solicitações que não refletem mais o estado atual da rede. O deslizamento de hora em hora, por sua vez, mostrou-se adequado para capturar a evolução nos comportamentos dos participantes da rede, sem impor uma sobrecarga maior ao processo de cálculo do grau de confiança.

Para avaliar cenários em que a rede encontra-se sob ataque *Sybil*, foram injetadas

artificialmente solicitações maliciosas de identidades, considerando duas estratégias diferentes. Na primeira, o atacante lança um ataque *Sybil* a partir de uma única fonte. A segunda estratégia, por sua vez, considerou que o atacante possui a sua disposição um determinado número de fontes. Nesse caso, o ataque *Sybil* realizado é distribuído, com solicitações partindo de cada uma das fontes sob o controle do atacante – cada fonte solicita uma quantidade pequena de identidades, de modo que não sejam classificadas como suspeitas. Em ambos os casos, busca-se avaliar a quantidade de identidades que um atacante consegue solicitar por meio do ataque *versus* a dificuldade do desafio que o sistema atribui para cada nova solicitação vinda de uma das fontes envolvidas no ataque.

4.2. Resultados Obtidos e Análise

Para organizar a discussão dos resultados obtidos, primeiramente é discutida a sobrecarga causada a usuários legítimos, em situações em que não há ocorrência de ataques *Sybil* na rede P2P. Em seguida, é avaliada a efetividade da solução na contenção de ataques *Sybil*, e o impacto que estes causam nos desafios com os quais usuários legítimos terão de arcar. Por fim, é analisada a sua resiliência em situações em que diversos atacantes agem em conluio, com o propósito específico de atacar a própria solução.

Sobrecarga Causada a Usuários Legítimos na Ausência de Ataques *Sybil*

A Figura 2 exibe a função cumulativa complementar de distribuição (*Complementary Cumulative Distribution Function*, *CCDF*) das confianças calculadas para as solitações de identidades partindo das fontes (consideradas legítimas) do traço estudado. É importante ressaltar que esse resultado refere-se somente às solicitações de identidade presentes no traço original, não tendo sido perturbado pela ocorrência de ataques *Sybil*.

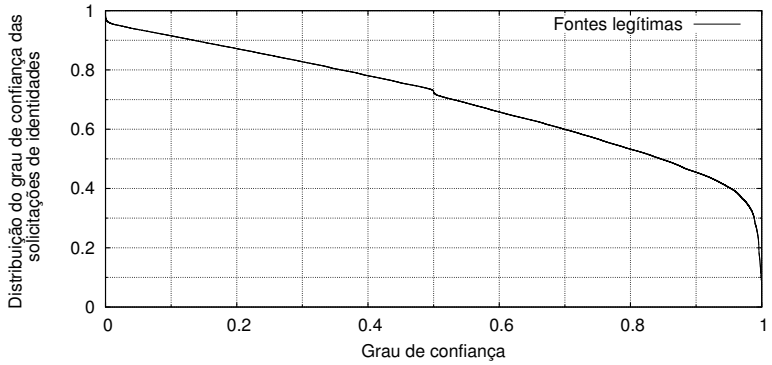


Figura 2. *CCDF* do grau de confiança de solicitações de identidades originadas por fontes legítimas

É possível notar no gráfico da Figura 2 que a maioria das solicitações de identidade geradas pelas fontes é de alta confiança. Por exemplo, aproximadamente 45% das solicitações de identidades foram realizadas por usuários oriundos de fontes com confiança maior ou igual a $0,9$. Em outras palavras, existe um grau de confiança igual ou maior do que $0,9$, para aproximadamente 45% das solicitações, de que as mesmas não estejam relacionadas a um ataque *Sybil*. Esse percentual aumenta para 60% se considerarmos as solicitações com confiança maior ou igual a $0,7$, e para aproximadamente 75% se considerarmos aquelas com confiança maior ou igual a $0,5$. Com esses valores de confiança obtidos, uma significativa fração das fontes obterá desafios computacionais de menor complexidade, logo causando mínimo impacto para os respectivos usuários.

Um aspecto importante a ser discutido sobre os resultados da Figura 2 diz respeito aos 25% das fontes com confiança menor que $0,5$. Embora as fontes contidas no traço sejam presumidamente legítimas (isto é, não lançaram algum ataque *Sybil* contra a rede P2P), existem casos em que as fontes podem recorrer mais vezes que a média da rede para solicitar identidades. Esse é o caso, por exemplo, em que vários usuários acessam a Internet através de redes utilizando o mecanismo de NAT, o qual faz com que os mesmos sejam associados à uma única fonte. De qualquer forma, o número de usuários afetados no experimento executado foi mínimo. Pouco mais de 10% das fontes alcançou valores de confiança menores ou iguais a $0,2$.

Impacto Causado a Potenciais Atacantes

A Figura 3 apresenta os resultados obtidos a partir do desdobramento do cenário ilustrado na Figura 2 em cinco novos cenários, cada um sob os efeitos de um ataque *Sybil* gerado artificialmente. Os ataques, em cada um dos cenários, são orquestrados por uma única fonte (maliciosa). A principal diferença entre os mesmos reside nas taxas de solicitação de identidades adotadas: 1; $1,25$; $1,5$; 2; e $2,5$ solicitações por hora, respectivamente.

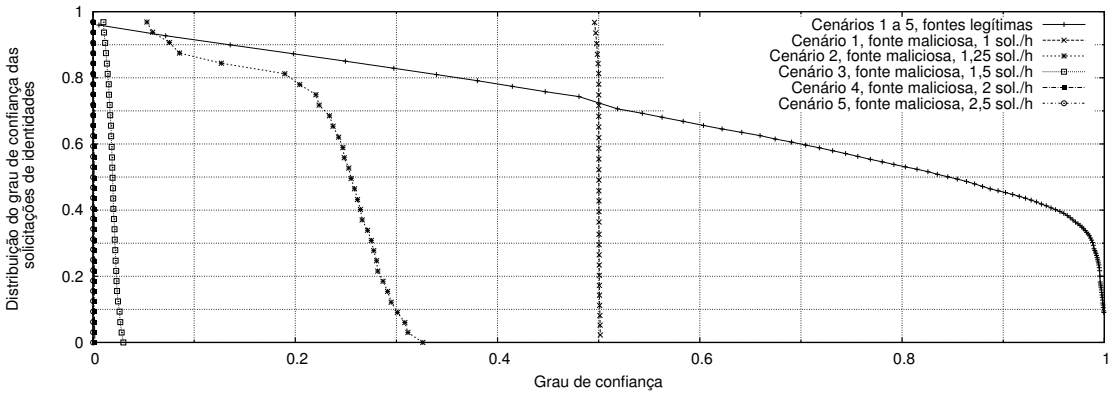


Figura 3. Resistência da solução proposta à ataques *Sybil* partindo de uma única fonte maliciosa, considerando diferentes taxas de recorrência da mesma

Uma observação importante em relação aos resultados apresentados na Figura 3 corresponde à influência do ataque *Sybil* sobre o grau de confiança obtido pelas fontes legítimas. Independente da taxa de solicitação de identidades adotada pelo atacante, as curvas que mostram a distribuição do grau de confiança dos pares legítimos mantêm-se inalteradas e idênticas. Tal se deve à resistência da média harmônica – medida empregada para calcular a taxa de recorrência da rede, conforme discutido na Seção 3.1 – à presença de taxas de recorrência com desvio significativo em relação as das demais fontes. Por questões de legibilidade, apenas uma curva é apresentada na Figura 3 para ilustrar a distribuição do grau de confiança das fontes legítimas.

Analisando os resultados da Figura 3 por uma perspectiva diferente, é possível observar que um aumento gradual na taxa de recorrência da fonte maliciosa é suficiente para que a mesma sofra quedas significativas no seu grau de confiança. Por exemplo, quando a taxa de recorrência da fonte maliciosa corresponde a 1 solicitação por hora (cenário 1), aproximadamente 100% das solicitações de identidades partindo daquela fonte obteve grau de confiança igual a $0,5$ (isto é, um grau de confiança de $0,5$ de que a solicitação não está relacionada a um ataque *Sybil*). Para a taxa de $1,25$, por sua vez, apenas 10% das solicitações de identidades obteve grau de confiança maior ou igual a $0,3$. No cenário 5,

o mais extremo ilustrado, a taxa de $2,5$ faz com que todas as solicitações de identidades partindo da fonte maliciosa sejam consideradas como parte de um ataque *Sybil* (uma vez que 100% das solicitações de identidade obteve grau de confiança 0). A consequência direta das quedas observadas é a imposição, aos usuários associados à fonte maliciosa, de desafios computacionais de complexidade computacional extrema.

Os resultados apresentados levam a duas conclusões distintas, dependendo da perspectiva pela qual são analisados. Por um lado, evidenciam que a solução proposta reage adequadamente ao aumento na taxa de recorrência das fontes, penalizando severamente aquelas que recorrem a uma taxa muito maior que a média observada na rede. Por outro lado, mostra que a solução compele as fontes a se "comportarem adequadamente" – isto é, recorrendo harmonicamente em relação às demais fontes – caso não desejem ser penalizadas com desafios computacionais mais complexos.

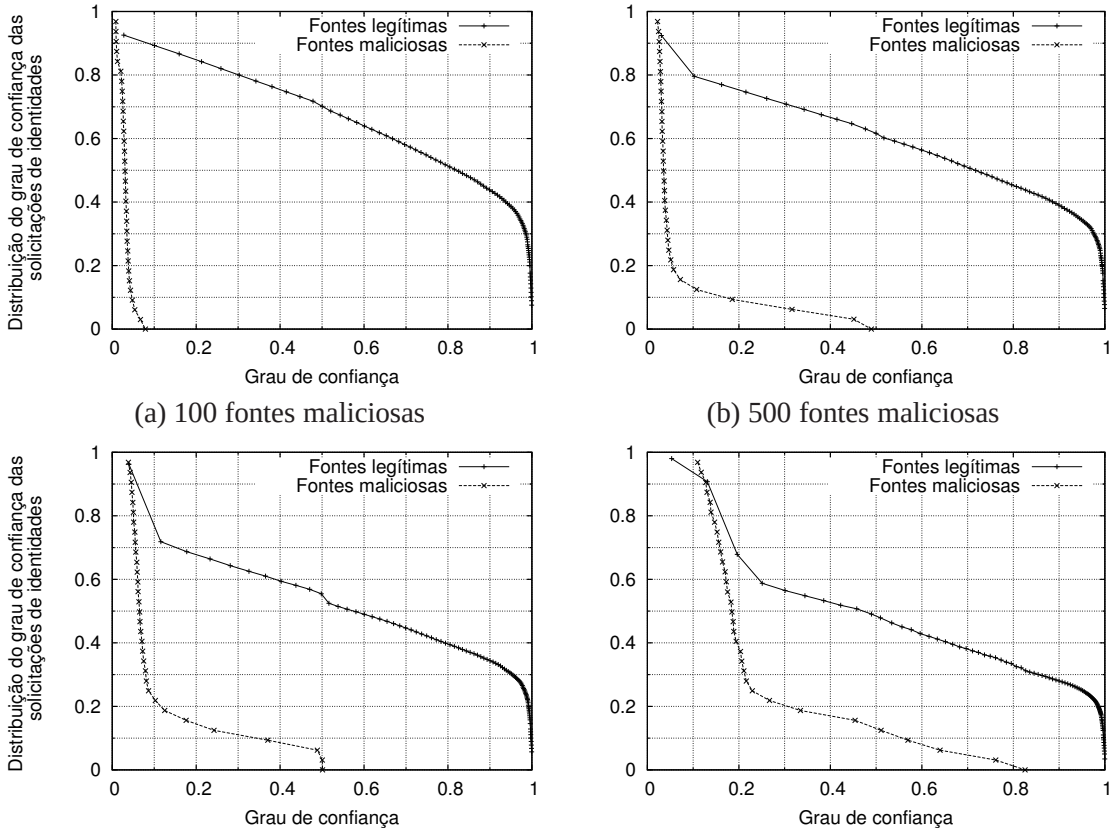
Resiliência da Solução Proposta a Ataques em Conluio

Após ter-se analisado o efeito de um ataque *Sybil* considerando uma única fonte, avaliou-se o efeito do ataque realizado de forma distribuída, isto é, considerando múltiplas fontes como origem das solicitações de identidade. Nesse caso, ao invés de aumentar a taxa de recorrência para obter mais identidades falsas, o atacante age em conluio com outros atacantes (ou lança mão de uma *botnet* formada por várias estações zumbi conectadas à Internet). São dois os objetivos do lançamento de um ataque *Sybil* em conluio. Primeiro, busca-se aumentar a velocidade com que o atacante obtém identidades falsas na rede P2P, sem ter de arcar com desafios mais complexos. Segundo, procura-se alterar a percepção de normalidade da rede. Em outras palavras, parte-se da idéia de que mais fontes maliciosas atuando com o mesmo comportamento tende a mudar a percepção sobre qual é, efetivamente, o comportamento da maioria das fontes na rede.

A Figura 4 apresenta os resultados obtidos considerando a nova estratégia de ataque. Quatro cenários distintos são considerados, cada um com um número distinto de fontes maliciosas à disposição do atacante: 100; 500; 1000; e 2000 fontes. Em todos os cenários, cada fonte maliciosa atua a uma taxa de $1,5$ solicitações de identidades por hora. Essa taxa foi escolhida porque permite ao atacante obter um número significativo de identidades e, ao mesmo tempo, passar mais despercebido como um atacante na rede (conforme evidenciado na análise anterior).

Observe na Figura 4 que, mesmo utilizando uma quantidade extremamente alta de fontes para lançar o ataque *Sybil*, o efeito que o atacante consegue exercer sobre o padrão de normalidade da rede é relativamente limitado. Por exemplo, na Figura 4 (a), 70% das solicitações de identidades foram realizadas por fontes com confiança maior ou igual a $0,5$. Esse percentual decresce para 61% na Figura 4 (b), 56% na Figura 4 (c) e aproximadamente 50% na Figura 4 (d).

Em contrapartida, as fontes associadas aos atacantes continuam a apresentar um comportamento discrepante em relação às demais fontes. Apesar de os atacantes conseguirem algum sucesso no ataque em conluio, estes continuam a obter baixíssimos valores de confiança (logo, desafios computacionais mais complexos). Com 100 fontes, nenhuma solicitação obtém grau de confiança maior ou igual a $0,05$. Embora haja um ganho considerável no ataque para o caso em que 500 fontes são empregadas, apenas 13% das solicitações obtiveram confiança maior ou igual a $0,1$. Para o caso com 1000 fontes, 15% das solicitações obtiveram confiança maior ou igual a $0,2$, e para o caso com 2000 fontes, 35% das solicitações. Esses resultados evidenciam, ao mesmo tempo,



(a) 100 fontes maliciosas (b) 500 fontes maliciosas

(c) 1000 fontes maliciosas (d) 2000 fontes maliciosas

Figura 4. Resiliência da solução proposta à ataques *Sybil* partindo de várias fontes maliciosas, considerando uma mesma taxa de recorrência

a robustez e a eficácia da solução proposta frente a ataques *Sybil*, mesmo quando estes ocorrem em conluio. Mais importante, mostra que o atacante precisa dedicar uma gigantesca quantidade de recursos para obter sucesso no ataques, tanto em termos de *fontes* distribuídas (para despistar o esquema de diferenciação por fontes de solicitação), como em termos de capacidade computacional (para resolver os desafios propostos).

5. Discussões sobre a Solução Proposta

Conforme apresentado na Seção 3, nossa solução baseia-se na exigência da resolução de desafios computacionais antes que potenciais usuários obtenham identidades que os permitam ingressar na rede P2P. É importante frisar que a implantação da solução requer mínimas alterações nas *entidades* que compõem a rede. Considerando a sua instanciação em um arcabouço P2P tal como o BitTorrent [Cohen 2003], por exemplo, uma sequência de passos para a obtenção de identidades seria: (*i*) usuário solicitar uma identidade juntamente ao *tracker*; (*ii*) *tracker* requisitar a resolução de um desafio computacional ao usuário solicitante; (*iii*) usuário responder ao *tracker* o desafio computacional proposto; e (*iv*) *tracker* conferir a resolução e continuar o processamento segundo o protocolo BitTorrent original, caso a resolução esteja correta. Nesse caso, as modificações mais significativas para instanciar a solução ficariam restritas ao *tracker* – de modo que este passasse a manter informações sobre taxas de recorrência dos usuários, calcular seus respectivos valores para grau de confiança, e definir a complexidade do desafio computacional a ser resolvido em função do grau de confiança atual.

Sobre os parâmetros da solução proposta (a, b, c, β, t_w, e t_d), atualizações na valoração dos mesmos são necessárias para refletir mudanças de mais longo prazo nas características da rede (por exemplo, aumento do número de usuários ou mudanças no perfil da comunidade). Nesse caso, ajustes nos parâmetros de tempos em tempos (na ordem de meses, por exemplo) pode ser feito pelo administrador da rede, considerando sua própria experiência ou com o apoio de heurísticas e/ou mecanismos automatizados. Por outro lado, é importante mencionar que ajustes para adaptar a solução à mudanças de mais curto prazo não são necessários, uma vez que as métricas *taxa recorrência da fonte* e *taxa de recorrência da rede* são capazes de capturar variações no comportamento na rede que ocorrem nesse período de tempo.

Focando agora no mapeamento do grau de *confiança* em complexidade do desafios, tal depende essencialmente da natureza do mesmo. Por exemplo, considere o desafio apresentado em [Douceur 2002]: dado um número aleatório suficientemente grande y, encontrar dois números x e z em um período de tempo limitado tal que a concatenação $x|y|z$, após processada por uma função *hash* segura, leva a um número cujos n bits menos significantes são todos 0. Uma vez que o tempo para resolver esse desafio é proporcional a 2^{n-1}, e o tempo para verificar a resolução é constante, uma estratégia de mapeamento seria adotar uma função $f(x)$, que recebe como parâmetro o grau de confiança, e retorna um número inteiro n que define a complexidade do desafio.

Por fim, em relação a materalização da noção de fonte, uma estratégia é considerar um endereço IP, uma sub-rede ou um sistema autônomo como uma fonte distinta. Outra estratégia seria o uso de sistemas de coordenadas de rede, por exemplo o Vivaldi [Dabek et al. 2004], para distinguir solicitações vindas de determinadas regiões, cidades, estados, ou mesmo países.

6. Considerações Finais

O emprego de desafios computacionais é uma alternativa que tem se mostrado promissora para combater a ocorrência de ataques *Sybils* em redes P2P, e que tem recebido massiva atenção da comunidade de pesquisa. Entretanto, a falta de mecanismos que permitam lidar adequadamente com situações em que existe significativa disparidade de poder computacional entre usuários legítimos e atacantes impede o seu uso mais efetivo e disseminado. Para lidar com essa limitação, nesse artigo foi proposto o uso de desafios computacionais adaptativos como limitante à disseminação de *Sybils*.

Os experimentos realizados, embora não exaustivos, evidenciaram a capacidade da solução proposta em diminuir a capacidade dos atacantes de criarem identidades falsas de forma indiscriminada, ao mesmo tempo sendo favorável a usuários legítimos, os quais foram, em geral, penalizados minimamente. Ao calcular valores de confiança menores a fontes com taxas de solicitação de identidade mais altas, os usuários (maliciosos) atrelados a essas fontes tiveram de arcar com desafios computacionais de maior complexidade. Por outro lado, os usuários associados a fontes presumidamente legítimas (e que recorreram menos vezes para solicitar identidades), receberam desafios computacionais menos complexos (dados os maiores valores de grau de confiança que as fontes em questão possuiam perante a rede P2P).

Como trabalhos futuros, pretende-se (*i*) estender a solução proposta para capturar o comportamento das fontes face o atraso associado à resolução dos desafios, posto que hoje tal não é considerado por questões de simplicidade; (*ii*) investigar um mecanismo que apóie a valoração mais adequada dos parâmetros da solução proposta considerando comunidades com características distintas; e (*iii*) instanciar a solução proposta em um arcabouço P2P, por exemplo o BitTorrent.

Agradecimentos

Agradecimentos ao Prof. Nazareno Andrade (UFCG), pela concessão dos traços históricos do Bitsoup.org utilizados na avaliação experimental apresentada neste artigo.

Referências

Aberer, K., Datta, A., and Hauswirth, M. (2005). A decentralized public key infrastructure for customer-to customer e-commerce. In *International Journal of Business Process Integration and Management*, pages 26–33.

Barcellos, M. P. and Gaspary, L. P. (2006). Fundamentos, tecnologias e tendencias rumo a redes p2p seguras. *Jornadas de Atualizações em Informática*, pages 187–244.

Bitsoup.org (2009). Bitsoup.org – the number one site for your torrent appetite. http://bitsoup.org/.

Borisov, N. (2006). Computational puzzles as sybil defenses. In *6th IEEE International Conference on Peer-to-Peer Computing (P2P 2006)*, pages 171–176.

Castro, M., Drushel, P., Ganesh, A., Rowstron, A., and Wallach, D. S. (2002). Secure routing for structured peer-to-peer overlay networks. In *5th Usenix Symposium on Operating Systems Design and Implementation (OSDI 2002)*, pages 299–314.

Cohen, B. (2003). Incentives Build Robustness in BitTorrent. http://citeseer.ist.psu.edu/579364.html.

Dabek, F., Cox, R., Kaashoek, F., and Morris, R. (2004). Vivaldi: a decentralized network coordinate system. *SIGCOMM Comput. Commun. Rev.*, 34(4):15–26.

Danezis, G., Lesniewski-Laas, C., Kaashoek, F. M., and Anderson, R. (2005). Sybil-resistant dht routing. pages 305–318.

Douceur, J. R. (2002). The sybil attack. In *1st International Workshop on Peer-to-Peer Systems (IPTPS 2002)*, pages 251–260.

Feldman, M., Papadimitriou, C., Chuang, J., and Stoica, I. (2006). Free-riding and whitewashing in peer-to-peer systems. *IEEE Journal on Selected Areas in Communications*, 24(5):1010–1019.

Morselli, R., Bhattacharjee, B., Katz, J., and Marsh, M. A. (2006). Keychains: A decentralized public-key infrastructure.

Rowaihy, H., Enck, W., McDaniel, P., and La Porta, T. (2007). Limiting sybil attacks in structured p2p networks. In *26th IEEE International Conference on Computer Communications (INFOCOM 2007)*, pages 2596–2600, Anchorage, Alaska , USA.

Singh, A., Ngan, T.-W., Druschel, P., and Wallach, D. S. (2006). Eclipse attacks on overlay networks: Threats and defenses. In *25th Conference on Computer Communications (INFOCOM 2006)*, pages 1–12, Barcelona, Catalunya, Spain.

Yu, H., Gibbons, P. B., Kaminsky, M., and Xiao, F. (2008). Sybillimit: A near-optimal social network defense against sybil attacks. In *IEEE Symposium on Security and Privacy*, pages 3–17. IEEE Computer Society.

Yu, H., Kaminsky, M., Gibbons, P. B., and Flaxman, A. (2006). Sybilguard: defending against sybil attacks via social networks. In *SIGCOMM '06: Proceedings of the 2006 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 267–278, New York, NY, USA. ACM Press.