E-ISSN: 2623-1700
P-ISSN: 1693-3672

ANALISIS ANOMALI TRAFIK JARINGAN ENTERPRISE BERBASIS LOG FIREWALL MENGGUNAKAN PROFILING STATISTIK DAN DETEKSI BERBASIS ENTITAS

Yudhi Biantoro[1], Yasmiati[2], Ramadhani Ulan Sari[3], Jenih[4], F.A. Ricky Bayu Styanto[5], Suharyanto[6]
[1,2,3,4,5,6]Universitas Respati Indonesia
[1]yudhi.biantoro@gmail.com, [2]yasmiati@urindo.ac.id, [3]ramadhani.ulansari@urindo.ac.id,
[4]jenih@urindo.ac.id, [5]ricky@urindo.ac.id, [6]suharyanto@urindo.ac.id

ABSTRAK

Deteksi anomali trafik jaringan berbasis *signature* cenderung tidak efektif terhadap pola ancaman baru, sementara lingkungan *enterprise* menghasilkan log yang besar, heterogen, dan sulit dianalisis secara manual. Studi ini menyajikan analisis empiris menggunakan data log aktual *firewall* dan *web filtering* pada infrastruktur organisasi untuk mengidentifikasi pola trafik menyimpang, konsentrasi aktivitas, serta indikator risiko keamanan. Melalui pendekatan eksploratif tanpa label (*unsupervised exploratory analysis*), dataset diolah melalui tahapan *Extract, Transform, Load* (ETL) yang deterministik, meliputi *parsing key-value*, normalisasi struktur, profiling statistik, dan analisis berbasis entitas (*source IP* dan *hostname*). Hasil profiling menunjukkan dominasi aktivitas yang sangat timpang (*concentration anomaly*), di mana sebagian kecil entitas mendominasi mayoritas *event* yang diblokir oleh kebijakan keamanan. Selain merekam tingginya trafik non-bisnis terkait aktivitas *gaming* dan *advertising* pada jam kerja, analisis ini berhasil mengisolasi klaster *endpoint* berisiko tinggi pada kategori *potentially unwanted program, proxy avoidance*, situs berbahaya, dan *cryptocurrency*. Penelitian ini menyimpulkan bahwa meskipun data bersifat *snapshot* jangka pendek, analisis *entity-centric* berbasis log *firewall* mampu menyediakan sinyal operasional yang kuat untuk kebutuhan triase keamanan awal (*initial security triage*) secara cepat. Riset selanjutnya disarankan mengintegrasikan analisis *time-series*, *unsupervised machine learning*, dan korelasi lintas-sumber log untuk membangun *baseline* perilaku jangka panjang yang adaptif.

Kata kunci: anomali trafik, log firewall, web filtering, enterprise security, behavior analytics.

ABSTRACT

Network traffic anomaly detection is becoming increasingly important because signature-based approaches tend to be ineffective against new patterns, while enterprise environments generate large, heterogeneous logs that are difficult to analyze manually. This study presents an empirical analysis using actual firewall and web filtering log data from an organizational infrastructure to identify deviant traffic patterns, activity concentrations, and security risk indicators. Through an unsupervised exploratory approach, the dataset is processed through a deterministic Extract, Transform, Load (ETL) pipeline involving key-value parsing, structural normalization, statistical profiling, and entity-based analysis (source IP and hostname). The profiling results reveal highly skewed activity distribution (concentration anomaly), where a small fraction of entities dominates the majority of events blocked by security policies. Aside from recording dense non-business traffic related to gaming and advertising during working hours, the analysis successfully

E-ISSN: 2623-1700
P-ISSN: 1693-3672

isolates high-risk endpoint clusters within the categories of potentially unwanted programs, proxy avoidance, malicious websites, and cryptocurrency. This study concludes that despite the short-term snapshot nature of the data, an entity-centric analysis of firewall logs provides strong operational signals for rapid initial security triage. Future research is recommended to integrate time-series analysis, unsupervised machine learning, and cross-source log correlation to establish an adaptive, long-term behavioral baseline.

Keywords: *traffic anomalies, firewall logs, web filtering, enterprise security, behavioral analytics.*

PENDAHULUAN

Anomali trafik jaringan secara umum dipahami sebagai penyimpangan signifikan dari pola normal, dan penyimpangan ini dapat berhubungan dengan penyalahgunaan jaringan, gangguan layanan, maupun aktivitas berbahaya. Literatur beberapa tahun terakhir menegaskan bahwa pendekatan signature-based tetap penting, tetapi memiliki keterbatasan untuk mendeteksi perilaku baru atau yang belum pernah dimodelkan sebelumnya, sehingga pendekatan anomaly-based tetap relevan dalam operasi keamanan modern [1], [2]. Di saat yang sama, log jaringan dan log firewall menyediakan jejak perilaku yang kaya, tetapi pemanfaatannya sangat bergantung pada kualitas parsing, pembentukan fitur, dan konteks operasional tempat log tersebut dihasilkan [3], [4].

Riset terbaru juga memperlihatkan bahwa analisis anomali tidak lagi berhenti pada statistik sederhana. Arah riset telah bergerak menuju pendekatan berbasis machine learning, analisis temporal, representasi spasiotemporal, serta model graf dan federated learning untuk menghadapi kompleksitas pola trafik modern [5]–[8]. Namun, studi berbasis data nyata organisasi tetap penting karena banyak pendekatan akademik diuji pada dataset publik atau data sintetis, sedangkan log operasional enterprise memiliki karakteristik lokal, kebijakan filtering, dan bias kategorisasi yang berbeda [4], [5].

Berdasarkan *research gap* yang teridentifikasi, kajian ini difokuskan untuk menjawab tiga pertanyaan utama. Pertama, pola anomali apa yang muncul pada log firewall yang dikaji. Kedua, apakah anomali lebih dominan berasal dari konsentrasi pengguna/host tertentu, kategori konten tertentu, atau kombinasi keduanya. Ketiga, peluang penelitian lanjutan apa yang paling relevan jika organisasi ingin menaikkan

E-ISSN: 2623-1700
P-ISSN: 1693-3672

kemampuan dari monitoring berbasis rule menjadi deteksi anomali yang lebih adaptif. Oleh karena itu, studi ini tidak berfokus pada pembangunan model prediktif baru, melainkan pada analisis empiris terhadap log firewall aktual untuk menghasilkan insight operasional. Selain itu, penelitian yang secara khusus mengkaji log firewall/web filtering sebagai sumber utama analisis anomali berbasis perilaku (behavioral anomaly) dan konsentrasi entitas (entity concentration) masih relatif terbatas. Lebih lanjut, belum banyak studi yang menekankan interpretasi operasional dari anomali—yaitu bagaimana hasil analisis dapat langsung digunakan untuk triase keamanan dan pengambilan keputusan di lingkungan organisasi. Oleh karena itu, kajian ini mengisi celah tersebut dengan menyajikan analisis empiris berbasis data nyata, dengan fokus pada interpretasi behavior anomaly, concentration anomaly, dan policy-driven risk exposure dalam konteks operasional enterprise.

PENELITIAN SEBELUMNYA

Dibandingkan penelitian terdahulu, studi ini menawarkan pendekatan yang lebih operasional melalui analisis log nyata pada lingkungan enterprise. Pertama, studi ini menggunakan data log firewall/web filtering dari suatu lingkungan infrastruktur enterprise, bukan dataset publik atau sintetis. Kedua, pendekatan yang digunakan tidak bergantung pada model machine learning kompleks, melainkan pada profiling statistik dan analisis berbasis entitas yang tetap mampu mengungkap pola anomali secara bermakna. Ketiga, kajian ini memperkenalkan klasifikasi anomali multi-dimensi yang mencakup behavior anomaly, concentration anomaly, dan policy-driven risk exposure, yang memberikan perspektif baru dalam interpretasi anomali jaringan. Keempat, analisis dilakukan pada level entitas (srcip dan hostname), sehingga lebih relevan untuk investigasi operasional. Oleh karena itu, studi ini menjembatani kesenjangan antara pendekatan akademik dan kebutuhan praktis dalam operasi keamanan jaringan. Penjelasan tentang penelitian-penelitian sebelumnya terlihat pada Tabel 1.

E-ISSN: 2623-1700
P-ISSN: 1693-3672

Tabel 1. Penelitian Sebelumnya yang Relevan

No	Penulis (Tahun)	Fokus Studi	Pendekatan	Jenis Data	Relevansi
1	Moustafa et al. (2019)[1]	Review sistem deteksi anomali jaringan	Survey konseptual	Dataset publik	Dasar teori anomaly detection
2	Abbasi et al. (2021)[2]	Monitoring trafik berbasis deep learning	Deep learning	Traffic/flow data	Perkembangan metode ML
3	Landauer et al. (2023)[3]	Deteksi anomali pada log data	Deep learning & log analysis	Log data	Relevan untuk preprocessing & log analysis
4	Komadina et al. (2024)[4]	Analisis anomali pada firewall logs	Comparative analysis	Firewall logs	Paling dekat dengan studi ini
5	Fosić et al. (2023)[5]	Deteksi anomali NetFlow	Supervised ML	NetFlow	Representasi pendekatan berbasis flow
6	Ji et al. (2024)[6]	Deteksi anomali berbasis spatiotemporal	Deep learning (attention)	Traffic data	Pendekatan temporal modern
7	Schummer et al. (2024)[7]	Implementasi ML untuk deteksi anomali	Machine learning	Traffic data	Sistem deteksi berbasis ML
8	Zhao et al. (2025)[8]	Deteksi anomali berbasis graph federated	Graph & federated learning	Network traffic	Pendekatan mutakhir

Dibandingkan dengan penelitian sebelumnya yang umumnya berfokus pada pengembangan model berbasis machine learning menggunakan dataset publik, studi

E-ISSN: 2623-1700
P-ISSN: 1693-3672

ini menekankan analisis empiris berbasis data nyata dari lingkungan enterprise. Pendekatan yang digunakan tidak bergantung pada label maupun kompleksitas model, melainkan pada profiling statistik dan analisis berbasis entitas yang lebih mudah diinterpretasikan secara operasional. Perbedaan utama terletak pada orientasi penelitian, di mana studi sebelumnya menitikberatkan pada akurasi deteksi serangan, sedangkan studi ini berfokus pada pemahaman pola anomali perilaku dan konsentrasi entitas yang relevan untuk kebutuhan triase keamanan. Oleh karena itu, studi ini melengkapi literatur yang ada dengan menyediakan perspektif praktis yang lebih dekat dengan implementasi di lingkungan organisasi nyata. Sintesis perbedaan studi ini dengan studi sebelumnya dapat dilihat pada Tabel 2.

Tabel 2. Sintesis Perbandingan Penelitian

Aspek	**Penelitian Sebelumnya**	**Penelitian Anda**
Sumber Data	Umumnya menggunakan dataset publik (KDD, UNSW-NB15, CICIDS) atau data sintetis	Menggunakan log firewall/web filtering dari konteks operasional organisasi nyata
Jenis Data	NetFlow, packet capture, atau flow-based data	Log aplikasi (firewall & web filtering) yang semi-terstruktur
Pendekatan Analisis	Dominan: machine learning (supervised & deep learning)	Statistical profiling + entity-based analysis + behavioral interpretation
Ketergantungan Label	Banyak bergantung pada data berlabel (supervised learning)	Tidak bergantung pada label (unsupervised exploratory analysis)
Fokus Deteksi	Deteksi serangan (DDoS, intrusion, malware)	Deteksi anomali perilaku (behavior anomaly), konsentrasi (concentration anomaly), dan policy-driven risk
Granularitas Analisis	Umumnya global (traffic-level atau flow-level)	Entity-level (srcip, hostname, kategori, relasi antar entitas)
Dimensi Waktu	Banyak menggunakan time-series	Snapshot analysis (±5 menit) dengan

E-ISSN: 2623-1700
P-ISSN: 1693-3672

Aspek	Penelitian Sebelumnya	Penelitian Anda
	dan temporal modeling	insight eksploratif
Kompleksitas Model	Tinggi (deep learning, graph learning, federated learning)	Relatif sederhana tetapi interpretatif dan operasional
Explainability	Cenderung rendah (black-box models)	Tinggi (interpretasi langsung berbasis pola nyata)
Tujuan Utama	Akurasi deteksi dan klasifikasi serangan	Pemahaman pola anomali dan implikasi operasional keamanan
Konteks Implementasi	Eksperimental / akademik	Konteks operasional enterprise (real-world applicability)
Kontribusi Utama	Pengembangan model deteksi baru	Insight empiris + framework interpretasi anomali berbasis log firewall
Keterbatasan	Kurang representatif terhadap kondisi nyata organisasi	Tidak memiliki baseline jangka panjang dan tidak menggunakan multi-source log
Implikasi Praktis	Sulit langsung diimplementasikan tanpa tuning	Langsung dapat digunakan untuk triase keamanan dan investigasi awal

Meskipun berbagai penelitian telah mengembangkan pendekatan deteksi anomali berbasis machine learning dan *deep learning*, sebagian besar studi masih bergantung pada dataset publik atau data sintetis yang belum tentu merepresentasikan kompleksitas infrastruktur *enterprise*. Selain itu, penelitian sebelumnya cenderung berfokus pada peningkatan akurasi model deteksi, dengan penekanan pada pendekatan supervised atau deep learning, sehingga kurang memberikan perhatian pada interpretasi operasional dari hasil analisis.

Lebih lanjut, sebagian besar studi melakukan analisis pada level trafik atau flow, tanpa mengeksplorasi pola anomali berbasis entitas seperti relasi antara source IP, hostname, dan kategori konten. Pendekatan yang ada juga umumnya menganggap anomali sebagai outlier atau serangan, tanpa membedakan dimensi

E-ISSN: 2623-1700
P-ISSN: 1693-3672

anomali seperti perilaku pengguna, konsentrasi aktivitas, dan eksposur risiko berbasis kebijakan.

Selain itu, masih terdapat keterbatasan dalam aspek explainability, di mana banyak model yang bersifat black-box sehingga sulit diterjemahkan ke dalam konteks operasional keamanan. Penelitian sebelumnya juga jarang mengaitkan hasil deteksi dengan kebijakan filtering yang aktif, serta belum mengeksplorasi potensi insight dari data snapshot jangka pendek.

Oleh karena itu, diperlukan penelitian yang tidak hanya mampu mendeteksi anomali, tetapi juga memberikan interpretasi yang kontekstual, berbasis data nyata, dan relevan untuk kebutuhan operasional keamanan jaringan.

METODE

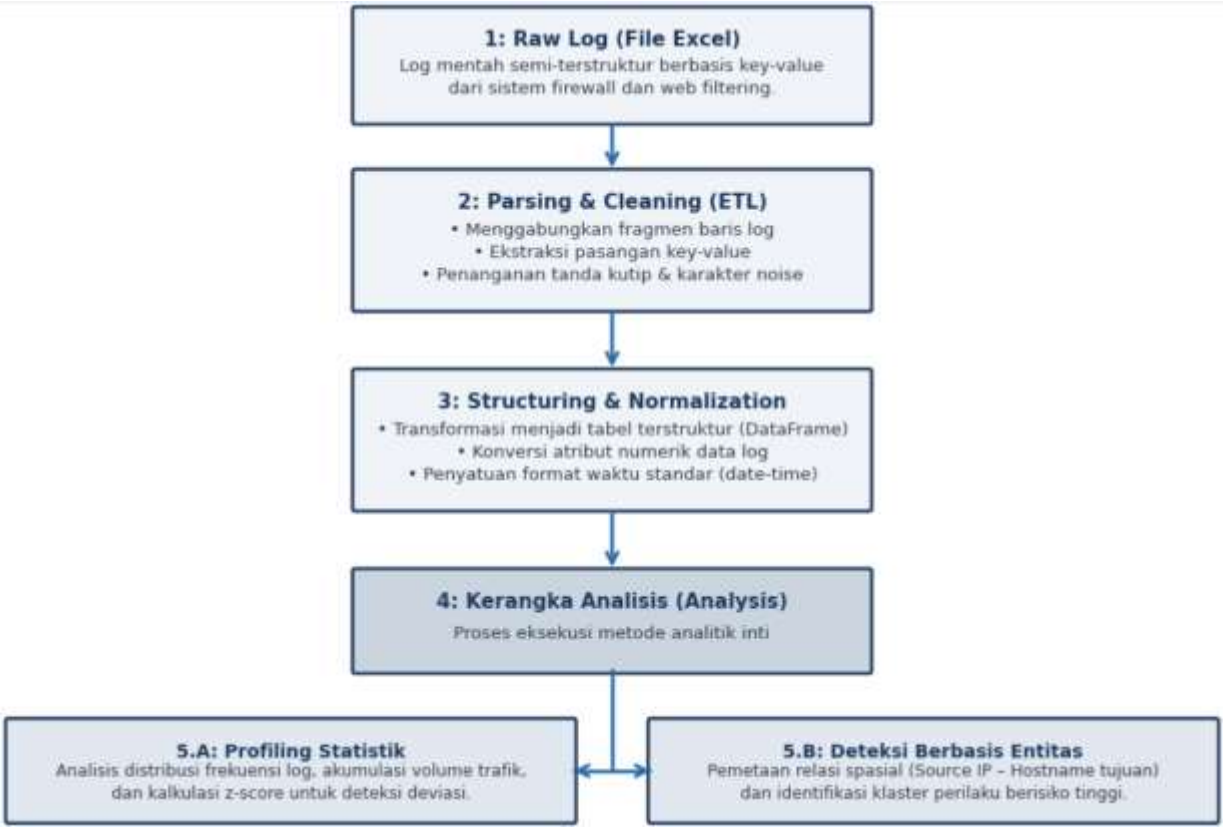
3.1 Data dan prapemrosesan

Data sumber berasal dari file Excel berisi log firewall/web filtering. Secara awal, struktur data tidak siap analisis karena setiap baris pada dasarnya merupakan string log mentah bertipe key-value yang terpecah ke banyak kolom. Dataset yang digunakan dipilih berdasarkan kriteria: (1) berasal dari infrastruktur *enterprise*, (2) mengandung atribut web filtering yang lengkap (kategori, hostname, action), (3) memiliki granularitas waktu hingga level detik, dan (4) tidak mengalami anonymization yang menghilangkan relasi antar entitas. Meskipun dataset bersifat snapshot, data tetap dianggap representatif untuk analisis pola anomali awal (initial anomaly triage).

Selanjutnya dilakukan proses *Extract*, *Transform*, Load (ETL) sebagai berikut: (1) *menggabungkan* fragmen per baris menjadi satu string log utuh, (2) mengekstrak pasangan key-value, termasuk nilai berformat kutip, (3) membentuk tabel terstruktur, (4) mengonversi atribut numerik, dan (5) membangun atribut waktu terpadu dari date dan time. Parsing, normalisasi, dan ekstraksi fitur dilakukan untuk memastikan log semi-terstruktur dapat dianalisis secara konsisten. Hal ini dikarenakan kualitas representasi data sangat memengaruhi performa deteksi [3], [4], [9].

Dengan demikian alur ETL dapat diurut yaitu: Raw Log, Parsing, Structuring, Cleaning, dan Analysis yang sejalan dengan tahapan preprocessing log dalam literatur anomaly detection [3], [4] dapat dilihat pada Gambar 1.

E-ISSN: 2623-1700
P-ISSN: 1693-3672



Gambar 1. Diagram Arsitektur Sistem

3.2 Desain analisis

Penelitian ini menggunakan dua kerangka utama analisis: (1) Profiling Statistik, untuk mengidentifikasi distribusi, konsentrasi, dan deviasi kuantitatif pada atribut utama log; (2) Deteksi Berbasis Entitas, untuk mengevaluasi pola perilaku spesifik melalui relasi antara source IP, hostname, dan kategori trafik.

Pendekatan *entity-centric* studi ini relevan dengan arah riset mutakhir yang menekankan bahwa anomali tidak selalu muncul sebagai lonjakan trafik global, melainkan sering terlihat sebagai deviasi perilaku pada host, flow, atau relasi antarnode tertentu [5]–[8].

Untuk memastikan reliabilitas proses ekstraksi dan parsing log, dilakukan validasi struktur data dengan cara: (1) pengecekan konsistensi jumlah event sebelum dan sesudah ETL, (2) verifikasi atribut kunci seperti srcip, hostname, action, dan category terhadap data mentah, serta (3) pengujian parsing terhadap sampel acak log

E-ISSN: 2623-1700
P-ISSN: 1693-3672

untuk memastikan tidak terjadi kehilangan atau distorsi informasi.

Selain itu, proses transformasi dilakukan secara deterministik menggunakan aturan parsing yang konsisten, sehingga hasil analisis dapat direplikasi.

3.3 Batasan penelitian

Studi ini hanya menggunakan satu sumber log, yaitu log firewall/web filtering, dan rentang waktunya sangat singkat. Karena itu, hasil penelitian lebih tepat dibaca sebagai analisis anomali snapshot daripada baseline perilaku jangka panjang. Literatur firewall-log anomaly detection juga mengingatkan bahwa konstruksi fitur, granularitas agregasi, dan ketersediaan label sangat menentukan hasil, terutama bila analisis dilakukan hanya dari firewall logs [4]. Oleh karena itu, studi ini secara logis menempatkan temuan sebagai indikator kuat untuk investigasi lanjutan, bukan sebagai bukti final insiden kompromi.

HASIL

4.1 Profil umum dataset

Setelah proses cleaning, dataset akhir berisi 9.092 event dan 47 kolom terstruktur. Rentang waktu event adalah 12 Maret 2026 pukul 09:36:14 hingga 09:41:13, sehingga observasi efektif mencakup sekitar 5 menit. Dalam periode singkat ini teridentifikasi 182 source IP dan 186 hostname tujuan. Ringkasan terkait dataset dapat dilihat pada Tabel 3.

Distribusi aksi menunjukkan: (a) blocked: 8.877 event (97,64%), (b) passthrough: 215 event (2,36%). Komposisi layanan: (a) HTTPS: 8.856 event, (b) HTTP: 236 event. Berdasarkan aspek sisi negara tujuan, trafik terutama mengarah ke: (a) Singapore: 3.975 event, (b) Indonesia: 3.933 event, (c) United States: 706 event.

Temuan ini menunjukkan bahwa dataset merekam trafik web outbound yang sangat terkonsentrasi pada koneksi HTTP/HTTPS dan sangat didominasi oleh event yang diblokir kebijakan.

Tabel 3. Ringkasan Dataset

Parameter	Nilai
Total Event	9092
Source IP Unik	182
Hostname Unik	186
Dominan Action	Blocked (97.64%)
Rentang Waktu	±5 menit

E-ISSN: 2623-1700
P-ISSN: 1693-3672

4.2 Dominasi kategori konten

Kategori konten terbesar adalah: (a) Games: 4.231 event (46,54%), (b) Advertising: 3.948 event (43,42%), (c) Potentially Unwanted Program: 245 event (2,69%), (d) Unrated: 201 event (2,21%), (e) Brokerage and Trading: 185 event (2,03%), (f) Proxy Avoidance: 139 event (1,53%), (g) Malicious Websites: 97 event (1,07%), (h) Spam URLs: 28 event (0,31%), (i) Cryptocurrency: 14 event (0,15%).

Hampir 90% trafik pada *snapshot* terkonsentrasi pada kategori Games dan Advertising. Hal ini menunjukkan bahwa anomali dominan dalam dataset bukanlah variasi tipis yang tersebar merata, tetapi pola perilaku yang sangat terfokus.

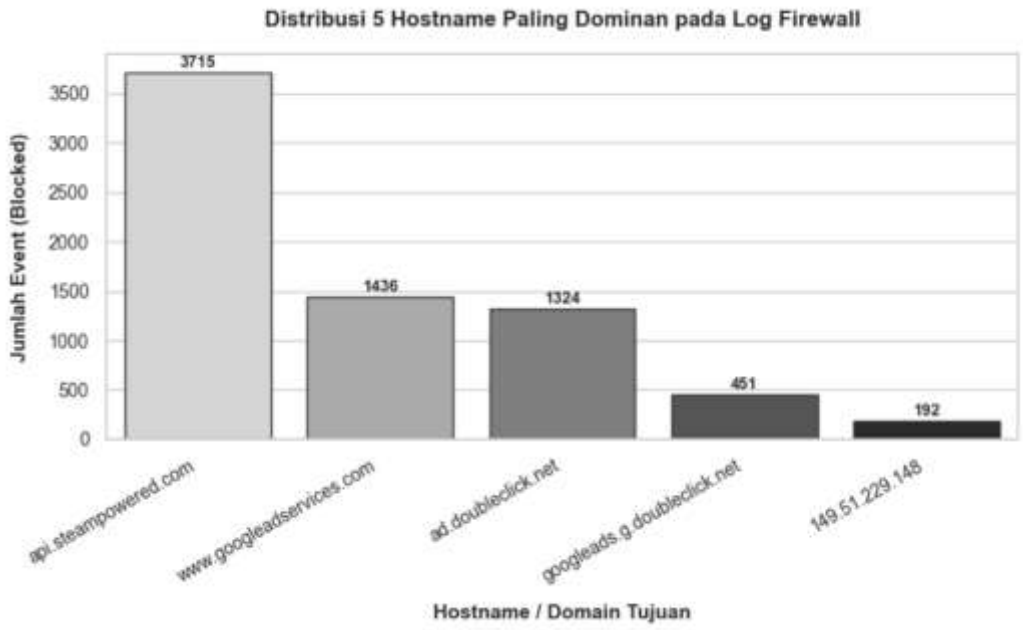
4.3 Dominasi hostname

Hostname paling dominan adalah: (a) api.steampowered.com sebanyak 3.715 event, (b) www.googleadservices.com sebanyak 1.436 event, (c) ad.doubleclick.net sebanyak 1.324 event, (d) googleads.g.doubleclick.net sebanyak 451 event, (e) 149.51.229.148 sebanyak 192 event. Tabel 4 menunjukkan top hostname (domain). Secara proporsional, host teratas menyumbang 40,86% dari seluruh event, dan lima (5) host teratas menyumbang 78,29% dari seluruh event. Angka tersebut menunjukkan konsentrasi tujuan yang sangat tinggi. Dalam pembacaan anomali, pola semacam ini biasanya layak diprioritaskan karena mempersempit himpunan entitas yang perlu diselidiki.

Tabel 4. Top Hostname (Domain)

Rank	Hostname	Event	%
1	api.steampowered.com	3715	40.86%
2	www.googleadservices.com	1436	15.80%
3	ad.doubleclick.net	1324	14.56%
4	googleads.g.doubleclick.net	451	4.96%
5	149.51.229.148	192	2.11%

E-ISSN: 2623-1700
P-ISSN: 1693-3672



Gambar 2. Grafik Top Domain

4.4 Dominasi source IP

Source IP paling dominan adalah: 10.17.120.175 sebanyak 2.684 event, 10.17.31.11 sebanyak 1.052 event, 10.17.30.47 sebanyak 965 event, 10.17.29.18 sebanyak 801 event, dan 10.17.120.178 sebanyak 626 event. Tabel 5 menunjukkan top source IP. Secara proporsional, IP teratas menyumbang 29,52% dari seluruh event, dan lima (5) IP teratas menyumbang 67,40% dari seluruh event.

Jika jumlah event per source IP dipandang sebagai distribusi, maka 10.17.120.175 merupakan outlier yang sangat kuat. Nilai z-score jumlah event untuk IP ini berada sekitar 11,11, jauh di atas mayoritas entitas lain. Pada level operasional, ini cukup untuk menempatkan IP tersebut sebagai kandidat investigasi prioritas.

Tabel 5. Top Source IP

Source IP	Jumlah
10.17.120.175	2684
10.17.31.11	1052
10.17.30.47	965
10.17.29.18	801
10.17.120.178	626

E-ISSN: 2623-1700
P-ISSN: 1693-3672

4.5 Pola perilaku per entitas

Analisis relasi srcip–hostname memperlihatkan klaster perilaku yang sangat jelas, yaitu: (a) IP 10.17.120.175 hampir sepenuhnya terkait domain iklan, yaitu: www.googleadservices.com sebanyak 1.408, ad.doubleclick.net sebanyak 1.276. Hal tersebut menunjukkan pola advertising-heavy yang sangat konsisten, (b) IP 10.17.31.11, 10.17.30.47, 10.17.29.18, 10.17.120.178, dan 10.17.30.11 sangat didominasi oleh api.steampowered.com di mana hal tersebut membentuk klaster gaming-related yang kuat, bukan noise acak. (c) IP 172.17.2.252 menampilkan kombinasi yang terdiri dari: 149.51.229.148 sebanyak 192, pushstream.tradingview.com sebanyak 64, notifications.tradingview.com sebanyak 53, data.tradingview.com sebanyak 34, telemetry.tradingview.com sebanyak 34. Kategori pada IP terebut terutama Potentially Unwanted Program dan Brokerage and Trading. (d) IP 10.17.110.18 dominan pada: c.csroutear.com sebanyak 28, socket-backup.earn.fm sebanyak 21, socket-prod.earn.fm sebanyak 20, di mana kategori utama adalah Proxy Avoidance dan Spam URLs. (e) IP 10.16.108.31 seluruh 38 event-nya dipetakan ke Malicious Websites, terutama v3.tiktokcdn.com. Temuan ini perlu dibaca hati-hati karena label kategori berasal dari mesin kategorisasi vendor, sehingga ini lebih tepat diperlakukan sebagai indikator kebijakan yang perlu diverifikasi ulang, bukan vonis kompromi.

4.6 Pola waktu

Walau seluruh data berada pada jam 09, distribusi per detik tetap memperlihatkan burst kecil. Dari 293 detik aktif, rata-rata volume adalah sekitar 31,03 event/detik, dengan maksimum 72 event/detik. Karena tidak tersedia baseline jam-jam lain, burst ini belum dapat disebut serangan volumetrik; namun ia menegaskan bahwa snapshot merekam aktivitas padat dan simultan pada beberapa klaster perilaku.

DISKUSI

Pada bagian diskusi akan dibahas jenis anomali yang muncul, interpretasi keamanan, posisi temuan terhadap literatur, kesimpulan substantif dari data, dan peluang penelitian selanjutnya.

5.1 Jenis anomali yang muncul

Berdasarkan data, anomali utama dalam snapshot ini dapat dibagi menjadi tiga jenis. Pertama, concentration anomaly. Sebagian

E-ISSN: 2623-1700
P-ISSN: 1693-3672

kecil IP dan host menyumbang mayoritas event. Pola ini menandakan ketidakseimbangan yang cukup ekstrem untuk layak dianggap sebagai deviasi operasional, khususnya ketika satu IP mendominasi hampir sepertiga event. Kedua, behavior anomaly. Klaster gaming dan advertising muncul sangat kuat. Pada konteks enterprise, trafik seperti ini tidak otomatis berarti serangan, tetapi dapat menunjukkan aktivitas non-produktif, perangkat dengan aplikasi latar yang agresif, atau endpoint yang menjalankan software dengan profil komunikasi yang tidak selaras dengan fungsi bisnis. Ketiga, policy-driven risk exposure. Kategori seperti Potentially Unwanted Program, Proxy Avoidance, Malicious Websites, Spam URLs, dan Cryptocurrency menunjukkan bahwa kebijakan firewall telah bersinggungan dengan trafik berisiko. Namun, karena data ini adalah log web filtering, label kategori harus dibaca sebagai sinyal triase, bukan bukti final insiden.

5.2 Interpretasi keamanan

Kebutuhan akan interpretasi hasil yang dapat dijelaskan (*explainability*) menjadi semakin penting dalam sistem deteksi anomali modern [9]. Temuan paling kuat dari snapshot ini adalah bahwa organisasi menghadapi dua lapis masalah yang berbeda. Lapisan temuan pertama menunjukkan dominasi trafik non-bisnis. Dominasi api.steampowered.com, title.mgt.xboxlive.com, dan hostname sejenis mengarah pada kemungkinan aktivitas gaming atau aplikasi terkait gaming pada jam kerja. Secara keamanan, ini relevan bukan semata karena produktivitas, tetapi karena endpoint yang menjalankan software hiburan atau launcher kerap membawa permukaan serangan tambahan, telemetri agresif, dan pola update yang berisik. Lapisan kedua adalah paparan trafik berisiko rendah hingga sedang yang tersebar pada kategori PUP, proxy avoidance, spam, unrated, dan cryptocurrency. Di sini, 172.17.2.252 dan 10.17.110.18 menjadi penting. Pada 172.17.2.252, keberadaan TradingView dan host 149.51.229.148 menunjukkan perilaku yang tidak sejalan dengan pola web bisnis umum, meski belum cukup untuk menyimpulkan malicious. Pada 10.17.110.18, kombinasi earn.fm dan Proxy Avoidance lebih layak diprioritaskan untuk pengecekan endpoint, ekstensi browser, atau aplikasi pihak ketiga.

E-ISSN: 2623-1700
P-ISSN: 1693-3672

5.3 Posisi temuan terhadap literatur

Hasil penelitian ini memperkuat paradigma bahwa anomali jaringan tidak selalu memanifestasikan diri sebagai lonjakan volume trafik global yang masif, melainkan sering kali muncul sebagai deviasi perilaku mikroskopis pada entitas atau relasi antarnode tertentu [5]–[8]. Pendekatan entity-centric yang diterapkan dalam studi ini terbukti sensitif dalam mengisolasi klaster aktivitas menyimpang tanpa bergantung pada ketersediaan label data (unsupervised exploratory analysis) [3], [4]. Hal ini membedakannya dari mayoritas literatur berbasis supervised learning—seperti pengujian algoritma pada dataset NetFlow—yang membutuhkan data berlabel rapi untuk mencapai akurasi klasifikasi serangan [5].

Meskipun demikian, keterbatasan analisis berbasis snapshot berdurasi pendek dalam studi ini melahirkan blind spot terhadap dinamika temporal jangka panjang. Untuk mengatasi hal tersebut, riset mutakhir di bidang keamanan siber sangat mendorong integrasi model spatiotemporal berbasis mekanisme attention yang mampu menangkap pola hubungan ruang dan waktu secara simultan [6]. Lebih lanjut, penerapan metode mutakhir seperti graph representation learning dan federated learning juga memegang peranan krusial dalam memetakan interaksi relasional yang kompleks serta terdistribusi pada infrastruktur enterprise modern [8]. Selain aspek akurasi deteksi, penguatan metodologi ke depan juga harus mempertimbangkan dimensi explainability (Explainable AI/XAI), sehingga output dari model analitik kompleks tidak lagi bersifat black-box, melainkan dapat diinterpretasikan secara langsung dan transparan oleh tim operasi keamanan dalam melakukan triase risiko insiden [9], [10].

5.4 Kesimpulan substantif dari data

Berdasarkan kajian menyeluruh terhadap data yang diunggah, terdapat beberapa hal yang dapat disimpulkan. Pertama, anomali paling dominan bukan serangan DDoS atau scanning masif, melainkan anomali perilaku dan konsentrasi. Snapshot didominasi oleh traffic gaming dan advertising dengan intensitas tinggi pada sedikit entitas. Kedua, terdapat endpoint prioritas investigasi. Secara khusus bahwa: 10.17.120.175 untuk pola advertising-heavy, 10.17.31.11, 10.17.30.47, 10.17.29.18, 10.17.120.178,

E-ISSN: 2623-1700
P-ISSN: 1693-3672

10.17.30.11 untuk pola gaming-heavy, 172.17.2.252 dan 10.17.110.18 untuk pola berisiko lebih tinggi. Ketiga, firewall policy bekerja aktif. Proporsi blocked yang sangat tinggi menunjukkan kebijakan filtering berjalan dan berhasil menahan mayoritas trafik bermasalah atau tidak diinginkan. Keempat, masih terdapat ruang blind spot. Keberadaan passthrough pada kategori Unrated dan Cryptocurrency memperlihatkan adanya trafik yang lolos kebijakan walaupun layak dipantau lebih dekat.

5.5 Peluang penelitian selanjutnya

Dari perspektif riset, data ini membuka beberapa jalur lanjutan yang nyata. Pertama, baseline longitudinal. Karena snapshot hanya lima menit, penelitian berikutnya perlu mengumpulkan data minimal 7–30 hari untuk membentuk baseline per user, per perangkat, per jam, dan per kategori. Ini penting agar "ramai" dapat dibedakan dari "anomali". Kedua, korelasi lintas-sumber. Log firewall sebaiknya dipadukan dengan DHCP, asset inventory, proxy log, DNS, EDR, dan autentikasi. Dengan begitu, source IP dapat dipetakan ke pengguna dan perangkat aktual, serta klasifikasi anomali menjadi lebih tajam. Ketiga, model unsupervised dan semi-supervised. Untuk lingkungan yang tidak memiliki label insiden rapi, pendekatan seperti isolation forest, clustering, autoencoder, atau one-class learning akan relevan. Literatur terbaru juga menunjukkan bahwa representasi spatiotemporal, explainable ML, dan model graf/federated learning menjanjikan untuk skenario trafik yang kompleks dan terdistribusi [6]–[8]. Keempat, evaluasi kualitas parsing dan feature engineering. Karena penelitian berbasis log sangat sensitif terhadap tahap parsing, studi lanjutan dapat membandingkan skema feature construction dari raw log, flow aggregation, dan entity graph untuk melihat mana yang paling stabil dan informatif [3], [4]. Kelima, pemisahan antara policy violation dan security incident. Penelitian berikutnya sebaiknya membangun taksonomi dua tingkat: (a) non-compliance/non-business traffic, dan (b) security-relevant anomalous traffic. Dengan begitu, tim operasi dapat membedakan tiket disiplin kebijakan dari investigasi keamanan yang benar-benar mendesak.

KESIMPULAN

E-ISSN: 2623-1700
P-ISSN: 1693-3672

Kajian studi ini menunjukkan bahwa log firewall/web filtering yang diunggah pengguna mengandung sinyal anomali yang cukup kuat meskipun hanya berupa snapshot singkat. Dari 9.092 event, mayoritas besar diblokir dan terkonsentrasi pada kategori Games dan Advertising, dengan dominasi mencolok oleh beberapa source IP dan hostname. Temuan ini menandakan bahwa lanskap anomali pada data yang dikaji lebih dekat dengan ketimpangan perilaku, dominan entitas tertentu, dan eksposur trafik berisiko berbasis kebijakan daripada serangan jaringan volumetrik klasik.

Secara praktis, organisasi sebaiknya memprioritaskan investigasi pada endpoint yang menampilkan pola advertising-heavy, gaming-heavy, proxy avoidance, dan potentially unwanted program. Hasil penelitian menunjukkan bahwa analisis log firewall dapat menjadi dasar awal dalam proses triase keamanan, namun akurasi dan kedalaman interpretasi akan jauh meningkat bila studi berikutnya menggunakan data longitudinal, korelasi lintas-log, dan model analitik yang lebih adaptif.

DAFTAR PUSTAKA

[1] Moustafa, N., Hu, J., & Slay, J. A Holistic Review of Network Anomaly Detection Systems: A Comprehensive Survey. *Journal of Network and Computer Applications*, 128, 33–55, 2019.

[2] Abbasi, M., Shahraki, A., & Taherkordi, A. Deep Learning for Network Traffic Monitoring and Analysis (NTMA): A Survey. *Computer Communications*, 170, 19–41, 2021.

[3] Landauer, M., Onder, S., Skopik, F., & Wurzenberger, M. Deep Learning for Anomaly Detection in Log Data: A Survey. *Machine Learning with Applications*, 12, 100470, 2023.

[4] Komadina, A., Kovačević, I., Štengl, B., & Groš, S. Comparative Analysis of Anomaly Detection Approaches in Firewall Logs: Integrating Light-Weight Synthesis of Security Logs and Artificially Generated Attack Detection. *Sensors*, 24(8), 2636, 2024.

E-ISSN: 2623-1700
P-ISSN: 1693-3672

[5] Fosić, I., Žagar, D., Grgić, K., & Križanović, V. Anomaly Detection in NetFlow NetworkTraffic Using Supervised Machine Learning Algorithms. *Journal of Industrial Information Integration*, 33, 100466, 2023.

[6] Ji, C., Yu, H., & Dai, W. Network Traffic Anomaly Detection Based on Spatiotemporal Feature Extraction and Channel Attention. *Processes*, 12(7), 1418, 2024.

[7] Schummer, P., del Rio, A., Serrano, J., Jimenez, D., Sánchez, G., & Llorente, Á. Machine Learning-Based Network Anomaly Detection: Design, Implementation, and Evaluation. *AI*, 5(4), 2967–2983, 2024.

[8] Zhao, Y., Liu, Z., & Pang, J. Anomaly Detection in Network Traffic via Cross-Domain Federated Graph Representation Learning. *Applied Sciences*, 15(11), 6258, 2025.

[9] Ahmed, M., Mahmood, A. N., & Hu, J. Network anomaly detection: A survey and comparative analysis of recent advances. *IEEE Communications Surveys & Tutorials*, 2023.

[10] Nguyen, T. T., Nguyen, T. G., Sohel, F., & Al-Zahrani, A. Explainable AI for network anomaly detection: Recent advances and challenges. *Future Generation Computer Systems*, 2024.