Análise de Séries Temporais Relacionadas a Vulnerabilidades de Software em Dispositivos Expostos à Internet

Carlos Eduardo de Schuller Banjar, Cainã Figueiredo Pereira, Daniel S. Menasché

[1]Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brasil

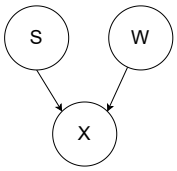
Abstract. *This article explores the relationship between the exposure of devices to the Internet and their susceptibility to software vulnerabilities exploitation. Highlighting the complexity of this phenomenon, we emphasize the importance of a rigorous approach to establish a clear relationship between exposure and exploitability. We then propose a perspective that considers three interconnected elements: the occurrence of real-world exploitation, the exposure of vulnerable devices, and the existence of* weapons *exploiting vulnerabilities. Despite the intuitive correlation between exposure and exploitability, we acknowledge the challenges in establishing a definitive relationship. To address these challenges, we gather public historical data from websites such as Shodan, EPSS, and inthewild.io, and present preliminary analyses on the relationship between such time series.*

Resumo. *Este artigo explora a relação entre a exposição de dispositivos à Internet e sua suscetibilidade à exploração de vulnerabilidades de software. Destacando a complexidade desse fenômeno, ressaltamos a importância de uma abordagem rigorosa para estabelecer uma relação clara entre exposição e explorabilidade. Propomos então uma visão que considera três elementos interligados: a ocorrência de exploração no mundo real, a exposição de dispositivos vulneráveis e a existência de armas que se aproveitem das vulnerabilidades. Apesar da correlação intuitiva entre exposição e explorabilidade, reconhecemos os desafios em estabelecer uma relação definitiva. Para contornar tais desafios, colhemos dados históricos públicos, de sítios públicos, como Shodan, EPSS e inthewild.io, e apresentamos análises preliminares sobre a relação entre tais séries temporais.*

1. Introdução

A relação entre a exposição de um tipo específico de dispositivo à Internet e sua suscetibilidade à exploração é um fenômeno complexo e multifacetado [Mazuera-Rozo et al. 2019, Gao et al. 2019]. É amplamente reconhecido que o nível de exposição à Internet desempenha um papel crucial na moldagem da explorabilidade de vulnerabilidades dentro desses dispositivos. No entanto, estabelecer uma relação clara e inequívoca entre exposição (*exposure*) e explorabilidade (*exploitability*) não é trivial.

A exposição à Internet implica em uma gama mais ampla de ameaças potenciais. Quando os dispositivos estão conectados à Internet, eles se tornam acessíveis a atores maliciosos que podem escanear vulnerabilidades e tentar explorá-las remotamente. Quanto mais exposto um dispositivo estiver, maior a probabilidade de ser alvo de ataques de indivíduos mal-intencionados que buscam aproveitar quaisquer fraquezas. Consequentemente, o risco de exploração bem-sucedida tende a aumentar à medida que o grau de exposição aumenta.



	Retrospectiva (histórico)	Predição (futuro)
Exploitation, X	inthewild.io, CISA KEV	EPSS
Weaponization, W	inthewild.io, NVD	ExpectedExploitability
Exposure, S	Shodan	N/A

Figura 1. *Exploitation in the wild*, denotado por X, depende de *exposure*, denotado por S e de *weaponization*, denotado por W, para ocorrer: para uma vulnerabilidade ser alvo de um ataque na Internet, é necessário que existam dispositivos expostos bem como um *weapon* para esta vulnerabilidade.

Embora essa correlação entre exposição e explorabilidade pareça intuitiva, quantificá-la é uma tarefa mais intrincada. Ao comparar diferentes tipos de dispositivos com níveis variados de exposição, podemos tentar isolar o impacto da exposição, controlando variáveis de confusão (*confounders*). Tais metodologias podem envolver a comparação de dispositivos idênticos em todos os aspectos, exceto pelo nível de exposição à Internet, imitando assim condições experimentais enquanto consideram as complexidades do mundo real.

No entanto, conduzir pesquisas quasi-experimentais no campo da cibersegurança é desafiador. Considerações éticas, o cenário de ameaças em constante evolução e a diversidade de dispositivos e vulnerabilidades representam obstáculos significativos. Além disso, é difícil eliminar completamente a influência de fatores confundidores, tornando difícil estabelecer uma relação definitiva. Também é necessário lidar com a natureza dinâmica das vulnerabilidades e explorações, que podem evoluir rapidamente com base nas táticas e motivações dos atacantes.

Neste trabalho, apresentamos medições colhidas de bases públicas da Internet sobre nossas três dimensões de interesse, a saber: explorabilidade (*exploitability*), exposição (*exposure*) e armamentização (*weaponization*). Em particular, apresentamos alguns gráficos mostrando vulnerabilidades para as quais a relação é mais clara, e outros nos quais mesmo que um dos componentes esteja presente, os demais não estão. Buscamos então possíveis hipóteses sobre os motivos de determinadas relações ocorrerem ou não.

O restante deste trabalho está organizado da seguinte forma. Na próxima seção apresentamos nossa visão. A Seção 3 apresenta a metodologia. Em seguida, a Seção 4 apresenta resultados preliminares, a Seção 5 trabalhos relacionados, e a Seção 6 conclui.

2. Nossa visão

Nossa visão é resumida na Figura 1. Podemos perceber que existem três elementos intimamente relacionados.

- *Exploitation in the wild*, X: ferramentas como Exploit Prediction Scoring System (EPSS) [Jacobs et al. 2021] tentam identificar se ocorrerá um uso, no mundo real, de uma determinada vulnerabilidade. Sobre o EPSS, é válido ressaltar que não temos acesso direto aos detalhes específicos sobre como ele determina a probabilidade de exploração de uma vulnerabilidade. Uma abordagem alternativa é usar dados de explorações passadas (*exploitations*), em vez de depender de um

modelo que prevê o futuro. Sítios como inthewild.io e bancos como CISA KEV prestam-se a esta finalidade de fornecer dados sobre explorações passadas.
- *Exposição*, S: ferramentas como Shodan [Matherly 2024, Zaidi et al. 2018] visam entender como ocorre a evolução do número de dispositivos expostos contendo uma determinada vulnerabilidade.
- *Weaponization*, W: inúmeros sítios na Internet, como inthewild.io[1], e o próprio NVD, contém informações sobre a existência de armas que se aproveitam de determinadas vulnerabilidades, para realização de ataques.

Embora os três elementos acima estejam intimamente relacionados, não é de nosso conhecimento nenhum trabalho anterior que tenha relacionado as séries temporais obtidas sobre as três dimensões acima [Wita et al. 2010, Martins et al. 2019]. Um de nossos propósitos é preencher esta lacuna.

É racional esperar alta chance de exploração em vulnerabilidades com alta exposição e para as quais existem armas (*weapons*) disponíveis. A primeira causa refere-se à atração dos hackers pelo potencial impacto gerado. Vulnerabilidades com alta exposição e *weapons* disponíveis tendem a atrair a atenção de hackers e cibercriminosos. Eles direcionam seus esforços para essas vulnerabilidades porque sabem que há um grande número de sistemas vulneráveis disponíveis para explorar, gerando alto impacto. A segunda causa refere-se à atração dos hackers pela facilidade de exploração. Alguns invasores são "work averse", o que significa que eles não irão desenvolver ou usar explorações novas e complexas se conseguirem alcançar seus objetivos com métodos já existentes [Allodi et al. 2022].

3. Metologia

Para compor nosso grupo de estudo, optamos por selecionar vulnerabilidades conhecidas que tenham sido amplamente expostas ao longo do tempo. Isso se deve ao fato de que muitas vulnerabilidades registradas nunca foram encontradas em dispositivos conectados à Internet, conforme indicado pelos dados do Shodan. Escolher vulnerabilidades com alta exposição ao longo do tempo é crucial para nosso objetivo de analisar a influência da exposição em outras variáveis.

Inicialmente, utilizamos a API de Tendências (Trends) do Shodan para identificar as CVEs mais frequentemente expostas em termos de quantidade de dispositivos afetados ao longo do tempo. A API de Tendências fornece uma análise mês a mês dos resultados históricos do Shodan. Após identificar as vulnerabilidades que já configuraram entre as mais expostas, fizemos requisições à API do Shodan para obter a série temporal de exposição de cada uma delas.

Posteriormente, coletamos informações sobre as armas cibernéticas (*weapons*)[2] e *exploitations* através da plataforma inthewild.io. Para *exploitation*, usamos também o EPSS, disponível na plataforma first.org.

Uma vez colhidos os dados, começamos a análise visualizando séries temporais. Em seguida, calculamos correlação de Pearson par a par entre três variáveis: *exposure* (S),

[1] `https://inthewild.io`

[2] *Weapons* também são conhecidas como *exploits*, mas neste artigo preferimos usar o termo *weapon* para se referir às armas, e *exploitation* para se referir ao uso destas armas em sistemas do mundo real.

weapons (W) e *exploitations* (X) para cada vulnerabilidade. A correlação de Pearson é comumente utilizada para determinar a relação linear entre duas séries temporais ao longo do tempo. Deixamos como trabalho futuro a análise de métricas que capturam relações não lineares, como informação mútua combinada ao Dynamic Time Warping (DTW).

Para construir as séries temporais de S, W e X, usadas para o cálculo do coeficiente de Pearson, consideramos valores instantâneos de S, e acumulados de W. Para X, consideramos duas alternativas: estimativa via dados preditivos (EPSS), com séries de valores instantâneos, ou dados históricos, sobre o número de *exploitations* de fato registrados até então (inthewild.io), com séries de valores cumulativos. Neste trabalho, para calcular a correlação de Pearson entre X e as demais variáveis, usamos a última alternativa, pois 1) evita lidarmos com as diferentes versões do EPSS, que evoluiu ao longo do tempo, e 2) o EPSS está disponível desde 2021, mas algumas de nossas séries começam antes desta data.

A correlação de Pearson entre duas variáveis não é definida quando uma das variáveis é uma constante. De um total de 336 vulnerabilidades consideradas, para 12 foi possível calcular a correlação par a par entre S, W e X. Escolhemos algumas delas para ilustrar estudos de caso na seção a seguir. Para as demais vulnerabilidades, ao menos uma das variáveis permaneceu constante ao longo do período analisado. A escolha da estratégia para produzir a série temporal de X, discutida no parágrafo acima, afeta diretamente o número de vulnerabilidades para as quais somos capazes de calcular a correlação par a par entre S, W e X. O uso do EPSS ao invés de inthewild.io favorece maior disponibilidade de dados, mas ainda assim escolhemos a última alternativa pelos motivos destacados no último parágrafo.

4. Resultados preliminares

4.1. A visualização das séries temporais apoiou a hipótese de que maior *exploitation* está relacionado com *weaponization* e *exposure*?

Observamos o padrão que esperávamos em algumas séries temporais. O padrão observado nas vulnerabilidades CVE-2023-44487, CVE-2019-0211 e CVE-2019-0211 confirmou nossas expectativas: elevada exposição e presença de armas cibernéticas associados com alto *exploitation*, seja este último medido de forma preditiva (EPSS) ou de forma histórica (*exploitations* publicados em inthewild.io). No entanto, apesar da consistência desse padrão em inúmeros cenários, encontramos casos, como o exemplificado pela CVE-2019-9639, nos quais, mesmo diante da exposição e presença de armas cibernéticas, o EPSS não alcançou níveis significativos, nem houve exploração.

Para título de ilustração, as vulnerabilidades CVE-2023-44487, CVE-2019-0211 e CVE-2021-40438 apresentaram altos valores de correlação de Pearson entre S, W e X, mas isso não ocorreu para CVE-2019-9639. Os casos de alta exposição combinada com a presença de armas cibernéticas, mas com baixo EPSS, ou sem nenhum *exploitation* registrado em inthewild.io, como CVE-2019-9639, podem ser atribuídos a uma série de fatores. Um deles pode ser a qualidade ou eficácia dos próprios *weapons* disponíveis. Mesmo que um sistema esteja altamente exposto e vulnerável, se os *weapons* disponíveis forem de baixa qualidade, alta complexidade ou baixa relevância para as vulnerabilidades presentes, a exploração dessas vulnerabilidades pode ser dificultada ou até mesmo inviável.

4.2. Qual a prevalência de alta explorabilidade e presença de *weapons*, para vulnerabilidades com alta exposição?

Com base nos dados obtidos do Shodan, identificamos 336 CVEs altamente expostas. Dentro desse conjunto, constatamos que 120 delas possuem *weapons* disponíveis.

Dentre as 336 CVEs consideradas, observamos que 54 alcançaram EPSS superior a 0.9 em algum momento, sendo que para 32 destas 54 CVEs fomos capazes de encontrar um *weapon* no sítio inthewild.io. Esses resultados nos levam à conclusão de que aproximadamente 60% das vezes em que o EPSS atingiu valores consideráveis, isso esteve associado à exposição e à presença de *weapons* pré-existentes, como mostra a Figura 2.

Figura 2. Resultado obtido analisando exposição, presença de *weapons* e EPSS.

4.3. Como se comportam as vulnerabilidades com baixa exposição?

Para explorar a relação entre a presença de *weapons* em vulnerabilidades com baixa exposição e seu impacto no EPSS, selecionamos aleatoriamente 336 vulnerabilidades que não estão na lista das "mais afetadas" pelo Shodan ($S \approx 0$) e que possuem armas cibernéticas ($W > 0$).

Os resultados revelam que, das 336 CVEs analisadas, apenas 11 (3%) delas apresentaram um EPSS superior a 0.9 em algum momento, enquanto a maioria, ou seja, 280 (83%) CVEs, tiveram um EPSS inferior a 0.1 em todos os pontos analisados.

Esses resultados indicam que, mesmo na presença de *weapons*, a maioria das vulnerabilidades com baixa exposição tende a ter um EPSS significativamente mais baixo. Isso sugere que a disponibilidade de *weapons* por si só pode não ser suficiente para impulsionar a exploração de vulnerabilidades em grande escala, especialmente se essas vulnerabilidades não estiverem amplamente expostas. Fatores como a visibilidade e a acessibilidade das vulnerabilidades podem desempenhar um papel crucial na determinação do EPSS, mesmo na presença de *weapons*. Isso destaca a importância de considerar não apenas a presença de *weapons*, mas também o contexto mais amplo em que as vulnerabilidades estão inseridas ao avaliar seu potencial de exploração.

4.4. Estudos de caso

As vulnerabilidades que serão analisadas serão CVE-2023-44487, CVE-2019-0211, CVE-2021-40438 e CVE-2019-9639. As três primeiras estão entre as 12 vulnerabilidades para as quais foi possível calcular a correlação par a par entre S, W e X. Para a quarta, não foi possível obter a correlação entre pares envolvendo *exploitation*, pois X manteve-se constante ao longo do período estudado (vide Seção 3). Os coeficientes de Pearson para essas vulnerabilidades estão disponíveis na Tabela 1.

No geral, a CVE-2023-44487 e a CVE-2021-40438 demonstraram fortes correlações positivas para todos os pares de variáveis. A vulnerabilidade CVE-2019-0211

Tabela 1. Coeficientes de Pearson par a par

CVE	X vs. S	W vs. S	X vs. W
CVE-2023-44487	0.865	0.9	0.994
CVE-2019-0211	0.153	0.452	0.908
CVE-2021-40438	0.831	0.840	0.960
CVE-2019-9639	N/A	-0.102	N/A

também apresentou correlações positivas, embora menos fortes do que as anteriores. A vulnerabilidade CVE-2019-9639 não teve valores de correlação calculáveis para alguns pares e, negativa, entre *Weaponization* e *Exploitation*.

Podemos comparar os resultados obtidos usando o coeficiente de Pearson com as tendências observadas nas séries temporais plotadas nos gráficos. Na Figura 3, podemos ver a evolução da métricas das quatro vulnerabilidades. A série temporal da CVE-2023-44487 (Figura 3(a)) indica a presença de 3 explorações (*exploitations*) e 7 *weapons*. Dos eventos, 8 ocorreram em outubro de 2023 (3 explorações e 5 weapons). A curva de exposição foi praticamente nula durante a maior parte do período, experimentando um aumento significativo em dezembro de 2023, quando a vulnerabilidade afetou mais de 7 milhões de dispositivos. O EPSS permaneceu em valores moderados. Além disso, todos os eventos (publicações de armas e explorações) ocorreram em um período próximo ao aumento da exposição. Portanto, tanto o gráfico quanto os coeficientes sugerem uma correlação entre as variáveis.

A série temporal da vulnerabilidade CVE-2021-40438 (Figura 3(b)) apresenta um intervalo de exposição maior do que a CVE-2023-44487 e os eventos parecem ocorrer dentro dessa janela. Houve quatro armas e uma exploração. Ao analisar os números (Tabela 1), as correlações positivas entre pares de variáveis existem, mas são mais fracas do que as da CVE-2023-44487. O EPSS manteve-se consistentemente alto.

A Figura 3(c) da CVE-2019-0211 mostra a presença de 4 armas e 1 exploração, juntamente com valores altos para o EPSS. Novamente, a imagem confirma os números citados e a ideia de que os eventos ocorrem durante o período de exposição. Além disso, o último evento registrado foi uma exploração, o que pode indicar que ocorreu influenciado pela presença de armas públicas já existentes e pela alta exposição no momento.

Na Figura 3(d) observamos um padrão diferente das CVEs anteriores. Esta CVE apresenta uma alta exposição, afetando até 7 milhões de dispositivos, e possui *weapons* públicos. No entanto, mesmo diante desse cenário, o EPSS permanece baixo e não há registro de *exploitations*. Uma possível explicação para tal comportamento é apresentada na Seção 4.1.

Resumo. Nossos resultados preliminares indicam haver, de fato, uma forte correlação entre a exposição, a presença de *weapons* e o *exploitation* (este último medido por dados históricos ou pelo EPSS). A constatação de que cerca de 60% das instâncias em que o EPSS alcançou valores elevados estavam associadas à presença simultânea de exposição e *weapons* sugere que a existência de *weapons* para uma vulnerabilidade pode aumentar sua exploração, especialmente quando combinada com uma ampla exposição dos dispositivos afetados.

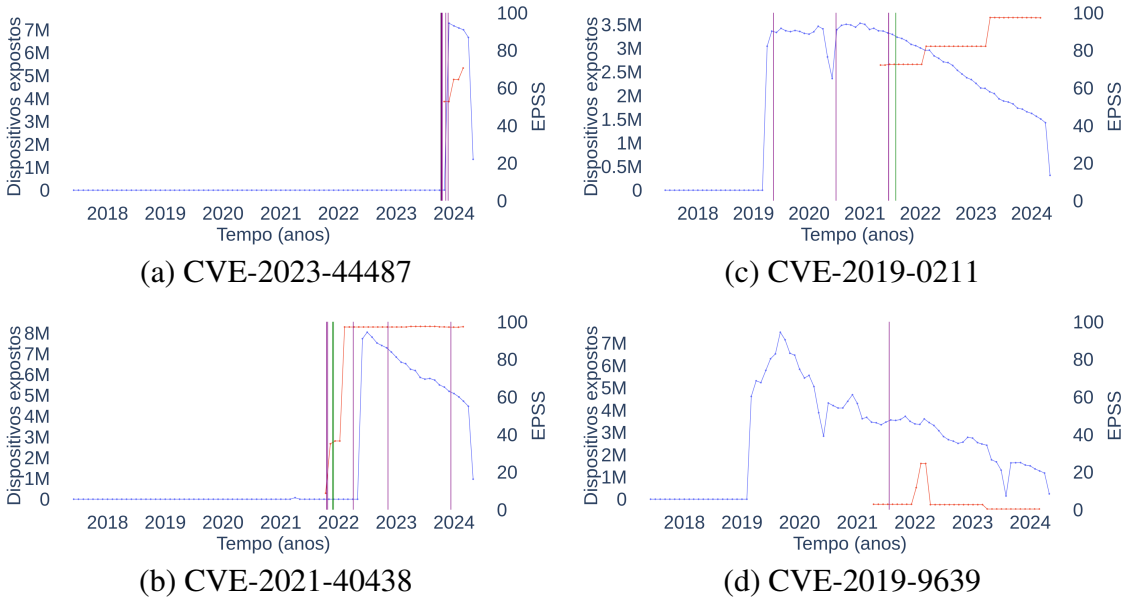


Figura 3. Exemplos de evolução de métricas associadas às vulnerabilidades: na figura mostramos curvas com chance de exploração *in the wild* (EPSS), em laranja, exposição no Shodan, em azul, e divulgação de armas (*weapons*) e *exploitations*, estas últimas como linhas verticais roxas e verdes.

5. Trabalhos relacionados

Existe uma ampla gama de trabalhos sobre o ciclo de vulnerabilidades, e sua relação com exposição [Pastrana et al. 2018, Bada and Pete 2020, Ponce et al. 2022]. Entretanto, não é de nosso conhecimento nenhum trabalho anterior que tenha relacionado Shodan, EPSS e fontes de dados sobre armas usadas para explorar vulnerabilidades.

Acreditamos que existem pelo menos duas abordagens viáveis para a integração das informações fornecidas pelo Shodan com os dados do EPSS. A primeira consiste em utilizar o Shodan como um modificador do EPSS, onde as informações obtidas através do Shodan são usadas para aprimorar ou ajustar as funcionalidades do EPSS. Isso pode envolver a utilização dos dados do Shodan para melhorar a precisão das análises realizadas pelo EPSS, identificar vulnerabilidades em tempo real ou aprimorar as capacidades de detecção de ameaças.

Por outro lado, a segunda abordagem envolve a integração direta do Shodan ao EPSS, permitindo que as informações coletadas pelo Shodan sejam integradas de forma nativa ao sistema. Neste cenário, as funcionalidades do Shodan seriam incorporadas ao EPSS, possibilitando o acesso às informações do Shodan dentro do próprio ambiente do EPSS. Isso pode facilitar a visualização e o gerenciamento de dados, oferecendo uma experiência mais integrada para os usuários do sistema.

Ambas as abordagens apresentam vantagens e desafios específicos, e a escolha entre elas dependerá das necessidades e requisitos específicos do ambiente de segurança cibernética em questão. No entanto, independentemente da abordagem escolhida, a integração entre o Shodan e o EPSS tem o potencial de fortalecer significativamente as capacidades de segurança cibernética, fornecendo insights mais abrangentes e permitindo uma resposta mais eficaz às ameaças em tempo real. Além disso, o fato de termos nos

concentrado exclusivamente em dados abertos pode, potencialmente, impulsionar o desenvolvimento de modelos transparentes, diferentemente de modelos que dependem de dados privados, como o EPSS.

6. Conclusão e trabalhos futuros

A relação entre a exposição à Internet e a explorabilidade de vulnerabilidades é uma consideração importante na cibersegurança. Embora a conexão intuitiva sugira que uma exposição mais alta aumenta o risco de exploração, estabelecer uma relação clara é não trivial. À medida que o cenário tecnológico continua a evoluir, acumular evidências empíricas pode aprimorar nossa compreensão de como a exposição impacta a explorabilidade, levando a estratégias e práticas de segurança mais informadas.

As análises realizadas no artigo não são suficientes para generalizar um comportamento para todas as CVEs. Dado que as CVEs possuem diversas características que influenciam sua explorabilidade, será necessário condicionar o estudo a algumas dessas características para estimar o comportamento de cada uma com maior precisão.

Referências

Allodi, L., Massacci, F., and Williams, J. (2022). The work-averse cyberattacker model: Theory and evidence from two million attack signatures. *Risk Analysis*.

Bada, M. and Pete, I. (2020). An exploration of the cybercrime ecosystem around Shodan. In *Int. conference on internet of things: Systems, management and security*, pages 1–8.

Gao, J., Li, L., Kong, P., Bissyandé, T. F., and Klein, J. (2019). Understanding the evolution of Android app vulnerabilities. *IEEE Transactions on Reliability*, 70(1):212–230.

Jacobs, J., Romanosky, S., Edwards, B., Adjerid, I., and Roytman, M. (2021). Exploit prediction scoring system (EPSS). *Digital Threats: Research and Practice*, 2(3):1–17.

Martins, M., Bicudo, M. A., Menasché, D., and de Aguiar, L. P. (2019). Análise temporal de risco de sistemas computacionais via modelagem de séries de eventos associados a vulnerabilidades. In *WPerformance*. SBC.

Matherly, J. (2024). Shodan. `https://www.shodan.io/`. Accessed: April 2, 2024.

Mazuera-Rozo, A., Bautista-Mora, J., Linares-Vásquez, M., Rueda, S., and Bavota, G. (2019). The Android OS stack and its vulnerabilities: an empirical study. *Empirical Software Engineering*, 24:2056–2101.

Pastrana, S., Hutchings, A., Caines, A., and Buttery, P. (2018). Characterizing eve: Analysing cybercrime actors in a large underground forum. In *RAID*, pages 207–227.

Ponce, L. M. S., Gimpel, M., Fazzion, E., Cunha, Í., Hoepers, C., Steding-Jessen, K., Chaves, M. H., Guedes, D., and Meira Jr, W. (2022). Caracterização escalável de vulnerabilidades de segurança: um estudo de caso na internet brasileira. *SBRC*.

Wita, R., Jiamnapanon, N., and Teng-Amnuay, Y. (2010). An ontology for vulnerability lifecycle. In *2010 Third International Symposium on Intelligent Information Technology and Security Informatics*, pages 553–557. IEEE.

Zaidi, N., Kaushik, H., Bablani, D., Bansal, R., and Kumar, P. (2018). A study of exposure of iot devices in India: Using Shodan search engine. In *Information Systems Design and Intelligent Applications*, pages 1044–1053. Springer.