NOD: Uncovering intense attackers' behavior through Nested Outlier Detection from SSH logs

Ghazal Abdollahi
University of Utah
ghazal.abdollahi@utah.edu

Hamid Asadi
University of Utah
hamid93asd@gmail.com

Robert Ricci
University of Utah
ricci@cs.utah.edu

Abstract—**Persistent, high-volume SSH brute-force activity frequently overwhelms security operations, yet current defenses often treat network telemetry as a terminal artifact for post-hoc diagnosis rather than a source for upstream investigation. These approaches focus on absolute volume suppression and binary alerts, often failing to provide population-aware rankings that are necessary to prioritize high-risk, relative outliers. This work addresses these gaps by introducing Nested Outlier Detection (NOD), a two-stage framework that transforms raw network telemetry into structured behavioral strata. By progressively filtering routine noise, NOD isolates "outliers of outliers"; statistically extreme behaviors. NOD provides interpretability by mapping these outliers to three intuitive dimensions; volume, reach, and credential diversity; enabling population-level reasoning. This tiered approach reveals distinct attacker phenotypes characterized by high volume, broad target reach, and a variety of credentials. Evaluation on large-scale datasets demonstrates that NOD compresses millions of logs into compact, interpretable structures, shifting the defensive focus from per-source classification to the graded, population-level reasoning required for scalable triage and longitudinal threat analysis.**

I. Introduction

Modern cybersecurity frameworks often operate under the assumption that threat severity correlates with attack volume and frequency. This view prioritizes the detection of high-frequency events and large-scale attacks while underrepresenting sophisticated adversaries that achieve strategic objectives through coordination and behavioral adaptation rather than computational scale. Contemporary threat actors increasingly employ distributed infrastructure and synchronized operations that generate minimal individual signatures while achieving significant collective impact through coordinated behavioral patterns. This shift underscores a broader evolution in networked threat dynamics, where shared infrastructure, programmability, and cross-layer dependencies enable adversaries to coordinate subtle yet systemic disruptions that elude traditional volume-based defenses [12].

While individual attack detection is effective for suppressing single sources, it provides limited support for after-the-fact investigation and reasoning. In environments where thousands of sources are simultaneously malicious, analysts need to understand how attackers behave as a population: how effort is distributed, how credentials are reused, and where coordination emerges. Examining individual logs or alerts in isolation obscures these patterns. This work enables investigators to extract population-level structure from raw logs, revealing meaningful attack behaviors without exhaustive per-source inspection.

Detection accuracy evaluates whether an alert fires correctly, but investigative value depends on whether alerts help explain attacker behavior. This work does not seek to improve per-source classification performance. Instead, it supports reasoning by organizing attackers into behaviorally coherent groups that expose differences in intensity, targeting strategy, and credential usage. The contribution lies in improving analysts' ability to interpret attack campaigns, not in triggering more or faster alerts. Conventional network security approaches focus on identifying individual malicious behaviors through rate-based monitoring, threshold analysis, and statistical outlier detection applied to discrete events. These methodologies excel at detecting isolated anomalies but offer limited visibility into the higher-order structure in which multiple actors employ synchronized strategies across distributed infrastructure. This strategy is particularly employed by existing SSH security frameworks, which focus mainly on single-source anomaly detection through rate-based blocking [4], [14], flow-based analysis [5], [10], and machine learning approaches [11], [18], leaving multi-actor relationships weakly characterized. Although honeypot studies [16], [22] capture extensive attack data revealing geographic clustering, quantitative tools for assessing coordination strength and separating coincidence from structured alignment remain under-specified in practice. This gap matters, particularly when sophisticated threats operate within nominal anomaly ranges to remain indistinguishable from background variability.

Our goal with this work is to provide a basis for threat intelligence that is based on methods for detecting likely coordination between attackers: that is, sets of attackers that have similar patterns that are unlikely to be mere coincidence. Although temporal outlier methods improve recognition of individual deviations [8], [18], they rarely elevate per-source deviations into reproducible behavioral relationships that inform reasoning about shared tactics or adaptive coordination. Consequently, analysis remains descriptive at the event level rather than articulating a population-level structure that sup-

Workshop on Attack Provenance, Reasoning, and Investigation for Security in the Monitored Environment (PRISM) 2026
23 February 2026, San Diego, CA, USA
ISBN 978-1-970672-00-8
https://dx.doi.org/10.14722/prism.2026.23008
www.ndss-symposium.org

ports forward-looking reasoning.

Singh et al. [22] conducted a four-year longitudinal analysis of SSH attacks and developed Dictionary-Based Blocking, achieving 99.5% attack coverage by using username dictionaries to identify and block attackers. Their work established that attackers exhibit recognizable and repeatable organization, raising practical questions about the measurable structure of such coordination: how to quantify it reproducibly, how its intensity varies across sources, and how its signatures evolve over time to support proactive reasoning rather than reactive classification.

In modern SSH deployments, brute-force activity is persistent rather than exceptional, producing a steady stream of alerts that rarely convey investigative meaning. The central challenge for defenders is therefore not detecting whether attacks exist, but understanding which behaviors reflect coordinated activity worth investigation, and which represent routine background noise. This work addresses that gap by restructuring large-scale SSH telemetry into population-level strata that surface rare, behaviorally extreme patterns that merit focused analysis. Our work contributes:

- **Structural Abstraction of Log Streams:** We propose a methodology to transform high-cardinality **raw logs** into high-fidelity **network telemetry**. By aggregating millions of events into compact, IP-level behavioral signatures, we enable tractable, population-level reasoning that remains performant at the scale of modern cloud infrastructure.
- **Nested Outlier Detection (NOD) Framework:** We introduce a two-stage analytical pipeline that isolates "*outliers of outliers*" within a multivariate feature space. This provides a graded metric of **statistical extremeness**, allowing defenders to distinguish between routine automated activity and highly deviant, coordinated behaviors without the fragility of binary thresholds.
- **Incidence-Based Coordination Discovery:** We develop a **multi-entity incidence representation** that encodes the structural relationships between source nodes, target reach, and credential usage, enabling the discovery of latent coordination and attacker phenotypes, such as distributed credential-stuffing and coordinated infrastructure swarms, transforming isolated telemetry points into a structured map of attackers' behavior evolution.

II. Related work

Singh et al. [22] present a comprehensive longitudinal study of SSH brute-force attacks (BFAs) to date, analyzing four years of sshd logs across 500 production servers in CloudLab. Their analysis highlights evolving attacker behavior, including shifts in targeted usernames, persistence across IP addresses, and the global distribution of compromised hosts. Crucially, they introduce Dictionary-Based Blocking (DBB), a defense that exploits attacker username dictionaries to achieve high coverage with low false positives. Earlier studies had a narrower scope: Owens and Matthews [19] examined 103K login attempts over three honeypots, showing administrator accounts dominate guesses, while Abdou et al. [1] confirmed the predominance of "*root*" and "admin" in guessing vectors. Work such as Ghiette et al. [7] has fingerprinted attacker tools (e.g., Ncrack [6], Hydra [23]) via SSH version strings and cipher negotiation, whereas honeypot deployments at scale [2], [13], [16] revealed how geography, hosting providers, and interaction fidelity influence attacker strategies. Collectively, these studies demonstrate that SSH remains a primary vector, despite best-practice recommendations for key-based authentication.

Traditional host-based tools (e.g., Fail2ban [14], Denyhosts [21], SSHGuard [4]) implement rate-based blocking, often at the cost of high false positives when legitimate users mistype credentials. Singh et al. demonstrate that such approaches can miss the majority of attacks. Alternatives include flow-based intrusion detection systems, such as SSHCure by Hellemons et al. [10], and NetFlow/IPFIX-based compromise detection by Hofstede et al. [5], which identify anomalous connection patterns. While effective at scale, these methods may still misclassify benign traffic spikes. More recent efforts explore ML and deep learning: Najafabadi et al. [18] and Hancock et al. [9] used aggregated NetFlow with classifiers to detect SSH/FTP BFAs, while Hossain et al. [11] applied LSTMs for sequence-based detection. These learning-based techniques enhance adaptability but often lack interpretability, which raises deployment challenges in production systems.

Prior SSH defenses and detection systems are evaluated based on alerting performance or blocking effectiveness for individual sources. In contrast, this work focuses on the structure of evidence for investigation. The framework operates upstream of enforcement by transforming raw telemetry into interpretable population-level representations that support reasoning about campaign structure. Because the output is not a detection decision but an investigative abstraction, direct apples-to-apples comparison with supervised or threshold-based detectors does not capture the contribution of this work.

The persistent challenge of alert overload and analyst fatigue stems from treating security events as terminal artifacts rather than reorganizable telemetry. While prior studies provide empirical evidence of the scale of this problem [24], they focus primarily on post-hoc diagnosis rather than structuring data upstream to support active investigation. Consequently, we lack a systematic way to rank behaviors that are rare in a relative sense, leaving the problem of transforming irregular alerts into actionable, population-level evidence largely unsolved.

Furthermore, current operational defenses are often limited by a focus on absolute volume suppression rather than longitudinal behavioral characterization. Systems like *CAUDIT* demonstrate effective automated blocking [3], yet they rely on absolute indicators that fail to provide a normalized view of how specific actors compare to the broader distribution. This creates a critical gap where low-frequency, high-risk behaviors, such as adaptive campaigns that stay just below blocking thresholds, blend into background noise, offering little support for prioritizing threats or reasoning about evolving attacker strategies.

Finally, even sophisticated frameworks designed to detect stealthy activity often prioritize binary detection over comparative reasoning. For example, while change-point analysis can identify stealthy epochs [15], it treats hosts primarily as discrete participants or non-participants rather than exposing a graded measure of behavioral extremeness. Without a method to organize detected activity into a relative risk hierarchy, defenders remain stuck with "*all-or-nothing*" alarms that lack the statistical structure needed for scalable interpretation, triage, and adaptive response in high-volume environments.

While longitudinal studies such as Munteanu et al. [17] establish a vital baseline for attacker intent, their methodology relies on interactive honeypots, a data source that is increasingly compromised by adversarial awareness. As the authors themselves observe, attackers now actively identify, evade, and even abuse honeypot infrastructure as proxies, rendering the resulting logs a potentially biased and polluted subset of true adversarial behavior. Furthermore, by treating these logs as terminal artifacts for post-hoc command classification, such research remains descriptive rather than structural. This focus on individual session interactions fails to capture the multi-entity relationships between source infrastructure, target reach, and the credential namespace. Consequently, while we understand what attackers do once they are trapped, a fundamental gap remains in transforming raw logs into a high-fidelity telemetry representation that can surface the population-scale coordination driving modern attack campaigns.

To address these gaps, we propose a two-stage analytical framework that transforms raw telemetry into a structured, population-level view of threat activity. By employing a nested outlier detection approach, the system isolates intense behaviors that translate into statistical outliers within a multi-dimensional feature space. This method moves beyond binary detection by progressively filtering routine noise to uncover "*outliers of outliers*", the most extreme and improbable behaviors that signal sophisticated or coordinated campaigns. Ultimately, this approach organizes massive event streams into compact, interpretable structures, providing the graded evidence necessary to prioritize analyst attention and support systematic reasoning under continuous attack conditions.

III. Dataset

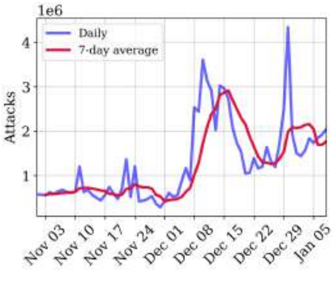


Fig. 1: Daily events over time. This figure shows the distribution of events across different times of day.

Our analysis is based on SSH authentication logs that originate from an anonymized research facility in North America. The raw dataset contains detailed records of SSH connection attempts, both successful and failed, captured across the network infrastructure. A representative sample of the raw data reveals the diversity of connection attempts across various source IP addresses, target nodes, and attempted usernames, including failed authentication attempts, which also highlights the diversity of connection attempts. To prepare the dataset for analysis, we applied several preprocessing steps. Each source IP address was mapped to its corresponding country using the GeoLite2-Country.mmdb database [20], providing valuable context for understanding the global distribution of SSH connection attempts. Following domain expertise recommendations, we removed descriptive fields (Tag, Message) to focus on the core behavioral and structural features relevant to anomaly detection. As Figure 1 shows, the dataset includes 68 days of SSH attack data collected from November 1, 2022, to January 8, 2023. Daily attacks fluctuate with sharp spikes; the 7-day average shows a pronounced surge in mid-December and a higher plateau into early January. This scale and diversity provide a rich foundation for behavioral analysis, capturing a wide range of connection patterns, geographic origins, and targeting strategies in real-world SSH traffic. Table I shows raw dataset features and post-processing characteristics.

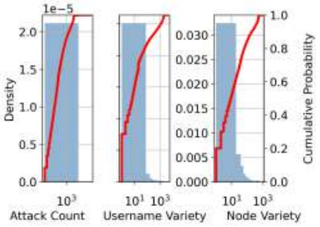


Fig. 2: Cumulative Probability. This figure illustrates the cumulative probability curves for the selected datasets.

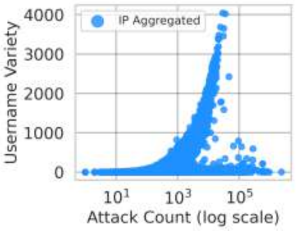


Fig. 3: Dataset aggregated by attacker source IP; two of ten dimensions shown.

Figure 2 shows the marginal distributions of attack count, username diversity, and node diversity on a logarithmic scale to span their wide ranges. Blue histograms depict probability density, while the red curves plot the empirical CDFs. The pairing makes the tail behavior explicit: densities are strongly right-skewed, and the CDFs rise steeply at small values (most observations are low) before flattening into long, shallow tails (a non-trivial fraction occupies much larger magnitudes). In practical terms, the CDF view clarifies percentiles; for example, the median lies near the lower end, whereas upper quantiles extend across orders of magnitude, highlighting both structural asymmetry and the presence of extreme outliers in all three variables.

IV. IP-based Aggregation

To enable population-level analysis at scale, we aggregate all attacks from the same source IP address into discrete behavioral profiles, with each source address corresponding to a single attacker. This architectural shift from event-level analysis to stable IP-level profiles is essential: conventional clustering methods (K-means, DBSCAN) fail on raw SSH telemetry because extreme volume variance causes density-based algorithms to group sources by connection frequency

TABLE I: Original Dataset Structure with Post-Processing Statistics

Feature	Description	Dataset Characteristics After Preprocessing	
		Attribute	Count
IP	Source IP address		
Date	Connection date		
Time	Connection hour	Unique Source IPs	77,886
Node	Target node		
Port	Source port number	Unique Usernames	88,482
PID	SSH daemon process ID		
Username	Attempted username	Unique Target Nodes	1,077
Tag	Connection status		
Message	Detailed log message	Unique Countries	221

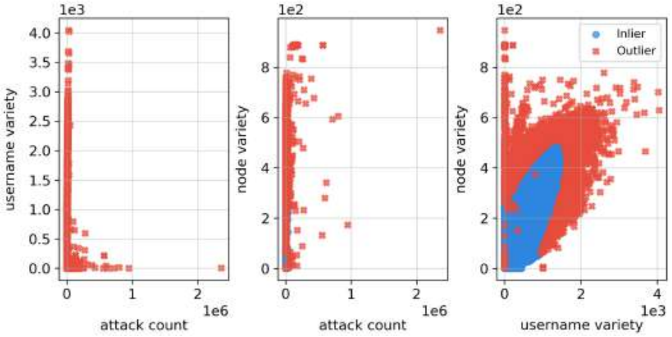


Fig. 4: Mahalanobis Tier1

rather than tactical similarity, obscuring sophisticated low-frequency campaigns within high-volume brute-force clusters.

Figure 3 illustrates the IP-level aggregation of SSH activity, showing two key dimensions: attack count and username variety. Each point represents a single IP summarized over the 68-day observation period. The plot reveals a strong positive association: higher-volume attackers typically probe a broader set of usernames, consistent with credential-sweeping behavior. A dense mid-range band ($\approx 10^3$–10^5 attempts) suggests coordinated scanners with sustained diversity, while low-volume IPs remain narrow in scope. This compresses 89M logs into 77,886 behavioral profiles suitable for an efficient statistical analysis.

V. Finding Attacker Coordination via Outlier Detection

We employ Mahalanobis distance for its simplicity and interpretability at scale. Applying this measure in a tiered manner enables progressive separation of routine activity from increasingly extreme behavior while preserving a clear rationale at each stage. Accordingly, the contribution lies in selecting and organizing a method that meets these practical requirements for population-level analysis, rather than in advancing the underlying distance formulation itself.

Single-tier detection captures broad anomalies but cannot distinguish between moderately unusual behavior and genuine coordination signatures. Two-tier analysis enables progressive refinement: isolating anomalies from routine activity, then identifying extreme coordination patterns within those anomalies, revealing that outliers themselves contain hierarchical structure. Extreme outliers, in particular, occupy the statistical frontier of highly improbable behavior, indicating behavior that may be designed to optimize attack power, evade defenses, or both. Additionally, the extreme outliers are few enough in number to permit manual inspection and to apply more targeted analyses that may not scale to the entire dataset.

As Table II shows, our two-tier methodology, applied to the 77,886 IP-aggregated dataset, identified 5,186 first-tier outliers (6.7% of the aggregated records), followed by 72 extreme outliers in the second tier. These 72 cases represent $8.08 \times 10^{-5}\%$ of the original 89,096,493 SSH connection attempts, occurring at approximately one extreme outlier per 1.24 million connection events across the 68-day observation period. The 99% Mahalanobis threshold establishes a statistical boundary corresponding to the 99th percentile of the chi-squared distribution in the three-dimensional feature space of attack count, username variety, and node variety. This threshold effectively identifies the top 1% most statistically improbable behavioral profiles, where ”improbable” refers to combinations of features that deviate significantly from the multivariate normal pattern exhibited by typical attackers. Unlike univariate approaches that examine features independently, the 99% Mahalanobis criterion considers the correlation structure between features, flagging cases where the joint probability of observing such feature combinations falls below the 1% threshold. We employ Mahalanobis distance to detect coordinated attacks that evade univariate thresholds. Unlike traditional detectors that ignore feature interdependencies, the Mahalanobis metric leverages the covariance matrix to identify anomalous multidimensional correlations in heterogeneous telemetry. This enables the detection of low-volume, sophisticated campaigns that appear benign when analyzed independently. Our nested multivariate framework replaces brittle, threshold-based heuristics with a mathematically principled approach for extracting actionable threat signatures from high-dimensional data.

Formal notation of the tier structure: Let $\mathbf{D}$ denote the full SSH dataset. From this set, we extract the *Tier-1 outliers* $O_1 \subset \mathbf{D}$. Within this layer, we further identify the *Tier-2 outliers* $O_2 \subset O_1$, which are therefore also contained in $\mathbf{D}$. The elements of Tier-1 that are not part of Tier-2 form the *Remainder*, defined as $R = O_1 \setminus O_2$. By construction, Tier-1 can be partitioned into two disjoint subsets, $O_1 = R \cup O_2$ with $R \cap O_2 = \varnothing$. In compact form, the nesting structure is summarized by the subset chain $\mathbf{D} \supseteq O_1 \supseteq O_2$.

A. Mahalanobis on Aggregated Data (Tier-1)

Tier-1 is the first Mahalanobis cut on the per-IP aggregate (*attack count*, *username variety*, *node variety*). In the scatter plot (Figure 4, the inliers form a compact, tilted ellipsoid, while Tier-1 points fall outside it, stretching the cloud mainly along *username variety* and, to a lesser extent, *node variety*, with only a mild push on *attack count*. Correlation analysis confirms this geometry: username vs. node variety shows a modest positive link ($r \approx 0.34$), while the attack count is weakly related to either axis ($r \approx 0.10$ and $r \approx -0.07$). Thus, as Figure 4 shows, Tier-1 separation is driven more by *breadth*

TABLE II: Feature Distribution Across Outlier Tiers

Feature	Total Dataset	Tier-1	Tier-2	Combined (T1+T2)	Tier-1 %	Tier-2 %	Combined %
Attack Count	89,090,860	59,721,573	15,974,515	75,696,088	67.0%	17.9%	84.9%
Username Variety	12,386,995	7,202,280	14,846	7,217,126	58.1%	0.1%	58.3%
Node Variety	5,043,482	2,440,703	40,059	2,480,762	48.4%	0.8%	49.2%
Number of IPs	77,886	5,186	72	5,258	6.7%	0.09%	6.8%

than by raw volume. Despite representing only 6.7% of IPs (5,186 out of 77,886), Tier-1 accounts for 67.0% of all attacks (59.7M out of 89.1M), 58.1% of username variety (7.20M out of 12.39M), and 48.4% of node variety (2.44M out of 5.04M). Tier-1 filtering exploits the heavy-tailed distribution of attack telemetry, where a sparse minority of sources drives the majority of system load and behavioral entropy. By isolating this "*exploratory plume*," our framework retains over 50% of feature diversity and bulk attack volume while drastically reducing the active analysis set. These Tier-1 sources exhibit high-cardinality username and target node exploration, allowing the system to capture sophisticated campaign signatures with minimal computational overhead compared to processing the undifferentiated inlier population.

B. *Outliers of Outliers (Tier-2)*

To find direct evidence of coordination between attackers, we apply a second outlier-detection pass, using a 99% Mahalanobis cut on the Tier-1 outliers. Tier-2 sources represent high-priority risks. As Figure 5 makes clear, this second pass does more than select "bigger points": it isolates *off–manifold* tails that are orthogonal to the inlier crescent high volume with little username breadth, high breadth with minimal node reach, and scattered extremes along the attack–node axes. Table III categorizes these departures into four subtypes, each with distinct means across attack volume, node reach, and credential breadth. Statistically, Tier-2 is important because it separates *heterogeneous tails* that would otherwise be pooled into a single "*outlier*" bucket: each subtype loads a different dimension of the covariance (volume VS. node VS. username), concentrates a disproportionate share of tail mass, and exerts strong influence on global summaries (means, variances, correlations). By partitioning these influential points, Tier-2 preserves the structure of the bulk distribution while yielding interpretable, high-leverage profiles for analysis and monitoring. The framework reduces high-cardinality attack telemetry into a tractable set of extreme behavioral profiles, mitigating alert fatigue. By collapsing thousands of raw sources into coordinated signatures based on shared temporal, credential, and targeting heuristics, the system enables high-fidelity attribution. This projection from undifferentiated logs to compact manifolds facilitates rapid analysis of infrastructure reuse and campaign evolution. As Table II shows, despite comprising only 0.09% of IPs, Tier-2 accounts for 17.9% of total attack volume (15.9M events). Their extreme statistical deviation and high resource-exhaustion footprint justify their status as high-priority anomalies for triage.

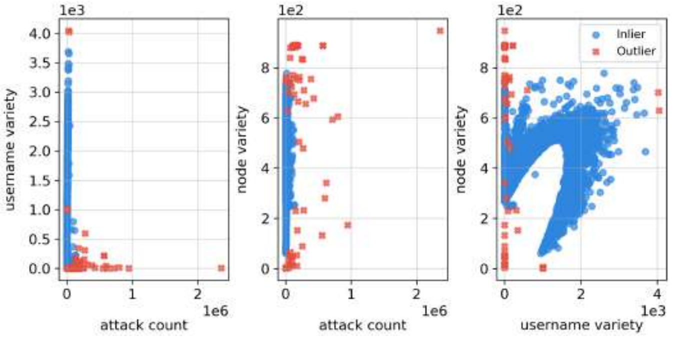


Fig. 5: Mahalanobis Tier2

1) Subcluster Structure (Tier-2): Within O_2, similarity metrics (e.g., Jaccard overlap of username sets, temporal alignment, and ASN/provider grouping) reveal that Tier-2 is not monolithic. Instead, several distinct sub-cluster patterns emerge, each providing insight into a different type of attacker coordination. In this section, we describe several techniques to further subdivide outliers into groups based on different types of evidence of coordination.

a) Clustering by Subnet & Autonomous Systems: Within Tier-2, we observe several IPs forming tight "twins," in which multiple attackers from different but related networks exhibit almost identical attack patterns across our three main dimensions (attack volume, username spread, and target spread). When enriched with IPinfo, these subnet groupings consistently map to the same autonomous system[1] (ASN) and provider. The largest example is a block of 22 addresses under Chinanet, all tied to the same backbone and clustered within neighboring prefixes. The uniformity of ASN and provider information indicates that these IPs are routed through a common national ISP backbone, making the cluster stand out as an infrastructural concentration rather than a collection of scattered singletons. Beyond this dominant group, we also detect smaller subnet clusters. Three IP addresses resolve to infrastructure in the same region as the facility from which our data originates, and they align under the same provider entry. Two IP addresses appear together in Moscow, sharing both the prefix and the ASN metadata. Four additional IP addresses are located in Amsterdam, and the hosting networks are grouped in a narrow address block. Each set illustrates the way subnet twins arise across different geographic and

[1] Roughly, the networks that comprise the Internet, such as backbones, end-user Internet Service Providers, large companies, major educational institutions, and other organizations that run their own networks.

TABLE III: Summary of Tiers' Characteristics

Cluster Type	Unique IPs	Avg. Attack Count	Avg. Username Variety	Avg. Node Variety
Subnet & Infrastructure Clusters	31	174,385.77	26.38	752.48
Single-Username Swarms	5	4,614.00	1.00	751.20
Distributed High-Username Attackers	4	5,732.50	1001.00	3.00
Extreme Anomalies (Profiles)	23	386,362.86	587.69	572.00

infrastructural contexts—sometimes within national carriers, sometimes within regional ISPs, and sometimes within commercial hosting providers. These subnet clusters likely indicate actors controlling multiple attacking devices on the same network; they can indicate multiple compromised machines in the same network, multiple VMs rented in the same cloud, or attacks launched by insiders within an organization.

b) Clustering by Username: Single-Username Swarms:: Within Tier-2, we identify a small but highly consistent swarm: five distinct IP addresses, all issuing attempts exclusively with the username *admin*. Unlike the typical brute-force background, where the username *root* dominates, this cluster stands out for its non-root focus and uniform behavior. Every member shows the same monotone signature, one credential repeated across multiple nodes, creating a tightly bound subcluster defined less by volume than by its exact replication of strategy. Enrichment with IPinfo reveals that all five swarm IPs originate from South Korea, specifically under the Korea Telecom infrastructure. The convergence of geographic location, provider, and username choice suggests that this swarm is not an accidental overlap but rather a coordinated pattern. Although small in size, the repetition across multiple independent IPs makes this a distinctive anomaly: a South Korean admin swarm in which identical credential focus is broadcast from separate addresses within the same telecom network.

c) Clustering by Username: Distributed High-Username Attackers: Another Tier-2 pattern lies at the opposite end of the swarm spectrum: sources with extremely high username variety but very limited target node coverage. We filtered Tier-2 for the top decile of username variety and combined it with the bottom quartile of node variety, isolating just four attacking IP sources. Each cycle traverses thousands of candidate credentials against only one or two targets, a behavior characteristic of dictionary injectors or credential stuffers. Their selectivity makes them stand out as concentrated probes that are wide in attempted credential space (exactly 1,001 usernames attempted by each) but narrow in target scope. A closer inspection reveals that these IP addresses geolocate to three different areas: one resolves to Sofia, another to Amsterdam, and the remaining two originate from Moscow networks. Despite geographic spread, the similarity in credential strategy and the partial prefix overlap suggest a degree of common tooling or distribution.

Using the same technique of filtering for high *username_variety*, we also found two attacking addresses that used over 4,000 usernames each ($4,045$ for one and $4,022$ for the other), as well as a close pair using 595 and 343 usernames, and smaller instances with 43, 18, and 6 usernames, respectively.

To test whether similarity in the size of username lists also implied similarity in content, we measured the *Jaccard similarity* of username sets:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}.$$

This metric, ranging from 0 (no overlap) to 1 (identical sets), quantifies how much two attackers' dictionaries coincide. The results show striking contrasts, indicating very different types of coordination: the 4045/4022 pair had almost no overlap ($J = 0.003$), while the 1001-username outliers matched perfectly ($J = 1.0$). The ten single-username IPs also aligned exactly ($J = 1.0$), and the 595/343 pair was nearly identical ($J = 0.99$). In contrast, the smaller sets at 43 and 18/6 showed only weak alignment ($J = 0.043$ and $J = 0.016$).

Taken together, these comparisons reveal that not all extreme outliers are alike. Some represent *independent extremes*, such as the 4045/4022 pair that scales up dictionary size without sharing actual content. This strongly suggests that the pair is operating under a shared username set that has been divided across multiple attacking devices.

Others form clear signs of working from the *same username set*, like the 1001-username injectors or the single-username swarm, where Jaccard values near 1.0 indicate reuse of the same credential lists. This indicates that these attackers are either dividing up the target space or choosing targets at random. The 595/343 pair stands between, suggesting a near-clone dictionary with only minor drift. These cases will serve as exemplar Tier-2 profiles in the analysis: a subnet twin cluster, a monotone swarm, a distributed high-username attacker, and one extreme but independent outlier.

All Four Phenotypes in a United Frame: Tier-2 splits into four distinct phenotypes with orthogonal extremes. Extreme Anomalies (23 IPs) dominate volume (avg. attack count $\approx 386\,000$), with broad node reach (avg. nodes ≈ 572) and high credential breadth (avg. usernames ≈ 588): these are attackers that use high volume against a spread set of targets. Subnet and Infrastructure clusters (31 IPs) emphasize target breadth (avg. nodes $\approx$ 752) at moderate volume ($\approx$ 174 000) and low credential variety (≈ 26), consistent with wide scanning. Single-username swarms (5 IPs) exhibit minimal credential breadth (1) yet very broad node reach ($\approx$ 751), indicating that focused dictionaries are widely disseminated. Distributed High-Username attackers (4 IPs) invert this: maximum credential breadth (1 001) with narrow targeting (≈ 3 nodes) and modest volume ($\approx$ 5 700). Statistically, the means separate cleanly along three axes: volume, node breadth, and credential breadth, indicating that ”*outlier*” does not have a single meaning. This matters because mitigation should be keyed to the dominant extreme: rate/pressure controls for Extreme Anomalies, destination-aware throttles for high-node clusters, and credential-list defenses (e.g., canaries, rapid lockouts) for high-username groups, rather than a one-size-fits-all rule.

C. Correlation of Outliers Within Tier-2

Our next set of analyses examines correlations among outliers within Tier 2 to identify additional actionable threat intelligence. Identifying high correlations between different types of attacks can facilitate quick decision-making for blocking. We use Pearson correlation to compare patterns across countries that appear as sources in this tier.

We plot these correlations using a colormap with the range $r \in [-1, 1]$ being shown with a purple→yellow scale: *purple/blue = negative or weak, orange/yellow = moderate to strong positive*. The diagonal is bright yellow (r=1) because each country is always perfectly correlated with itself.

Attack count (Figure 6a): The matrix is a near-solid warm block off the diagonal. Concretely, pairs such as *China–United States*, *China–Russia*, *United States– Russia*, *Germany–The Netherlands*, and *South Korea–Estonia* all sit in orange/yellow tiles. Only a few edge tiles (e.g., involving *Venezuela* with *South Korea* or *Estonia*) show slightly cooler orange, but the overall field remains uniformly warm.

Node variety (Figure 6b): Warm colors still dominate, but are more structured. Columns/rows for *China* and *United States* stay broadly warm against most partners (e.g., *China–Russia*, *United States–Germany*, *United States–The Netherlands*). In contrast, some pairs involving *Estonia* or *South Korea* display cooler orange patches (moderate r), and a few isolated tiles e.g., *Estonia* with *Venezuela* or *The Netherlands* dip toward cooler shades, indicating weaker linear association for node breadth in those pairings.

Username variety (Figure 6c): This panel is the most heterogeneous. While many pairs remain warm (e.g., *China–United States*, *United States–Russia*, *Russia–The Netherlands*), there are visible cooler cross-bands. Notably, tiles along *Germany* and *Estonia* with several others (e.g., *Germany–China*, *Germany–United States*, *Estonia–United States*, *Estonia–Russia*) shift toward cool orange/blue, and some pairings with *South Korea* also cool. These named rows/columns mark country pairs where changes in username breadth align less consistently.

Color summary: Country pairs are most uniformly warm for *attack count* (e.g., *China/US/Russia/–Germany/The Netherlands* among each other), become more mixed for *node variety* (cooler patches often involving *Estonia* or *South Korea*), and are most variable for *username variety* with cooler cross-bands especially through *Germany* and *Estonia*. These observations strictly describe the color patterns (linear co-variation) in the heatmaps without inferring behavior.

D. Volume–Diversity Profiles and Cross-Tier Correlations

Having closely examined the attackers within Tier-2, we now step back to examine Tier-1 and its inliers. Our intention is to apply the lessons learned from a small set of Tier-2 attackers at scale.

1) Analyzing the distribution of attack count vs. username variety via boxplot:: Figure 7a summarizes the distribution of

$$\frac{\#\text{attack count}}{\#\text{username variety}}$$

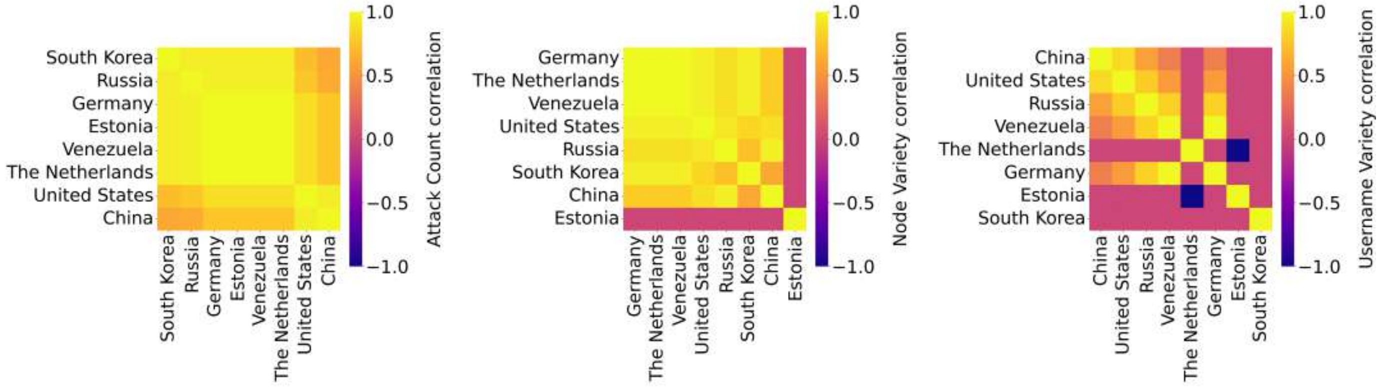
in three groups (Inliers, Tier-1, Tier-2). The Inlier and Tier-1 boxes are compressed near zero with short whiskers, indicating low attempts per username and limited spread. Tier-2 shows a markedly higher median and a much taller box-and-whisker plot, with numerous high-end points. Visually, Tier-2 concentrates far more attempts per distinct username than the other groups. This demonstrates that both Tier-1 and Tier-2 identify a set of attackers whose behavior is statistically distinct from the Inliers, but in different ways.

2) Analyzing attack count vs. node variety ratio (boxplot):: As with the previous analysis, this one looks at a ratio, this time between the total number of attacks from a source and the number of targets that source hits. Figure. 7b shows that Inliers and Tier-1 remain near zero with limited dispersion, though Tier-1 has a much tighter distribution. Tier-2 exhibits a substantially higher central tendency and broader spread, with many high-ratio points. Hence, the number of attempts per node is much larger for Tier-2 than for the other groups. Tier-1 also shows a much tighter distribution than the other two categories.

3) Analyzing the username variety distribution via histogram:: As Figure 7c shows, the density curves shift progressively to the right from Inliers to Tier-1 and again to Tier-2. Tier-1 already exhibits higher username variety than the background, and Tier-2 is further right-shifted with a tighter, high-variety mass. Qualitatively, the probability mass for Tier-2 sits at larger values of distinct usernames than Tier-1, which in turn exceeds the inlier background. Particular spikes can be seen in Tier-2 at around 1,000 and 4,000 usernames, corresponding to specific groupings discussed in Section V-B. Interestingly, Tier-1 shows a high density of attackers, with around 2,000 usernames; we will return to an analysis of this tier in a later section.

4) Analyzing the node variety distribution (histogram): As Figure. 7d shows that the density of target nodes shifts to the right from Inliers to Tier-1, with Tier-2 covering a broader range of attack targets. Again, we see that many sources in Tier-1 have a similar number of target counts, as they did with username variety; this further suggests that many or most of the attackers in Tier-1 may be coordinating with each other, a hypothesis we will return to.

5) Pearson view (score–score): Figure 7e shows a comparison of continuous scores from the two Tiers (e.g., Tier-1 score on the x-axis vs. Tier-2 score on the y-axis) with a fitted least-squares line. Points fall nearly on a straight line with a slope close to unity, and the reported R^2 is near 1.0. Thus, as



(a) countries correlation based on attack count

(b) countries correlation based on node variety

(c) countries correlation based on username variety

Fig. 6: Correlation between outlier countries in Tier-2 based on attack count, node variety, and username variety

a continuous metric, the two scoring schemes agree linearly to a very high degree.

6) Spearman view (rank–rank): As Figure 7f shows, scores are converted to ranks before comparison. The cloud aligns closely with the diagonal, indicating a strong *monotonic* relationship between orders of the two levels. Small deviations, especially near the extremes, reflect local rank swaps and ties, but the overall rank agreement remains high.

Remark: Altogether, in ratio boxplots, Tier-2 has the highest attempts-per-username and attempts-per-node, with the widest spread. Both diversity distributions are progressively right-shifted from Inliers to Tier-1 to Tier-2, indicating an increase in distinct usernames and nodes. Finally, the Pearson view shows near-linear agreement between tier scores, while the Spearman view confirms that this agreement also holds at the rank (monotonic) level.

7) Pearson Shifts (Figures 8a and 8b), Username–Node Coupling Collapses in Tier-2:: Across the entire set (Tier-1), the only notable relationship is a moderately positive link between username variety and node variety (r $\approx$ 0.34). This suggests that sources that diversify the names they use also tend to connect to more nodes. All other pairwise correlations are weak, so the attack count is largely independent of both varieties at the population level.

In Tier-2, that structure collapses: all off-diagonal entries are near zero $|r| \leq 0.14$, and the username–node coupling essentially disappears (slightly negative). Tier-2, therefore, appears nearly orthogonal; the three features convey distinct information with minimal linear redundancy. Practically, there is no single dominant axis; variation is distributed across multiple dimensions. Looking feature-by-feature, attack count vs. username variety shifts from almost neutral (r$\approx$–0.07) to a small negative (r$\approx$–0.14). In this tier, higher counts tend to come from repeating fewer names rather than broadening the breadth of names. The attack count versus node variety remains a small positive in both tiers ($\approx$+0.10…+0.12), indicating that total volume is only a weak proxy for how widely nodes are targeted. The pivotal change is *username variety vs. node variety*: what is moderately coupled in Tier-1 decouples in Tier-2. This means that extreme cases no longer follow the population pattern that "*more name breadth → more node breadth*." Instead, sources can specialize along one axis without moving on the other, yielding heterogeneous modes that linear correlation cannot combine into a single trend. The implications of this analysis are tier-specific: in Tier-1, it can be assumed that some co-movement will occur between the two variety measures; in Tier-2, three features must be treated as independent axes. Ranking on any single feature will miss critical structure; multivariate or interaction-aware methods are required to separate behaviors within the extreme subset.

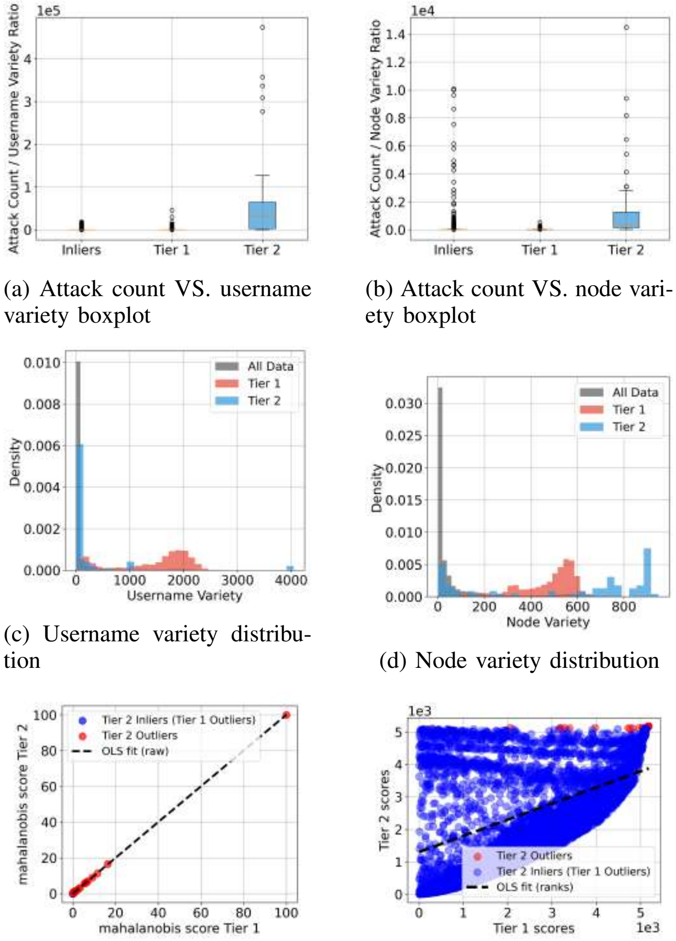
VI. Coverage of Three High-Similarity Username Subsets Across All IPs

We focus on *three* username subsets (excluding "root" username) from the Tier-2 extreme-outliers subcluster, the 1,001-username set ($J = 1.0$), the $\sim$216-username set ($J \approx 0.99$), and the single-username set ($J = 1.0$), and measure their usage across all source IPs. Let $\mathcal{I}$ be the 77,886 IPs and $\mathcal{U}$ the global username vocabulary. We build a sparse binary incidence matrix $X \in \{0,1\}^{|\mathcal{I}|\times|\mathcal{U}|}$ where $X_{i,u} = 1$ iff IP i attempted username u. For each subset $S_k \subseteq \mathcal{U}$ (with indicator $\mathbf{s}^{(k)} \in \{0,1\}^{|\mathcal{U}|}$), the coverage vector $\mathbf{c}^{(k)} = X\,\mathbf{s}^{(k)} \in \mathbb{N}^{|\mathcal{I}|}$ counts, per IP, how many usernames from S_k are present. Full-membership can be defined as $m_i^{(k)} = \mathbf{1}\{c_i^{(k)} = |S_k|\}$ (or $c_i^{(k)} \geq \tau$ for a partial threshold τ). Aggregating $m^{(k)}$ yields per-subset adoption counts and fractions over $\mathcal{I}$, and logical intersections (e.g., $m^{(a)} \wedge m^{(b)}$) quantify overlap—computed directly via sparse matrix–vector products without quadratic pairwise set comparisons.

Matrix cardinality: The binary matrix has $|X| = |\mathcal{I}| \times |\mathcal{U}| =$ 77,886 $\times\ |\mathcal{U}|$ entries, with nonzeros $\mathrm{nnz}(X) = \sum_{i\in\mathcal{I}} |\mathcal{U}_i|$, where $\mathcal{U}_i$ is the username set observed for IP i.

TABLE IV: Per-IP statistics for two high-similarity username subsets (Tier-2).

Subset	IP	Attacks	Node Variety	Username Variety
1001-usernames	193.37.69.221	11,740	7	1001
1001-usernames	185.73.124.15	6,091	2	1001
1001-usernames	185.73.125.81	2,823	2	1001
1001-usernames	87.251.75.121	2,276	1	1001
Subset-1 totals / mean attacks		*22,930 / 5,732.5*	*range 1–7*	*fixed*
$\sim$216-usernames	152.89.196.220	569,287	887	216



(a) Attack count VS. username variety boxplot

(b) Attack count VS. node variety boxplot

(c) Username variety distribution

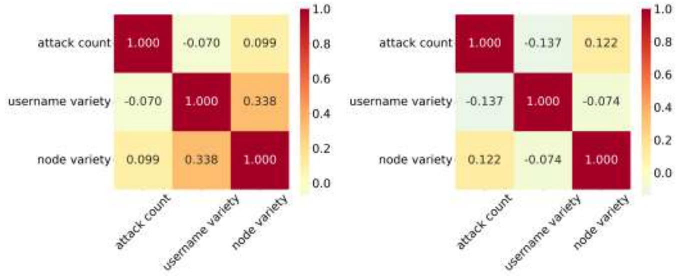
(d) Node variety distribution

(e) Pearson correlation view

(f) Spearman correlation view

Fig. 7: Comparison of attack count, username/node variety distributions, and correlation views (Pearson and Spearman).

Understanding how these high-similarity username subsets spread across the population shows whether they are niche artifacts tied to a few sources or widely reused dictionaries. Broad coverage points to reusable lists worth prioritizing for rules and throttling, and for canary credentials; concentrated coverage favors targeted actions, such as per-source filtering or takedowns. Measuring overlap reveals compounding risk when the same attacking IP sources adopt multiple subsets.



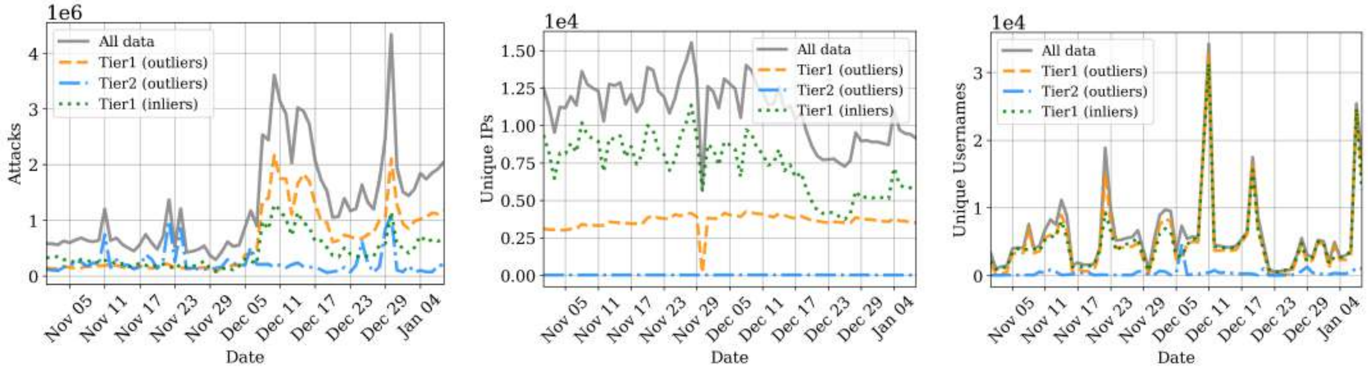
(a) Tier-1 Pearson correlation

(b) Tier-2 Pearson correlation

Fig. 8: Pearson correlation matrices for Tier-1 and Tier-2.

The coverage matrix comprises 9,405 rows (IPs) and three columns corresponding to the Tier-2–derived username subsets (1,001-username, $\sim$216-username, and single-username). Column sums show a highly skewed distribution: $\sum$subset 1 $= 4$, $\sum$subset 2 $= 1$, and $\sum$subset 3 $= 9{,}400$. These counts represent only the IPs that showed 100% usage of each subset, that is, sources whose username lists matched exactly the given subset. Thus, only a handful of sources engage with the larger dictionaries (about 0.04% and 0.01% of IPs, respectively), whereas virtually all activity aligns with the single-credential subset ($\approx 99.95\%$). This stark asymmetry suggests that extensive credential lists are employed by rare, high-complexity actors, while the majority of IPs exhibit uniform, single-credential behavior, converting qualitative coordination signals into a compact, reproducible statistic.

As Table IV shows, Subset 1 (1,001 usernames) appears as depth-focused probing: four sources generate moderate volumes (2.3k–11.7k attacks each) while concentrating on a few nodes (1–7), consistent with iterative dictionary cycling against limited targets. In contrast, Subset 2 ($\sim$216 usernames) presents breadth-focused spraying: a single source launches 569k+ attempts across 887 nodes, pushing a mid-sized list horizontally. Additionally, Subset 3 is single-credential: every attempt uses the username "*admin*" (9,400 IPs show full membership). "*admin*" is the most universally targeted privileged username. Compromising this username yields immediate, high-impact control. Altogether, these profiles illustrate two ends of the Tier-2 spectrum—intensive per-host brute forcing versus large-scale distributed spraying—highlighting that dictionary size alone does not determine volume; the operational topology (node fan-out) is the differentiator.



(a) attack pattern through time

(b) unique IPs pattern through time

(c) unique usernames pattern through time

Fig. 9: patterns through time

VII. Using time-series to find coordination

Finally, we track the evolution of key signals over time: total attacks, distinct sources (IP addresses), and distinct usernames, decomposing the aggregate into statistically defined strata.

Notation. Let D be the full dataset, O_1 the Tier-1 outliers, and $O_2 \subset O_1$ the Tier-2 outliers. In all plots, Tier-1 (outliers) denotes $O_1 \backslash O_2$ (Tier-1 after removing Tier-2), Tier-2 (outliers) denotes O_2, and Tier-1 (Inliers) denotes $D \setminus O_1$.

Attack pattern through time: Figure 9a separates the global time series into *Tier-1 (Inliers)*, *Tier-1 (outliers)*, and *Tier-2 (outliers)*. The Tier-1 (outlier) series correlates and co-moves with "*all data*" but with larger amplitudes on surge days, indicating that system-level waves are primarily intensity ramps among already active outliers. The Tier-1 (Inliers) line forms a tempered baseline with only mild deflection during peaks, delineating background variability from the outlier signal. Tier-2 remains low in magnitude yet persistent. A few extreme sources that are active but not determinative of the aggregate shape.

Unique IPs pattern through time: Figure 9b shows a broadly stable Tier-1 (Outliers) population, including a brief, steep mid-period dip that quickly reverts. This is remarkable, as it shows that all Tier-1 outliers abruptly stopped attacks for a single day, then resumed the following day. The Tier-1 Inliers do show a dip on the same day, but it is nowhere near as dramatic as it appears, given the level of day-to-day noise common among the Inliers. Read alongside the Figure. 9a and the evidence in Figure. 7, this suggests a high likelihood that the Tier-1 outliers form a coordinated botnet comprising over 5,000 attackers. Most source IP addresses in this set (a) use a similar number of usernames (around 2,000), (b) attack a similar number of targets (400–600), and (c) cease attacking for the same single day. The fact that the Inliers drop somewhat, but not as dramatically, on the same day suggests that the set of Tier-1 outliers captures most, though not all, members of the campaign.

Unique usernames pattern through time: Figure 9c shows bursts in credential breadth that co-occur with volume surges. Tier-1 (outliers) carries most of this variation, alternating between phases of dictionary reuse and expansion, whereas Tier-2 exhibits sparse, distinct pulses consistent with compact, high-value sets identified elsewhere in the study. The Tier-1 (Inliers) line mirrors the overall trend but with reduced amplitude, representing baseline activity without an extreme surge. Taken together, the alignment of username bursts with attack waves indicates structured changes in credential use within the outlier strata, without additional turnover in the participating IPs.

VIII. Limitations

Our framework distinguishes between statistical congruence and causal attribution, treating multidimensional coordination as a lead-generation heuristic rather than proof of common provenance. While the system surfaces high-fidelity clusters based on shared artifacts, these outputs serve as an investigative substrate requiring external validation. We acknowledge that behavioral correlations across infrastructure and credentials suggest coordination, but do not definitively confirm latent actor intent.

IX. Conclusion

This work introduces a scalable framework for transforming high-cardinality SSH telemetry into interpretable behavioral signatures. By employing a two-tier Mahalanobis filter, we prune undifferentiated background noise to isolate a sparse, high-signal minority that drives disproportionate network load. This structural data reduction enables the downstream forensic examination of coordinated infrastructure, which is typically obscured by the raw log volume, and reveals it through high Jaccard similarity and temporal alignment. We demonstrate that operational risk is defined by multidimensional exploratory patterns rather than univariate totals, providing a generalizable pipeline for prioritizing adversarial behavior in high-throughput environments.

X. Acknowledgment

This material is based upon work supported by the National Science Foundation under Award No. 2027208.

References

[1] AbdelRahman Abdou, David Barrera, and Paul C. van Oorschot. What lies beneath? analyzing automated ssh bruteforce attacks. In Frank Stajano, Stig F. Mjølsnes, Graeme Jenkinson, and Per Thorsheim, editors, *Technology and Practice of Passwords*, pages 72–91, Cham, 2016. Springer International Publishing.

[2] T. Barron and N. Nikiforakis. Picky attackers: Quantifying the role of system properties on intruder behavior. In *Proceedings of the 33rd Annual Computer Security Applications Conference (ACSAC '17)*, pages 387–398, New York, NY, USA, 2017. Association for Computing Machinery.

[3] Phuong M. Cao, Yuming Wu, Subho S. Banerjee, Justin Azoff, Alexander Withers, Zbigniew T. Kalbarczyk, and Ravishankar K. Iyer. CAUDIT: Continuous auditing of SSH servers to mitigate brute-force attacks. In *Proceedings of the 16th USENIX Symposium on Networked Systems Design and Implementation (NSDI '19)*, Boston, MA, USA, 2019. USENIX Association.

[4] SSHGuard Contributors. Sshguard: Brute-force attack protection for unix-like systems. https://www.sshguard.net/, 2024. Accessed: 2025-08-28.

[5] Dinei Florêncio, Cormac Herley, and Paul C. van Oorschot. A large-scale study of web password habits. In *Proceedings of the 23rd USENIX Security Symposium*, New York, NY, USA, 2014. Association for Computing Machinery.

[6] Fyodor and the Nmap Project. Ncrack: High-speed network authentication cracking tool. https://nmap.org/ncrack/, 2024. Accessed: 2025-08-28.

[7] Vincent Ghiette, Harm Griffioen, and Christian Doerr. Fingerprinting tooling used for SSH compromisation attempts. In *22nd International Symposium on Research in Attacks, Intrusions and Defenses (RAID 2019)*, pages 61–71, Chaoyang District, Beijing, September 2019. USENIX Association.

[8] M. Gupta, J. Gao, C. C. Aggarwal, and J. Han. Outlier detection for temporal data: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 26(9):2250–2267, 2013.

[9] John Hancock, Taghi M. Khoshgoftaar, and Joffrey L. Leevy. Detecting ssh and ftp brute force attacks in big data. In *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 760–765, 2021.

[10] Laurens Hellemons, Luuk Hendriks, Rick Hofstede, Anna Sperotto, Ramin Sadre, and Aiko Pras. Sshcure: A flow-based ssh intrusion detection system. In Ramin Sadre, Jiří Novotný, Pavel Čeleda, Martin Waldburger, and Burkhard Stiller, editors, *Dependable Networks and Services*, volume 7279 of *Lecture Notes in Computer Science*, pages 86–97, Berlin, Heidelberg, 2012. Springer.

[11] Md Delwar Hossain, Hideya Ochiai, Fall Doudou, and Youki Kadobayashi. Ssh and ftp brute-force attacks detection in computer networks: Lstm and machine learning approaches. In *2020 5th International Conference on Computer and Communication Systems (ICCCS)*, pages 491–497, 2020.

[12] Yudi Huang, Yilei Lin, and Ting He. Optimized cross-path attacks via adversarial reconnaissance. *Proc. ACM Meas. Anal. Comput. Syst.*, 7(3):58:1–58:30, December 2023.

[13] L. Izhikevich, M. Tran, M. Kallitsis, A. Fass, and Z. Durumeric. Cloud watching: Understanding attacks against cloud-hosted services. In *Proceedings of the 2023 ACM Internet Measurement Conference (IMC '23)*, pages 313–327, New York, NY, USA, 2023. Association for Computing Machinery.

[14] Cyril Jaquier and Fail2Ban Contributors. Fail2ban: Ban hosts that cause multiple authentication errors. https://github.com/fail2ban/fail2ban, 2024. Accessed: 2025-08-28.

[15] Mobin Younas Javed and Vern Paxson. Detecting stealthy, distributed SSH brute-forcing. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer and Communications Security (CCS)*, Berlin, Germany, 2013. ACM.

[16] C. Munteanu, S. J. Saidi, O. Gasser, G. Smaragdakis, and A. Feldmann. Fifteen months in the life of a honeyfarm. In *Proceedings of the 2023 ACM Internet Measurement Conference (IMC '23)*, pages 282–296, New York, NY, USA, 2023. Association for Computing Machinery.

[17] Cristian Munteanu, Yogesh Bhargav Suriyanarayanan, Georgios Smaragdakis, Anja Feldmann, and Tobias Fiebig. Attacks come to those who wait: Long-term observations in an ssh honeynet. In *Proceedings of the 2025 ACM Internet Measurement Conference (IMC '25)*, Madison, WI, USA, October 2025. ACM.

[18] Maryam M. Najafabadi, Taghi M. Khoshgoftaar, Chad Calvert, and Clifford Kemp. Detection of ssh brute force attacks using aggregated netflow data. In *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, pages 283–288, 2015.

[19] John Owens and Jeremy Matthews. A study of passwords and methods used in brute-force ssh attacks. In *Proceedings of the USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET)*. USENIX Association, 2008.

[20] P3TERX. Geolite.mmdb. https://github.com/P3TERX/GeoLite.mmdb, 2023. Accessed: 2025-10-14.

[21] Phil Schwartz and DenyHosts Contributors. Denyhosts: Ssh attack prevention and host blocking tool. https://denyhosts.sourceforge.net/, 2024. Accessed: 2025-08-28.

[22] Sachin Kumar Singh, Shreeman Gautam, Cameron Cartier, Sameer Patil, and Robert Ricci. Where the wild things are: Brute-Force SSH attacks in the wild and how to stop them. In *21st USENIX Symposium on Networked Systems Design and Implementation (NSDI 24)*, pages 1731–1750, Santa Clara, CA, April 2024. USENIX Association.

[23] van Hauser/THC and David Maciejak. Hydra: A fast network logon cracker. https://www.kali.org/tools/hydra/, 2024. Accessed: 2025-08-28.

[24] Limin Yang et al. Understanding and mitigating alert overload in security operations centers. In *Proceedings of the 33rd USENIX Security Symposium (USENIX Security 24)*, Philadelphia, PA, USA, 2024. USENIX Association.