RESSALVA

Atendendo solicitação do
autor, o texto completo
desta dissertação será
disponibilizado somente
a partir de 20/02/2021.

UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de São José do Rio Preto

Leandro Bertini Lara Gonçalves

Abordagem para Geração Automática de Assinatura de Ataques baseada em Fluxos de Redes de Computadores

São José do Rio Preto
2019

Leandro Bertini Lara Gonçalves

Abordagem para Geração Automática de Assinatura de Ataques baseada em Fluxos de Redes de Computadores

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação, junto ao Programa de Pós-Graduação em Ciência da Computação, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista "Júlio de Mesquita Filho", Câmpus de São José do Rio Preto.

Financiadora: CAPES

Orientador:

Prof. Dr. Adriano Mauro Cansian

São José do Rio Preto
2019

G635a

Gonçalves, Leandro Bertini Lara
Abordagem para Geração Automática de Assinatura de Ataques baseada em Fluxos de Redes de Computadores / Leandro Bertini Lara Gonçalves. -- São José do Rio Preto, 2019
74 f. : il., tabs.

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto
Orientador: Adriano Mauro Cansian

1. Ciência da computação. 2. Computadores Medidas de segurança. 3. Crime por computador. 4. Reconhecimento de padrões. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Leandro Bertini Lara Gonçalves

Abordagem para Geração Automática de Assinatura de Ataques baseada em Fluxos de Redes de Computadores

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Ciência da Computação, junto ao Programa de Pós-Graduação em Ciência da Computação, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista "Júlio de Mesquita Filho", Câmpus de São José do Rio Preto.

Financiadora: CAPES

Comissão Examinadora

Prof. Dr. Adriano Mauro Cansian
Unesp - Câmpus São José do Rio Preto
Orientador

Prof. Dr. Geraldo Francisco Donegá Zafalon
Unesp - Câmpus São José do Rio Preto

Prof. Dr. Robson de Oliveira Albuquerque
UnB - Universidade de Brasília

São José do Rio Preto
20 de Fevereiro de 2019

Agradecimentos

Agradeço, primeiramente, ao Laboratório ACME! pois sem ele não haveria a possibilidade de realizar o estudo nesta monografia descrito. Além disso, agradeço aos integrantes desse laboratório, em especial aos membros Raphael Campos Silva, Rafael Stefanini Carreira, Vinícius Oliveira Ferreira, Vinícius Vassoler Galhardi, Amanda Barbosa Sobrinho, Bruno Ferreira Leal e Pedro Ferracini de Barros por toda a ajuda e dúvidas tiradas durante as pesquisas e testes e pelos laços de amizades ali construídos. Assim como agradeço ao meu orientador, Prof. Dr. Adriano Mauro Cansian, pela paciência no processo e por dividir comigo seus conhecimentos e experiências.

Agradeço à minha namorada por todo o apoio incondicional dedicado, aos meus amigos e familiares pelo apoio constante dedicado à mim e às minhas ideias.

Em especial, agradeço meu amigo Matheus Gonçalves Ribeiro pela amizade, companhia e apoio, dividindo muitas experiências e conversas desde os primeiros dias da graduação. Agradeço aos meus amigos Guilherme Freire Roberto e Matheus Carreira Andrade por todo o apoio e suporte a mim sempre prontamente oferecidos, pelas ótimas conversas e pelo companheirismo.

Agradeço à Sra. Olga Maria Rissi Ferreira por todo o incentivo e pelos sempre excelentes conselhos que me motivaram durante o curso. Agradeço à minha amiga Adriana Felix Roberto Ártico pelo constante apoio, incentivo e conselhos desde o início da minha graduação.

Agradeço, também, à Universidade Estadual Paulista "Júlio de Mesquita Filho", campus de São José do Rio Preto, pelos cursos de graduação e de pós-graduação ali ministrados por professores de excelência.

O presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, à qual agradeço.

"Dietro ogni problema c’è un’opportunità."

-Galilei, Galileo

Resumo

Este trabalho apresenta um método para a geração automática de assinaturas de ataques em redes de computadores a partir de fluxos de redes. Nessa abordagem, utiliza-se um modelo de tráfego legítimo para a comparação dos dados e identificação dos elementos significativos. Esse modelo é composto de duas partes: uma formada por *clusters* e funções de distribuição acumuladas e outra formada por tráfegos pré-processados. A partir desse modelo composto, dispensa-se a necessidade do fornecimento de dados de múltiplas ocorrências do mesmo evento para a análise e extração de características. O sistema foi testado, quanto a sua capacidade de geração de assinaturas, para seis diferentes ataques e para tráfegos legítimos. O melhor resultado de assinatura gerado obteve F1-*Score* de 0.9801, o que é considerado bom pela literatura. Os tempos de geração de assinaturas para tráfegos maliciosos foram, em sua maioria, considerados rápidos, permanecendo abaixo de um minuto.

Palavras-chave: fluxos de rede, geração de assinaturas, segurança de redes.

Abstract

In this work we present a method for automatic signature generation of network malicious activity based in its network flow data. In this approach, we proposed a model of legitimate data for data comparison and significant features identification. Such model is composed of two parts: one formed by clusters and cumulative distribution functions, and another formed by preprocessed data. With this model, there is no need to provide data of multiple occurrences of the same event so the characteristics can be extracted. The system was tested for six different attacks and for legitimate traffic, so we could obtain its signature generation capacity. The best signature result has its F1-Score equal to 0.9801, which is good according to the literature. The signature generation time for malicious traffic was, in its majority, considered fast enough, remaining below one minute.

Keywords: *network flows, network security, signature generation.*

Lista de Figuras

Lista de Tabelas

Lista de Abreviações

APT *Advanced Persistent Threats*

BIC *Bayesian Information Criterion* (Critério de Informação Bayesiano)

CERT.Br Centro de Estudos, Resposta e Tratamento de Incidentes de Segurança no Brasil

DDoS *Distributed Denial of Service* (Negação de Serviço Distribuída)

DoS *Denial of Service* (Negação de Serviço)

E Especificidade

FDA Função de Distribuição de Probabilidade

FDP Função de Densidade de Probabilidade

FN Falso Negativo

FP Falso Postivo

HTTP *HyperText Transfer Protocol*

IDS *Intrusion Detection System* (Sistemas de Detecção de Intrusão)

IETF *Internet Engineering Task Force*

IP *Internet Protocol*

JSON *Javascript Object Notation*

NAT *Network Address Translator*

NIDS *Network Intrusion Detection System* (Sistema de Detecção de Intrução em Redes)

S Sensibilidade

SGBD Sistema Gerenciador de Banco de Dados

TFN Taxa de Falso Negativo

TFP Taxa de Falso Positivo

TVN Taxa de Verdadeiro Negativo

TVP Taxa de Verdadeiro Positivo

VN Verdadeiro Negativo

VP Verdadeiro Positivo

XSS *Cross Site Scripting*

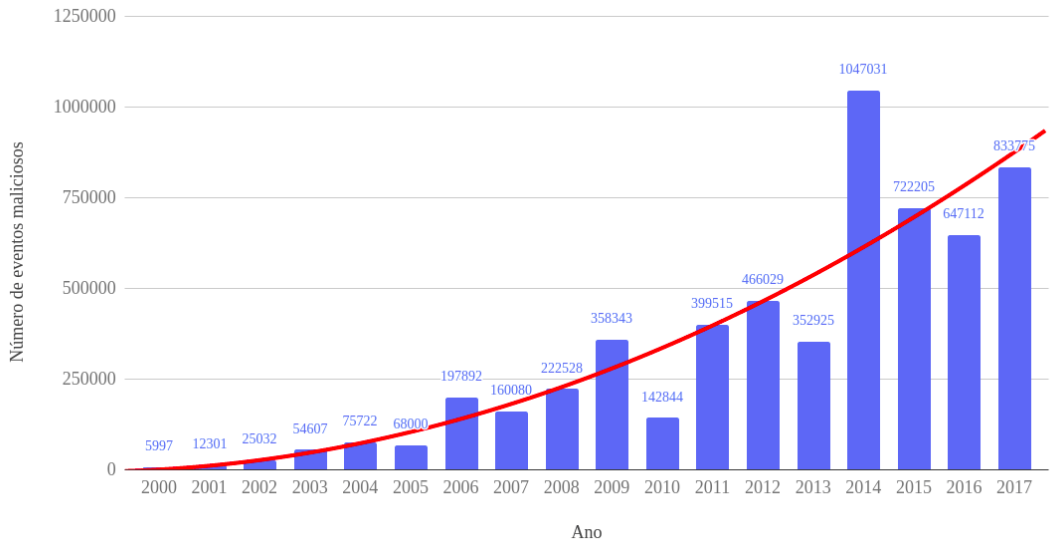
Sumário

Capítulo 1

Introdução

Eventos de segurança da informação têm se mantido como uma tendência de alta na última década, como pode ser observado no gráfico de incidentes reportados para o CERT.Br (2019), na Figura 1.1. O aumento desses incidentes de segurança requer maior quantidade de mão de obra para seu tratamento (GARCIA-TEODORO et al., 2015), o que implica em uma elevação de custos operacionais e exige uma significativa quantidade de tempo para se lidar com carga de trabalho (AFEK; BREMLER-BARR; FEIBISH, 2013).

Figura 1.1: Números de ataques reportados para o CERT.Br por ano.



Fonte: Adaptado de CERT.Br (2019).

Assim, a automatização de atividades de segurança se apresenta como uma solução atrativa para enfrentar a elevação do número de ocorrência de atividades maliciosas.

Dentre as funções passíveis de automatização pode-se incluir a geração de assinaturas para eventos maliciosos em redes de computadores.

Assinaturas para incidentes de segurança são padrões que podem ser utilizados para a identificação de eventos maliciosos. Esses padrões podem ser obtidos pela análise de dados de dois tipos: maliciosos, provenientes de ataques que possibilitam a geração de assinaturas de ataque, e dados provenientes de abusos de redes, que possibilitam a geração de assinaturas de abuso (CORRêA, 2009). Este trabalho concentra-se na geração de assinaturas desse último tipo.

A análise de dados para a identificação dos padrões para a formação de uma assinatura é um processo que demanda tempo e requer indivíduos altamente especializados. Por via de regra, pode-se dizer que a qualidade de uma assinatura pode variar de acordo com a proficiência do profissional que a gerou (SHIM et al., 2017). Também, a tendência crescente de ataques, como observado em CERT.Br (2019), e o aumento da diversidade de ataques dificultam a aplicação de políticas e controle de redes, uma vez que indivíduos maliciosos constantemente buscam passar pelas medidas de proteção implementadas em seus sistemas ou serviços alvos. Nesse cenário, a redução do tempo de resposta ou a automatização do processo de identificação de aspectos relevantes em atividades maliciosas podem trazer um diferencial para a aplicação de contramedidas ante os diversos eventos maliciosos.

1.1 Motivação e Justificativas

No tratamento de atividades maliciosas, variações de ataques a redes de computadores impõem um significativo problema de segurança e exigem a solução ativa de profissionais especializados, o que acarreta custos e pode restringir a capacidade de aplicação de práticas de segurança de qualidade.

Neste tocante, a geração automática de assinaturas torna uma parte do processo de aplicação de contramedidas mais simples para os profissionais da área. Aumentando a eficiência do combate a eventos maliciosos no geral. Para o tratamento de *malwares*, diversos estudos sobre geração automática de assinaturas foram desenvolvidos, resultando em métodos capazes de identificar e classificar comportamentos nocivos realizados por esses *softwares* maliciosos, como Kreibich e Crowcroft (2004), David e Netanyahu (2015) e Szynkiewicz e Kozakiewicz (2017). Em contrapartida, para ataques a redes de computadores, comparativamente, poucos estudos foram desenvolvidos sobre a geração e classificação automática de eventos maliciosos. Com isso, este trabalho tem como intuito preencher parte deste espaço, buscando contribuir para uma

melhor aplicação de segurança em redes de computadores.

Diversos trabalhos observados na literatura têm o propósito de identificar padrões em dados de ataques previamente conhecidos, dentre esses Afek, Bremler-Barr e Feibish (2013), Fallahi, Sami e Tajbakhsh (2016). Esses trabalhos encontram invariâncias entre os tráfegos de ataques e as utiliza como assinaturas. Contudo, ataques a redes de computadores podem ter trechos que não são representados por invariâncias, mas sim por intervalos. O uso desses intervalos aumenta a representação da assinatura, mas requer outros métodos de identificação.

Neste trabalho, estudou-se a possibilidade de se aplicar um modelo de tráfegos legítimos para a seleção de atributos característicos de tráfegos maliciosos. Este modelo é composto de duas partes: um modelo de *clusters* e Funções de Distribuição Acumuladas e uma base de dados de comparação.

Uma vez que métodos de clusterização foram bem sucedidamente utilizados para o agrupamento de fluxos de rede nos trabalhos Ferreira (2016), Galhardi (2017), Ravale, Marathe e Padiya (2015), espera-se que sejam capazes de formar um modelo confiável para a extração de padrões e posteriormente probabilidades de ocorrência para ponderar a relevância dos atributos analisados.

Assim, como tráfegos maliciosos tendem a apresentar características que os distinguem dos demais tráfegos idôneos em um dado momento, é possível a identificação e o isolamento dessas características para a identificação de demais tráfegos maliciosos semelhantes. Isso pode ser feito em duas etapas: a primeira consiste em identificar atributos invariantes no tráfego malicioso, a segunda é feita com a ordenação por discrepância dos atributos ante a tráfegos legítimo e sua posterior seleção por filtragem de dados.

1.2 Objetivos

Este trabalho teve como objetivo o desenvolvimento e a análise de uma abordagem para a geração de assinaturas de atividades maliciosas pela análise dos fluxos de tráfegos de rede considerados abusivos. Como requisito, essa geração deve ser feita em menos tempo do que manualmente e com F1-*Score* representando uma boa qualidade de assinaturas de acordo com a literatura, ou seja, acima de 0.9.

Como objetivos específicos desenvolvidos durante a execução deste trabalho, pode-se considerar: *a*) estudo sobre o processo de extração de assinaturas a partir de fluxos maliciosos; e *b*) o desenvolvimento de um método rápido para a análise e seleção de dados maliciosos em fluxos de redes.

Capítulo 5

Conclusões

Tráfegos presentes em redes de computadores podem ser sumarizados em fluxos de rede, esses fluxos contêm diversas métricas que podem ser utilizadas para, entre outras, a aplicação de segurança em redes de computadores. Tráfegos maliciosos podem definir valores característicos nos fluxos de redes, possibilitando, portanto, sua identificação.

Neste trabalho foi desenvolvida uma abordagem para a identificação automática desses atributos maliciosos e, com isso, a formação de uma assinatura possibilitando a identificação de outras ocorrências dessa atividade na rede. Nesse escopo, utilizou-se a análise de invariâncias e comparação com dados legítimos para a identificação de atributos maliciosos relevantes.

O alto desempenho no processamento de dados maliciosos era um requisito do objetivo do projeto e este foi alcançado com o tempo máximo de geração de saída de dados na ordem de minutos e com 70% dos ataques com tempo de produção de até 41 segundos.

Observou-se, contudo, que a qualidade das assinaturas geradas pelo sistema em sua maioria é baixa, com F1-*Score* menor do que 0.8. Os principais fatores que levam a este problema são a seleção de atributos com alto grau de distinção, mas com baixa significância - como, por exemplo, a escolha do número de portas de origem da camada de transporte ao invés do número de fluxos em uma assinatura de varredura - o que reduz a caracterização do ataque e a seleção de conjuntos de atributos que pode atender o critério de não identificação de fluxos legítimos, mas também não é capaz de descrever o ataque.

Ainda que em sua maioria as assinaturas não tenham atingido altos níveis de qualidade, o sistema foi capaz de gerar algumas assinaturas com F1-*Score* acima de 0.9. Comparando com o trabalho Fallahi, Sami e Tajbakhsh (2016), o ataque DoS conse-

guiu desempenho superior na geração de assinaturas e o ataque de força bruta SSH desempenho praticamente equivalente. Essas assinaturas têm como diferencial a seleção dos atributos que permitiram a significância para o ataque analisado. Esses resultados embasam a hipótese de que os problemas de seleção de características maliciosas significativas comentados anteriormente possam ser tratados com a modificação do critério de seleção de atributos quantitativos, seja, por exemplo, pela alteração da ordenação dos atributos submetidos ao teste de seleção, ou pela forma de unificação de atributos quantitativos. Isso pode ser feito com a aplicação de técnicas de aprendizado de máquina ou métodos estatísticos e é considerado como um dos trabalhos futuros a este.

Uma característica intrínseca à geração de assinaturas baseadas em uma amostra de eventos maliciosos é a excessiva especialização da assinatura gerada. São esperados resultados com alto grau de especificidade para a variação do ataque analisada. Isso requer métodos de generalização tanto para atributos quanto para valores de atributos. Tais métodos estão sendo desenvolvidos em paralelo a este trabalho no Laboratório ACME!.

A presença de tráfegos legítimos não foi totalmente ignorada pelo sistema, mesmo com a comparação com esse mesmo tipo de tráfego. Esse resultado tem consequências negativas para a aplicação do sistema em produção, uma vez que é esperado esse tipo de tráfego em meio ao fluxo malicioso. Uma proposta para se contornar esse problema é a segmentação do tráfego por IP de origem e destino, fragmentando os dados antes de seu processamento pelo sistema e com isso, possibilitando o processamento do tráfego malicioso com menos ou nenhuma interferência.

A análise e a aplicação de diferentes parâmetros para o K-Means, assim como outros algoritmos de clusterização, podem ser exploradas em trabalhos futuros para a melhoria da verosimilhança dos valores de probabilidade gerados, o que pode ter impacto positivo na seleção de atributos e melhorar a qualidade média das assinaturas produzidas pelo sistema. A composição dos métodos de associação de probabilidade e seleção de atributos significativos foram os elementos de maior complexidade de desenvolvimento na elaboração deste trabalho. Apresentam, além disso, um vasto conjunto de possibilidades de desenvolvimentos alternativos com suas próprias vantagens e desvantagens cujo comportamento para a aplicação deste trabalho ainda não é empiricamente conhecida.

Assim, este trabalho atinge os seus objetivos com sucesso parcial, sendo capaz de gerar automaticamente e em décimos de segundos assinaturas satisfatórias para os ataques de varredura de reconhecimento, DoS e força bruta SSH. Além de abrir

margens para desenvolvimentos futuros quanto à melhoria da qualidade das assinaturas geradas e quanto ao aumento da capacidade de tratamento de tipos de ataques.

Referências Bibliográficas

ABDOU, A.; BARRERA, D.; OORSCHOT, P. C. V. What lies beneath? analyzing automated ssh bruteforce attacks. In: SPRINGER. *International Conference on Passwords*. [S.l.], 2015. p. 72–91.

AFEK, Y.; BREMLER-BARR, A.; FEIBISH, S. L. Automated signature extraction for high volume attacks. In: IEEE PRESS. *Proceedings of the ninth ACM/IEEE symposium on Architectures for networking and communications systems*. [S.l.], 2013. p. 147–156.

AHMAD, M. A.; WOODHEAD, S.; GAN, D. Early containment of fast network worm malware. In: IEEE. *Information and Computer Science (NICS), 2016 3rd National Foundation for Science and Technology Development Conference on*. [S.l.], 2016. p. 195–201.

ALTWAIJRY, H.; SHAHBAR, K. (whasg) automatic snort signatures generation by using honeypot. *JCP*, v. 8, p. 3280–3286, 2013.

ALWAN, Z. S.; YOUNIS, M. F. Detection and prevention of sql injection attack: A survey. *International Journal of Computer Science and Mobile Computing*, v. 6, n. 8, p. 5–17, 2017.

BARABAS, M.; DROZD, M.; HANACEK, P. Behavioral signature generation using shadow honeypot. *World Academy of Science, Engineering and Technology, ser*, n. 65, p. 829–833, 2012.

BRAY, T. *The javascript object notation (json) data interchange format*. [S.l.], 2017.

CERT.BR. *Estatísticas dos Incidentes Reportados ao CERT.br*. 2019. <https://www.cert.br/stats/incidentes/>. Accessed on January 15, 2019.

CHANDOLA, V.; BANERJEE, A.; KUMAR, V. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, ACM, v. 41, n. 3, p. 15, 2009.

CORRêA, J. L. *Um modelo de detecção de eventos em redes baseado no rastreamento de fluxos*. Universidade Estadual Paulista, 2009. Disponível em: <http://www.dominiopublico.gov.br/pesquisa/DetalheObraForm.do?select_action=&co_obra=153259>.

DAVID, O. E.; NETANYAHU, N. S. Deepsign: Deep learning for automatic malware signature generation and classification. In: *2015 International Joint Conference on Neural Networks (IJCNN)*. [S.l.: s.n.], 2015. p. 1–8. ISSN 2161-4407.

DAVIS, J.; GOADRICH, M. The relationship between precision-recall and roc curves. In: ACM. *Proceedings of the 23rd international conference on Machine learning*. [S.l.], 2006. p. 233–240.

DERI, L.; SPA, N. nprobe: an open source netflow probe for gigabit networks. In: *TERENA Networking Conference*. [S.l.: s.n.], 2003.

DESAI, S. P.; HADULE, P. R.; DUDHGAONKAR, A. A. Denial of service attack defense techniques. 2017.

FALLAHI, N.; SAMI, A.; TAJBAKHSH, M. Automated flow-based rule generation for network intrusion detection systems. In: *2016 24th Iranian Conference on Electrical Engineering (ICEE)*. [S.l.: s.n.], 2016. p. 1948–1953.

FERREIRA, V. O. Classificaçao de anomalias e reduçao de falsos positivos em sistemas de detecçao de intrusao baseados em rede utilizando métodos de agrupamento. Universidade Estadual Paulista (UNESP), 2016.

GALHARDI, V. V. Detecção adaptativa de anomalias em redes de computadores utilizando técnicas não supervisionadas. Universidade Estadual Paulista (UNESP), 2017.

GARCIA-TEODORO, P.; DIAZ-VERDEJO, J. E.; TAPIADOR, J. E.; SALAZAR-HERNÁNDEZ, R. Automatic generation of http intrusion signatures by selective identification of anomalies. *Computers & Security*, Elsevier, v. 55, p. 159–174, 2015.

GÓMEZ, J.; GIL, C.; BAÑOS, R.; MÁRQUEZ, A. L.; MONTOYA, F. G.; MONTOYA, M. A pareto-based multi-objective evolutionary algorithm for automatic rule generation in network intrusion detection systems. *Soft Computing*, Springer, v. 17, n. 2, p. 255–263, 2013.

GU, G.; PERDISCI, R.; ZHANG, J.; LEE, W. Botminer: Clustering analysis of network traffic for protocol-and structure-independent botnet detection. 2008.

GUPTA, S.; ARBELAEZ, P.; MALIK, J. Perceptual organization and recognition of indoor scenes from rgb-d images. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2013. p. 564–571.

HE, D.; CHEN, X.; ZOU, D.; PEI, L.; JIANG, L. An improved kernel clustering algorithm used in computer network intrusion detection. In: IEEE. *Circuits and Systems (ISCAS), 2018 IEEE International Symposium on*. [S.l.], 2018. p. 1–5.

HEO, H.; SHIN, S. Who is knocking on the telnet port: A large-scale empirical study of network scanning. In: ACM. *Proceedings of the 2018 on Asia Conference on Computer and Communications Security*. [S.l.], 2018. p. 625–636.

HOFSTEDE, R.; ČELEDA, P.; TRAMMELL, B.; DRAGO, I.; SADRE, R.; SPEROTTO, A.; PRAS, A. Flow monitoring explained: From packet capture to data analysis with netflow and ipfix. *IEEE Communications Surveys & Tutorials*, IEEE, v. 16, n. 4, p. 2037–2064, 2014.

HOQUE, N.; BHUYAN, M. H.; BAISHYA, R. C.; BHATTACHARYYA, D. K.; KALITA, J. K. Network attacks: Taxonomy, tools and systems. *Journal of Network and Computer Applications*, Elsevier, v. 40, p. 307–324, 2014.

INAYAT, Z.; GANI, A.; ANUAR, N. B.; KHAN, M. K.; ANWAR, S. Intrusion response systems: Foundations, design, and challenges. *Journal of Network and Computer Applications*, Elsevier, v. 62, p. 53–74, 2016.

IOFFE, S. Improved consistent sampling, weighted minhash and l1 sketching. In: IEEE. *Data Mining (ICDM), 2010 IEEE 10th International Conference on*. [S.l.], 2010. p. 246–255.

JAIGANESH, V.; MANGAYARKARASI, S.; SUMATHI, D. P. Intrusion detection systems: A survey and analysis of classification techniques. *International Journal of Advanced Research in Computer and Communication Engineering*, v. 2, n. 3, p. 1629–1635, 2013.

JAIN, A. K. Data clustering: 50 years beyond k-means. *Pattern recognition letters*, Elsevier, v. 31, n. 8, p. 651–666, 2010.

JOHARI, R.; SHARMA, P. A survey on web application vulnerabilities (sqlia, xss) exploitation and security engine for sql injection. In: IEEE. *2012 International Conference on Communication Systems and Network Technologies*. [S.l.], 2012. p. 453–458.

KINDY, D. A.; PATHAN, A. K. A survey on sql injection: Vulnerabilities, attacks, and prevention techniques. In: *2011 IEEE 15th International Symposium on Consumer Electronics (ISCE)*. [S.l.: s.n.], 2011. p. 468–471. ISSN 2159-1423.

KREIBICH, C.; CROWCROFT, J. Honeycomb - creating intrusion detection signatures using honeypots. *Computer Communication Review*, v. 34, p. 51–56, 01 2004.

LEYDESDORFF, L. On the normalization and visualization of author co-citation data: Salton's cosine versus the jaccard index. *Journal of the American Society for Information Science and Technology*, Wiley Online Library, v. 59, n. 1, p. 77–85, 2008.

LI, B.; SPRINGER, J.; BEBIS, G.; GUNES, M. H. A survey of network flow applications. *Journal of Network and Computer Applications*, Elsevier, v. 36, n. 2, p. 567–581, 2013.

LIAO, H.-J.; LIN, C.-H. R.; LIN, Y.-C.; TUNG, K.-Y. Intrusion detection system: A comprehensive review. *Journal of Network and Computer Applications*, Elsevier, v. 36, n. 1, p. 16–24, 2013.

MARCHETTI, M.; PIERAZZI, F.; COLAJANNI, M.; GUIDO, A. Analysis of high volumes of network traffic for advanced persistent threat detection. *Computer Networks*, Elsevier, v. 109, p. 127–141, 2016.

MEHMOOD, A.; UMAR, M. M.; SONG, H. Icmds: Secure inter-cluster multiple-key distribution scheme for wireless sensor networks. *Ad Hoc Networks*, Elsevier, v. 55, p. 97–106, 2017.

MEYER, P. L. Probabilidade: aplicações à estatística. In: *Probabilidade: aplicações à estatística*. [S.l.]: Livro Técnico, 1970.

MUDA, Z.; YASSIN, W.; SULAIMAN, M.; UDZIR, N. K-means clustering and naive bayes classification for intrusion detection. *Journal of IT in Asia*, v. 4, n. 1, p. 13–25, 2016.

NAIDU, V.; WHALLEY, J.; NARAYANAN, A. Generating rule-based signatures for detecting polymorphic variants using data mining and sequence alignment approaches. *Journal of Information Security*, n. 9, p. 265–298, 2018.

NITHYA, V.; PANDIAN, S. L.; MALARVIZHI, C. A survey on detection and prevention of cross-site scripting attack. *International Journal of Security and Its Applications*, v. 9, n. 3, p. 139–52, 2015.

OCAMPO, F. B. C. D.; CASTILLO, T. M. L. D.; GOMEZ, M. A. N. Automated signature creator for a signature based intrusion detection system with network attack detection capabilities (pancakes). *International Journal of Cyber-Security and Digital Forensics*, The Society of Digital Information and Wireless Communications, v. 2, n. 1, p. 25–36, 2013.

OLEJNIK, L.; CASTELLUCCIA, C. Towards web-based biometric systems using personal browsing interests. In: IEEE. *Availability, Reliability and Security (ARES), 2013 Eighth International Conference on*. [S.l.], 2013. p. 274–280.

PELLEG, D.; MOORE, A. W. et al. X-means: Extending k-means with efficient estimation of the number of clusters. In: *Icml*. [S.l.: s.n.], 2000. v. 1, p. 727–734.

RAHMAN, A.; KAWSHIK, K. R.; SOURAV, A. A.; GAJI, A. Advanced network scanning. *American Journal of Engineering Research (AJER)*, v. 5, n. 6, p. 38–42, 2016.

RAVALE, U.; MARATHE, N.; PADIYA, P. Feature selection based hybrid anomaly intrusion detection system using k means and rbf kernel function. *Procedia Computer Science*, Elsevier, v. 45, p. 428–435, 2015.

SADASIVAM, G. K.; HOTA, C.; ANAND, B. Honeynet data analysis and distributed ssh brute-force attacks. In: *Towards Extensible and Adaptable Methods in Computing*. [S.l.]: Springer, 2018. p. 107–118.

SAGALA, A. Automatic snort ids rule generation based on honeypot log. In: IEEE. *Information Technology and Electrical Engineering (ICITEE), 2015 7th International Conference on*. [S.l.], 2015. p. 576–580.

SCOTT, D. W. Multivariate density estimation and visualization. In: *Handbook of computational statistics*. [S.l.]: Springer, 2012. p. 549–569.

SHI, R.; NGAN, K. N.; LI, S. Jaccard index compensation for object segmentation evaluation. In: IEEE. *Image Processing (ICIP), 2014 IEEE International Conference on*. [S.l.], 2014. p. 4457–4461.

SHIM, K.-S.; YOON, S.-H.; LEE, S.-K.; KIM, M.-S. Sigbox: Automatic signature generation method for fine-grained traffic identification. *J. Inf. Sci. Eng.*, v. 33, n. 2, p. 537–569, 2017.

SILVERMAN, B. W. *Density estimation for statistics and data analysis*. [S.l.]: Routledge, 2018.

SUN, K.; PENG, P.; NING, P.; WANG, C. Secure distributed cluster formation in wireless sensor networks. In: IEEE. *Computer Security Applications Conference, 2006. ACSAC'06. 22nd Annual*. [S.l.], 2006. p. 131–140.

SZYNKIEWICZ, P.; KOZAKIEWICZ, A. Design and evaluation of a system for network threat signatures generation. *Journal of computational science*, Elsevier, v. 22, p. 187–197, 2017.

TRAMMELL, B.; BOSCHI, E. Bidirectional flow export using ip flow information export (ipfix)", rfc 5103. Internet Engineer Task Force, 2008.

USSATH, M.; CHENG, F.; MEINEL, C. Automatic multi-step signature derivation from taint graphs. In: IEEE. *Computational Intelligence (SSCI), 2016 IEEE Symposium Series on*. [S.l.], 2016. p. 1–8.

WEBB, A. R. *Statistical pattern recognition*. [S.l.]: John Wiley & Sons, 2003.

WERNER, T.; FUCHS, C.; GERHARDS-PADILLA, E.; MARTINI, P. Nebula-generating syntactical network intrusion signatures. In: IEEE. *Malicious and Unwanted Software (MALWARE), 2009 4th International Conference on*. [S.l.], 2009. p. 31–38.

WU, S. X.; BANZHAF, W. The use of computational intelligence in intrusion detection systems: A review. *Applied Soft Computing*, v. 10, n. 1, p. 1 – 35, 2010. ISSN 1568-4946. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1568494609000908>.