VISUALISASI SERANGAN GERAKAN LATERAL (*LATERAL MOVEMENT ATTACK*) DENGAN MENERAPKAN METODE *K-MEANS CLUSTERING*

SKRIPSI

Oleh:

SEPTIANI KUSUMA NINGRUM
09011182025018

JURUSAN SISTEM KOMPUTER

FAKULTAS ILMU KOMPUTER

UNIVERSITAS SRIWIJAYA

2023

LEMBAR PENGESAHAN

VISUALISASI SERANGAN GERAKAN LATERAL (*LATERAL MOVEMENT ATTACK*) DENGAN MENERAPKAN METODE *K-MEANS CLUSTERING*

SKRIPSI

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Program Studi Sistem Komputer
Jenjang S1

Oleh:

SEPTIANI KUSUMA NINGRUM
09011182025018

Palembang, Desember 2023

Pembimbing I	**Pembimbing II**
Prof. Deris Stiawan, M.T., Ph.D.	**Nurul Afifah, M.Kom.**
NIP. 197806172006041002	NIP. 199211102023212049

Mengetahui, 11/1/24
Ketua Jurusan Sistem Komputer

Dr. Ir. H. Sukemi, M.T.
NIP. 196612032006041001

ii

LEMBAR PERSETUJUAN

Telah diuji dan lulus pada :

Hari : Senin

Tanggal : 08 Januari 2024

Tim Penguji :

1. Ketua : Ahmad Fali Oklilas, M.T.
2. Sekretaris : Iman Saladin B. Azhar, M.MSI.
3. Penguji : Ahmad Heryanto, M.T.
4. Pembimbing I : Prof. Deris Stiawan, M.T., Ph.D.
5. Pembimbing II : Nurul Afifah, M.Kom

Mengetahui,
Ketua Jurusan Sistem Komputer

Dr. Ir. H. Sukemi, M.T.
NIP. 196612032006041001

HALAMAN PERNYATAAN

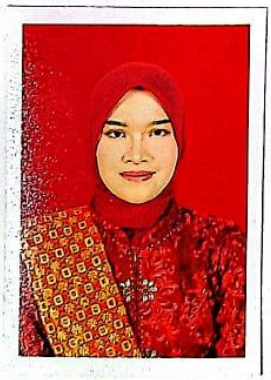
Yang bertanda tangan dibawah ini :

Nama	:	Septiani Kusuma Ningrum
NIM	:	09011182025018
Judul	:	Visualisasi Serangan Gerakan Lateral (*Lateral Movement Attack*) Dengan Menerapkan Metode *K-Means Clustering*

Hasil Pengecekan Software Turnitin : 3%

Menyatakan bahwa laporan tugas akhir saya merupakan hasil karya sendiri dan bukan hasil penjiplakan atau plagiat. Apabila ditemukan unsur penjiplakan atau plagiat dalam laporan tugas akhir ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya.

Demikian, pernyataan ini saya buat dalam keadaan sadar dan tanpa paksaan dari siapapun.



Palembang, Januari 2024

Septiani Kusuma Ningrum
NIM. 09011182025018

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Puji dan syukur penulis panjatkan kehadirat Allah SWT yang telah memberikan segala nikmat dan karunia-Nya sehingga penulis dapat menyelesaikan Skripsi dengan judul "**Visualisasi Serangan Gerakan Lateral (*Lateral Movement Attack*) Dengan Menerapkan Metode *K-Means Clustering***". Shalawat beriringkan salam senantiasa tercurahkan kepada Nabi Muhammad Shallallahu 'Alaihi Wasallam yang telah membawa kedamaian dan rahmat untuk semesta alam serta menjadi suri tauladan bagi umatnya.

Skripsi ini merupakan salah satu syarat untuk memenuhi sebagian kurikulum dan syarat kelulusan Mata Kuliah Skripsi pada Jurusan Sistem Komputer, Universitas Sriwijaya. Dalam penyusunan Skripsi ini, penulis telah banyak memperoleh bantuan, bimbingan dan saran baik moril maupun materil dari berbagai pihak. Oleh karena itu, ucapan terima kasih sebesar-besarnya diberikan kepada:

1. Allah SWT. yang selalu memberikan rahmat dan karunia-Nya
2. Ibu, Bapak, Adik-adik penulis tercinta serta seluruh keluarga besar yang telah banyak memberikan do'a, nasihat, serta motivasi kepada penulis selama ini.
3. Bapak Prof. Dr. Erwin, S.Si., M.Si. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Bapak Dr. Ir. H. Sukemi, M.T. selaku Ketua Jurusan Sistem Komputer Universitas Sriwijaya.
5. Bapak Prof. Deris Stiawan, M.T., PH.D., IPU., ASEAN ENG. selaku Dosen Pembimbing I Tugas Akhir dan juga Dosen Pembimbing Akademik.
6. Ibu Nurul Afifah, M.Kom. selaku Dosen Pembimbing II Tugas Akhir.
7. Mbak Renny Virgasari dan Bapak Yopi selaku admin Jurusan Sistem Komputer.
8. Rizky Elinda Sari, Nurul Fitria, Vijiantika Fajaria Sastri, Zulianti selaku sahabat seperjuangan yang selalu ada saat susah maupun senang.

9. Kakak - Kakak tingkat SK Unggulan 2019 yang termasuk tim riset COMNETS.
10. Teman teman seperjuangan Jurusan Sistem Komputer Angkatan 2020 terkhusus kelas B.
11. Seluruh pihak yang membantu dalam menyelesaikan laporan ini yang tidak bisa disebutkan satu persatu.
12. Almamater

Penulis menyadari bahwa masih terdapat kekurangan dalam penulisan Skripsi ini, baik dari materi maupun teknik penyajiannya, mengingat kurangnya pengetahuan dan pengalaman penulis. Untuk itu, penulis mengharapkan adanya kritik dan saran yang membangun agar dapat memperbaiki kekurangan-kekurangan tersebut kedepannya nanti.

Penulis berharap semoga penulisan Skripsi ini dapat menjadi tambahan wawasan dan ilmu pengetahuan bagi mahasiswa yang memerlukan khususnya mahasiswa Fakultas Ilmu Komputer Universitas Sriwijaya secara langsung maupun tidak langsung sebagai sumbangan pikiran dalam peningkatan mutu pembelajaran.

Palembang, Januari 2024
Penulis,

Septiani Kusuma Ningrum
NIM. 09011182025018

VISUALISASI SERANGAN GERAKAN LATERAL (*LATERAL MOVEMENT ATTACK*) DENGAN MENERAPKAN METODE *K-MEANS CLUSTERING*

SEPTIANI KUSUMA NINGRUM (09011182025018)

Jurusan Sistem Komputer, Fakultas Ilmu Komputer

Universitas Sriwijaya

Email: 09011182025018@student.unsri.ac.id

ABSTRAK

Gerakan lateral adalah salah satu fase terpenting dalam serangan *Advanced Persistent Threat* (APT), yang bertujuan untuk memasuki sumber daya lain dan memperoleh hak istimewa yang lebih besar di jaringan target. Pelaku serangan biasanya memanfaatkan teknik rekayasa sosial (misalnya, *phishing*, *pretexting*, *baiting*) untuk memperdaya orang dalam jaringan agar menjalankan kode berbahaya atau menyerahkan kredensial. Ini me-mungkinkan penyerang untuk mendapatkan akses ke komputer korban dan secara bertahap mencari informasi berharga dengan memanfaatkan kerentanan entitas intranet lainnya. Penggunaan metode K-means untuk *clustering* aktivitas *benign* dan aktivitas *malicious*, yang dikombinasikan dengan metode *Principal Component Analysis* (PCA). Pendekatan ini memberikan kinerja yang baik dalam menvisualisasikan serangan yang bergerak secara *lateral*. Mengkombinasikan ke empat dataset diantaranya dataset connection, dataset files, dataset DNS, dan dataset HTTP dapat memberikan visualisasi yang baik untuk menggambarkan perbedaan berdasarkan aktivitas yang dilakukan oleh user. Melalui penggabungan data ini, menciptakan pemahaman yang lebih baik tentang aktivitas normal dan *malicious* dalam jaringan. Penggunaan metode validasi yakni *Elbow method* yang membuktikan keberadaan *elbow point* yang terlihat jelas, maka *cluster* optimal mudah untuk diidentifikasi, sementara grafik yang tidak memiliki *elbow point* yang jelas akan memberikan penilaian yang tidak dapat diandalkan. Sehingga penelitian ini mengindikasikan bahwa penggunaan *silhouette method* merupakan metode evaluasi yang efektif dalam mengukur kualitas *cluster* yang ideal.

Kata Kunci: *Lateral Movement, K-Means Clustering, Principal Component Analysis*

VISUALIZATION OF LATERAL MOVEMENT ATTACKS USING K-MEANS CLUSTERING METHOD

SEPTIANI KUSUMA NINGRUM (09011182025018)

Department of Computer Systems, Faculty of Computer Science

Sriwijaya University

Email: 09011182025018@student.unsri.ac.id

ABSTRACT

The lateral movement is one of the most crucial phases in an Advanced Persistent Threat (APT) attack, aiming to penetrate other resources and gain greater privileges within the target network. Attackers typically exploit social engineering techniques (such as phishing, pretexting, baiting) to deceive individuals within the network into running malicious code or surrendering credentials. This enables the attacker to gain access to the victim's computer and gradually seek valuable information by exploiting vulnerabilities in other intranet entities. Utilizing the K-means method for clustering benign and malicious activities, combined with Principal Component Analysis (PCA), this approach delivers good performance in visualizing laterally moving attacks. Combining the four datasets, connection dataset, files dataset, DNS dataset, and HTTP dataset, can provide a clear visualization to illustrate differences based on user activities. Through this data integration, it fosters a better understanding of normal and malicious activities within the network. The use of validation methods like the Elbow method proves the existence of a clear elbow point, making it easy to identify the optimal clusters. On the other hand, graphs lacking a clear elbow point might provide unreliable assessments. Hence, this research indicates that using the silhouette method is an effective evaluation technique for measuring the quality of an ideal cluster.

Keywords: *Lateral Movement, K-Means Clustering, Principal Component Analysis*

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

BAB I
PENDAHULUAN

1.1 Latar Belakang

Advanced Persistent Threats (APTs) adalah salah satu ancaman *modern* yang paling umum terhadap keamanan informasi penting bagi pemerintah, organisasi, dan perusahaan [1]. Perusahaan keamanan siber seperti *Kaspersky* dan *Fireeye* melaporkan secara terbuka sejumlah besar peristiwa APT setiap tahun [2], [3]. Pelaku serangan menggunakan teknik yang canggih untuk melampaui sistem pertahanan dan terus menerus meretas aset inti *internal*. Meskipun strategi monetasi berbeda-beda, semua kerusakan terhadap aset digital penting perusahaan dapat menyebabkan kerugian yang merusak baik secara ekonomi maupun reputasi [4].

Serangan dimulai dengan mengorbankan beberapa host atau akun pengguna dalam jaringan, dan meninggalkan *backdoors* untuk mendapatkan akses terus menerus ke aset internal, jenis serangan ini umumnya dikenal sebagai *Advanced Persistent Threat* (APT) [5]. Pelaku APT biasanya menggunakan serangan *spear phishing* atau *watering hole* untuk menemukan pijakan dalam jaringan target sebagai batu loncatan untuk mencapai sistem lain sampai mendapatkan akses ke sistem kritis, seperti *server file* yang berisi dokumen rahasia, gerakan inkremental menuju sistem kritis disebut sebagai gerakan *lateral* (*Lateral Movement*) [6].

Serangan Gerakan Lateral (*Lateral Movement Attack*) adalah faktor utama yang bertanggung jawab atas banyak Ancaman Persisten Lanjutan (*Advanced Persistent Threat*) saat ini [7]. Gerakan lateral terdiri dari teknik yang digunakan pelaku untuk memasuki dan mengendalikan sistem jarak jauh di jaringan [8]. Setelah memasuki jaringan, pelaku mempertahankan akses berkelanjutan dengan bergerak melalui lingkungan yang dikompromikan dan mendapatkan peningkatan hak istimewa menggunakan berbagai alat [7].

Pendeteksian gerakan lateral memiliki dua komponen. Pertama, penyerang melakukan serangan secara perlahan dan jarang agar tidak terdeteksi [5]. Kedua, banyak metode penyerangan yang dapat digunakan, seperti *Pass the Hash*, *Remote Desktop Protocol*, *Remote Service*, dan lain lain dapat digunakan selama *Lateral Movement* [9]. Pada penelitian [5] mengungkapkan solusi untuk masalah kedua dengan menerapkan algoritma *Random Forest* untuk mengeksplorasi fitur berbasis

grafik yang diekstrak dari log autentikasi host serta melakukan rekayasa fitur untuk memfasilitasi deteksi dini *lateral movement*. Model tersebut mendapatkan nilai *recall* 93,04%, *precision* 97,02%, dan *F1 Score* 95%.

Serangan berbahaya telah menjadi lebih canggih, sehingga pelaku serangan menggunakan berbagai teknik untuk menghindar dan menyembunyikan informasi guna mencegah pendeteksian oleh IDS. IDS (*Intrusion Detection System*) adalah sistem yang dapat mendeteksi aktivitas berbahaya dengan menyesuaikan pola serangan yang diketahui (yaitu berbasis *signature*) atau mengamati aktivitas anomali [10]. Pada IDS berbasis anomali, pendeteksian dilakukan dengan membandingkan aktivitas yang dipantau dengan profil dasar. Profil dasar ini dibuat selama pelatihan dan ada ambang batas yang ditetapkan. Setiap penyimpangan aktivitas dari profil dasar dianggap sebagai perilaku mencurigakan [11].

Pada penelitian [12] menunjukkan adanya pembelajaran tanpa pengawasan (*unsupervised learning*) dalam deteksi *lateral movement*, memungkinkan pendeteksian perilaku anomali yang lebih kompleks dalam data autentikasi. Eksperimen dengan *XGBoost* tersebut menghasilkan recall 86,51%.

Pada penelitian lainnya [13] menerapkan *LogitBoost* dalam mendeteksi *lateral movement* berbasis *Remote Desktop Protocol* (RDP), *Remote Desktop Protocol* adalah metode yang digunakan dalam *lateral movement* untuk melakukan autentikasi ke *host* yang tidak diizinkan sehingga meninggalkan catatan pada *log host* dan jaringan. Penelitian ini menghasilkan akurasi 99,99%, recall 99,47% dan f1 score 99,7%.

Supervised learning juga dapat digunakan dalam menyelidiki kemampuan mendeteksi lateral movement di *log windows active directory* [14]. Investigasi yang dilakukan pada penelitian tersebut dengan mengevaluasi serta membandingkan kinerja *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN), *Decision Tree* (DT), *Random Forest* (RF), dan *Artificial Neural Network* (ANN) dalam pengaturan klasifikasi *multiclass*. Sehingga penelitian tersebut menyimpulkan bahwa kelima algoritma bekerja dengan akurasi 98%. *Random Forest* (RF) menampilkan f1 *score* 88% dan *recall* 85,8%, *Support Vector Machine* (SVM) menghasilkan presisi metrik kinerja 97,2%.

Salah satu pembelajaran tanpa pengawasan (*unsupervised learning*) yaitu *k-means* dan *spherical k*-means diusulkan pada penelitian ini [15] guna mendeteksi gerakan lateral (*lateral movement*) dengan mengidentifikasi koneksi antar-sistem yang anomali. Pada penelitian [16] *k-means* dipilih karena memiliki aturan – aturan yang mudah dipahami, ditafsirkan dan divisualisasikan. Selain itu, algoritma *k-means* dengan mudah menangani data heterogen yang berisi berbagai fitur yang dihasilkan oleh sumber yang berbeda.

Pada penelitian [17] mengevaluasi dua varian metode pembelajaran grafik yaitu *Graph Learning with Local View* (GL-LV) dan *Graph Learning with Global View* (GL-GV) dalam menunjukkan kekuatan topologi grafik untuk mendeteksi gerakan lateral. Penelitian ini mencapai 80% TPR serta mampu mengurangi FPR menjadi 0%. Namun penelitian ini mengalami kesulitan dalam pelabelan suatu aktivitas sebagai anomali sehingga membutuhkan pembelajaran mesin dan algoritma AI. Sehingga penulis melakukan penelitian dengan judul "**Visualisasi Serangan Gerakan Lateral (*Lateral Movement Attack*) Dengan Menerapkan Metode *K-Means Clustering***" dengan harapan metode *k-means* mampu menyelesaikan persoalan pelabelan aktivitas anomali serta persoalan fitur pada penelitian sebelumnya.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang ada, permasalahan yang akan dibahas pada penelitian ini ada tiga yaitu:

1. Keterbatasan analisis fitur pada data serangan gerakan *lateral* (*lateral movement attack*) yang berisi lalu lintas komunikasi *internal* dan eksternal perusahaan.
2. Kompleksitas dalam mengidentifikasi aktivitas normal dan serangan berbasis gerakan *lateral*.
3. Pengujian *clustering* dalam mengidentifikasi model *cluster* yang ideal.

1.3 Batasan Masalah

Berikut batasan masalah pada penelitian ini, yaitu:

1. Penelitian ini hanya menggunakan *dataset Picodomain* yang berasal dari lingkungan berbasis *Windows* kecil yang hanya mengandung elemen-elemen paling penting yang umumnya ditemukan dalam domain tingkat perusahaan.
2. Penelitian ini hanya menggunakan 5 macam log dari 17 macam tipe log yang berbeda beda, seperti *Connection, Files, Domain Name System, HTTP*.

1.4 Tujuan

1. Menerapkan metode *clustering* menggunakan algoritma *k-means* pada data serangan *lateral movement*.
2. Memvisualisasikan hasil *clustering* sehingga mampu membedakan aktivitas normal dan aktivitas mencurigakan yang dipicu dengan gerakan *lateral*.
3. Mendapatkan model *cluster ideal* dengan menggunakan metode validasi *Silhouette* dan *Elbow*.

1.5 Manfaat

1. Penyajian indikator *lateral movement* pada penelitian ini diharapkan berkontribusi dalam pengembangan IDS yang andal.
2. Informasi yang diperoleh dapat membantu perusahaan-perusahaan di berbagai industri dalam upaya antisipasi dini terhadap ancaman keamanan data.
3. Pengujian *cluster* menggunakan *Silhouette* dan *Elbow method* digunakan untuk memahami kualitas serta model *cluster* yang ideal.

1.6 Metodologi Penelitian

Metodologi yang diterapkan dalam penulisan penelitian ini melalui beberapa tahapan sebagai berikut:

1. Studi Pustaka / Literatur

Tahap ini diawali dengan mencari informasi dan masalah yang sesuai dan relevan untuk diangkat sebagai penelitian. Kemudian mencari beberapa referensi

seperti artikel, publikasi ilmiah, buku dan sumber lain yang relevan dan berkaitan langsung dengan topik Skripsi ini.

2. Perancangan Sistem

Tahap ini membahas masalah proses bagaimana cara membangun metode atau pendekatan tertentu, perangkat lunak maupun perangkat keras yang digunakan untuk konfigurasi sistem.

3. Pengujian

Pada tahap ini dilakukan pengujian berdasarkan metode yang digunakan dalam penelitian dan metode-metode yang digunakan oleh penelitian sebelumnya, sehingga didapatkan hasil yang sesuai.

4. Analisa

Tahap ini dilakukan pengolahan data dan analisa data yang didapatkan dari hasil pengujian yang dilakukan sebelumnya untuk mendapatkan data yang aktual. Kemudian hasil akan dianalisis dengan tujuan untuk mengetahui kekurangan pada hasil perancangan sistem tersebut.

5. Kesimpulan dan Saran

Tahapan ini merupakan langkah akhir, hasil dari semua langkah yang dilakukan sebelumnya akan dirumuskan menjadi suatu kesimpulan serta saran dibutuhkan untuk mengetahui kekurangan pada hasil perancangan dan faktor penyebabnya.

1.7 Sistematika Penulisan

Adapun sistematika dalam penulisan Skripsi ini adalah sebagai berikut:

BAB I PENDAHULUAN

Pada bab ini merupakan penjelasan berisi mengenai Latar belakang, Rumusan Masalah, Batasan Masalah, Tujuan, Manfaat, Metodologi penelitian dan Sistematika penulisan yang digunakan dalam Skripsi ini.

BAB II TINJAUAN PUSTAKA

Pada bab ini berisi mengenai bacaan *literature* yang menjadi referensi serta penjelasan pendukung dari penelitian visualisasi serangan gerakan lateral (*Lateral*

Movement Attack).

BAB III METODOLOGI PENELITIAN

Pada bab ini membahas proses penelitian, kerangka penelitian, serta menjelaskan metodologi penelitian.

BAB IV HASIL DAN ANALISA

Pada bab ini menjelaskan hasil dari penelitian dan menvisualisasikan serangan gerakan lateral (*Lateral Movement Attack*) dengan menerapkan metode *K-Means Clustering*.

BAB V KESIMPULAN DAN SARAN

Pada bab ini berisi tentang kesimpulan dari penelitian yang dilakukan serta saran untuk penelitian selanjutnya dimasa mendatang.

DAFTAR PUSTAKA

[1] Y. Fang, C. Wang, Z. Fang, and C. Huang, "LMTracker: Lateral movement path detection based on heterogeneous graph embedding," *Neurocomputing*, vol. 474, pp. 37–47, 2022, doi: 10.1016/j.neucom.2021.12.026.

[2] K. Inc., "Kaspersky apt trends report q1 2020," 2020. https://securelist.com/apt-trends-report-q1-2020/96826/

[3] FireEye Inc., "a dual espionage and cyber crime operation," 2019. https://content.fireeye.com/apt-41/rpt-apt41/

[4] C. Dong, J. Yang, S. Liu, Z. Wang, Y. Liu, and Z. Lu, "C-BEDIM and S-BEDIM: Lateral movement detection in enterprise network through behavior deviation measurement: C-BEDIM and S-BEDIM," *Comput. Secur.*, vol. 130, p. 103267, 2023, doi: 10.1016/j.cose.2023.103267.

[5] H. Bian, T. Bai, M. A. Salahuddin, N. Limam, A. A. Daya, and R. Boutaba, "Uncovering Lateral Movement Using Authentication Logs," *IEEE Trans. Netw. Serv. Manag.*, vol. 18, no. 1, pp. 1049–1063, 2021, doi: 10.1109/TNSM.2021.3054356.

[6] A. Niakanlahiji, J. Wei, M. R. Alam, Q. Wang, and B. T. Chu, "ShadowMove: A stealthy lateral movement strategy," *Proc. 29th USENIX Secur. Symp.*, pp. 559–576, 2020.

[7] CrowdStrike, "Lateral Movement Explained," 2021. https://www.crowdstrike.com/cybersecurity-101/lateral-movement/

[8] GitBook, "Lateral Movement," 2021. https://dmcxblue.gitbook.io/red-team-notes-2-0/

[9] T. M. Corporation, "Lateral Movement," 2019. https://attack.mitre.org/

[10] A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, "Survey of intrusion detection systems: techniques, datasets and challenges," *Cybersecurity*, vol. 2, no. 1, Dec. 2019, doi: 10.1186/s42400-019-0038-7.

[11] I. Ghafir *et al.*, "Hidden markov models and alert correlations for the prediction of advanced persistent threats," *IEEE Access*, vol. 7, pp. 99508–99520, 2019, doi: 10.1109/ACCESS.2019.2930200.

[12] D. Kushwaha *et al.*, "Lateral Movement Detection Using User Behavioral Analysis," vol. d, pp. 1–20, 2022, [Online]. Available:

http://arxiv.org/abs/2208.13524

[13] T. Bai, H. Bian, A. A. Daya, M. A. Salahuddin, N. Limam, and R. Boutaba, "A Machine Learning Approach for RDP-based Lateral Movement Detection," *Proc. - Conf. Local Comput. Networks, LCN*, vol. 2019-Octob, pp. 242–245, 2019, doi: 10.1109/LCN44214.2019.8990853.

[14] V. Uppströmer and H. Råberg, "Detecting Lateral Movement in Microsoft Active Directory Log Files A supervised machine learning approach," no. June, 2019, [Online]. Available: www.bth.se

[15] B. A. Powell, "Role-based lateral movement detection with unsupervised learning," *Intell. Syst. with Appl.*, vol. 16, no. July, p. 200106, 2022, doi: 10.1016/j.iswa.2022.200106.

[16] J. Henriques, F. Caldeira, T. Cruz, and P. Simões, "Combining k-means and xgboost models for anomaly detection using log datasets," *Electron.*, vol. 9, no. 7, pp. 1–17, 2020, doi: 10.3390/electronics9071164.

[17] B. Bowman, C. Laprade, Y. Ji, and H. H. Huang, "Detecting lateral movement in enterprise computer networks with unsupervised graph AI," *RAID 2020 Proc. - 23rd Int. Symp. Res. Attacks, Intrusions Defenses*, pp. 257–268, 2020.

[18] M. Chen, Y. Yao, J. Liu, B. Jiang, L. Su, and Z. Lu, "A novel approach for identifying lateral movement attacks based on network embedding," *Proc. - 16th IEEE Int. Symp. Parallel Distrib. Process. with Appl. 17th IEEE Int. Conf. Ubiquitous Comput. Commun. 8th IEEE Int. Conf. Big Data Cloud Comput. 11t*, no. December, pp. 708–715, 2019, doi: 10.1109/BDCloud.2018.00107.

[19] I. J. King and H. H. Huang, "Euler: Detecting Network Lateral Movement via Scalable Temporal Graph Link Prediction," 2022, doi: 10.14722/ndss.2022.24107.

[20] Mandiant, "Exposing One of China's Cyber Espionage Units", [Online]. Available: www.mandiant.com

[21] T. Saranya, S. Sridevi, C. Deisy, T. D. Chung, and M. K. A. A. Khan, "Performance Analysis of Machine Learning Algorithms in Intrusion Detection System: A Review," in *Procedia Computer Science*, 2020, vol.

171, pp. 1251–1260. doi: 10.1016/j.procs.2020.04.133.

[22] Z. Ahmad, A. Shahid Khan, C. Wai Shiang, J. Abdullah, and F. Ahmad, "Network intrusion detection system: A systematic study of machine learning and deep learning approaches," *Trans. Emerg. Telecommun. Technol.*, vol. 32, no. 1, p. e4150, 2021.

[23] J. Díaz-Verdejo, J. Muñoz-Calle, A. E. Alonso, R. E. Alonso, and G. Madinabeitia, "On the Detection Capabilities of Signature-Based Intrusion Detection Systems in the Context of Web Attacks," *Appl. Sci.*, vol. 12, no. 2, 2022, doi: 10.3390/app12020852.

[24] J. J. Singh, H. Samuel, and P. Zavarsky, "Impact of Paranoia Levels on the Effectiveness of the ModSecurity Web Application Firewall," in *2018 1st International Conference on Data Intelligence and Security (ICDIS)*, 2018, pp. 141–144. doi: 10.1109/ICDIS.2018.00030.

[25] Syed Ameer Sohail, "Feature Scaling in Machine Learning: Robust Scaler and MinMax Scaler with K-Means Clustering," 2021. https://medium.com/@syedar.sohail/outlier-handling-using-robust-scaler-a-python-tutorial-613d174b58eb (accessed Jul. 26, 2023).

[26] T. H. Sandhu, "Machine Learning and Natural Language Processing – a Review," *Int. J. Adv. Res. Comput. Sci.*, vol. 9, no. 2, pp. 582–584, 2018, doi: 10.26483/ijarcs.v9i2.5799.

[27] M. W. Berry, A. Mohamed, and B. W. Yap, *Supervised and unsupervised learning for data science*. Springer, 2019.

[28] G. Kovács, "An empirical comparison and evaluation of minority oversampling techniques on a large number of imbalanced datasets," *Appl. Soft Comput.*, vol. 83, p. 105662, 2019.

[29] J. Brandt and E. Lanzén, "A Comparative Review of SMOTE and ADASYN in Imbalanced Data Classification," p. 42, 2020.

[30] S. Miao, L. Zheng, J. Liu, and and H. Jin, "K-means Clustering Based Feature Consistency Alignment for Label-free Model Evaluation," 2023, [Online]. Available: http://arxiv.org/abs/2304.09758

[31] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and

advances in the era of big data," *Inf. Sci. (Ny).*, vol. 622, pp. 178–210, 2023, doi: https://doi.org/10.1016/j.ins.2022.11.139.

[32] C. Shi, B. Wei, S. Wei, W. Wang, H. Liu, and J. Liu, "A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm," *Eurasip J. Wirel. Commun. Netw.*, vol. 2021, no. 1, 2021, doi: 10.1186/s13638-021-01910-w.

[33] A. Maravillas, "Integration of K-means Algorithm and Elbow Method in Clustering the Bivalve Species." 2023.

[34] M. Cui, "Introduction to the K-Means Clustering Algorithm Based on the Elbow Method," *Clausius Sci. Press*, vol. 1, no. 1, pp. 5–8, 2020, doi: 10.23977/accaf.2020.010102.

[35] B. Pena, L. Blakely, and M. Reno, "Parameter Tuning Analysis for Phase Identification Algorithms in Distribution System Model Calibration," 2021. doi: 10.1109/KPEC51835.2021.9446218.

[36] J. Gabry, D. Simpson, A. Vehtari, M. Betancourt, and A. Gelman, "Visualization in Bayesian Workflow," *J. R. Stat. Soc. Ser. A Stat. Soc.*, vol. 182, no. 2, pp. 389–402, 2019, doi: 10.1111/rssa.12378.

[37] marastats., "General marathon stats," 2019. https://marastats.com/marathon/

[38] H. De Plaen and J. A. K. Suykens, "A Dual Formulation for Probabilistic Principal Component Analysis," 2023, [Online]. Available: http://arxiv.org/abs/2307.10078

[39] Cisco, "Cybersecurity Operational." www.cisco.com

[40] S. Bagui *et al.*, "Detecting Reconnaissance and Discovery Tactics from the MITRE ATT&CK Framework in Zeek Conn Logs Using Spark's Machine Learning in the Big Data Framework," *Sensors (Basel).*, vol. 22, no. 20, 2022, doi: 10.3390/s22207999.

[41] Y. Januzaj, E. Beqiri, and A. Luma, "Determining the Optimal Number of Clusters using Silhouette Score as a Data Mining Technique.," *Int. J. Online Biomed. Eng.*, vol. 19, no. 4, pp. 174–182, 2023, [Online]. Available: http://e-resources.perpusnas.go.id:2048/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=edo&AN=162884448&site=eds-live

[42] A. Punhani, N. Faujdar, K. K. Mishra, and M. Subramanian, “Binning-Based Silhouette Approach to Find the Optimal Cluster Using K-Means,” *IEEE Access*, vol. 10, pp. 115025–115032, 2022, doi: 10.1109/ACCESS.2022.3215568.

[43] K. R. Shahapure and C. Nicholas, “Cluster Quality Analysis Using Silhouette Score,” in *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, 2020, pp. 747–748. doi: 10.1109/DSAA49011.2020.00096.