Task 7: An Analysis of the Hyper-Volumetric DDoS Threat Landscape and a Case Study of the Record-Breaking 7.3 Tbps Attack of May 2025

R S Abhinav

October 16, 2025

Abstract

The distributed denial-of-service (DDoS) threat landscape underwent a paradigm shift in 2024 and 2025, characterized by an exponential increase in attack frequency and the normalization of hyper-volumetric assaults exceeding multiple terabits per second. This report provides a comprehensive analysis of this new era of digital sieges. It begins with a survey of five significant DDoS incidents that exemplify the diversity of modern attack methodologies, from sophisticated application-layer campaigns to brute-force volumetric floods. The core of the report is an in-depth case study of the record-breaking 7.3 Tbps multi-vector attack mitigated in May 2025. This analysis deconstructs the attack's anatomy, including its volumetric characteristics, vector composition, and global botnet infrastructure. It further explores the attacker's potential motives, assesses the impact, and details the advanced, automated defensive strategies that proved critical for successful mitigation. The report concludes with strategic imperatives for achieving DDoS resilience in an environment where multi-terabit attacks are no longer anomalous but an operational reality.

Contents

1 Introduction: The Paradigm Shift in Digital Sieges

The years 2024 and 2025 marked a fundamental and unprecedented escalation in the scale, frequency, and strategic complexity of distributed denial-of-service (DDoS) attacks. The threat evolved from a disruptive nuisance into a strategic weapon capable of inflicting significant economic and reputational damage. This transformation is not merely an incremental increase but a paradigm shift, redefining the operational reality for any organization with an online presence.

1.1 The Statistical Explosion

The quantitative data from the period paints a stark picture of this new reality. In the first quarter of 2025 alone, the web infrastructure and security company Cloudflare reported mitigating 20.5 million DDoS attacks. This figure is staggering when contextualized: it represents 96

1.2 The Normalization of Hyper-Volumetric Attacks

Perhaps the most defining characteristic of this new era is the normalization of hyper-volumetric attacks, defined as those exceeding 1 terabit per second (Tbps) or 1 billion packets per second (Bpps). Once considered rare, "black swan" events, such assaults became a regular feature of the threat landscape. In Q1 2025, Cloudflare mitigated an average of eight such attacks every day.[1, 3] The rapid succession of record-breaking volumetric attacks illustrates this trend: a 5.6 Tbps attack was mitigated in late 2024, followed by a 6.5 Tbps attack in early 2025, culminating in the 7.3 Tbps event in May 2025 that is the subject of this report's case study.[1, 2, 4] This progression demonstrates that the upper ceiling of attack magnitude is continuously being pushed, making multi-terabit defense a baseline requirement for infrastructure providers.

1.3 The Strategic Evolution of DDoS

The purpose and application of DDoS attacks have evolved in lockstep with their technical capabilities. They have transitioned from simple acts of digital vandalism or hacktivism to a primary tool in a sophisticated cybercriminal's arsenal. The motives are now predominantly commercial and strategic:

- **Extortion:** Ransom DDoS (RDDoS) has become a primary driver. In the fourth quarter of 2024, reports from Cloudflare customers of receiving a ransom note or threat in conjunction with a DDoS attack increased by 78

- **Business Disruption:** Attacks are increasingly used for direct commercial sabotage. When surveyed, 39

- **Diversionary Tactic:** Most alarmingly, DDoS attacks are now frequently employed as a smokescreen for more sinister activities, such as data exfiltration or the deployment of ransomware.[1] A large-scale volumetric attack is designed to consume the full attention and resources of a security operations center (SOC). While the security team is engaged in an "all hands on deck" effort to restore service availability, the attackers can exploit other, quieter channels to achieve their primary objective of stealing data or encrypting systems. This transforms a DDoS event from a simple availability crisis into a potential indicator of a much deeper compromise of data confidentiality and integrity.

The commoditization of the tools required to launch these attacks is a key enabler of this strategic evolution. The proliferation of powerful botnets like Aisuru and the widespread availability of DDoS-for-hire (or "booter") services have collapsed the barrier to entry.[5, 6] What once may have required the resources of a state-level actor can now be purchased for a nominal fee, making hyper-volumetric attacks accessible to a much broader range of threat actors with diverse motivations.

1.4 The Dual-Pronged Technological Threat

The modern DDoS landscape is characterized by a bifurcation of attack methodologies, forcing defenders to contend with two distinct but equally dangerous types of threats:

1. **Brute-Force Volumetric Floods:** These attacks, which constitute the hyper-volumetric events described above, aim to saturate a target's network bandwidth or overwhelm its network hardware with sheer traffic volume. They are typically generated by large-scale botnets of compromised Internet of Things (IoT) devices and servers.[1]

2. **Stealthy Application-Layer (L7) Attacks:** In contrast to volumetric floods, L7 attacks are more sophisticated and surgical. They target specific functions or APIs within an application, often using very low traffic volumes that mimic legitimate user behavior. This allows them to bypass defenses that are tuned only to detect volumetric anomalies.[1, 7] Attacks targeting APIs are a growing concern; in the financial services sector, for instance, L7 DDoS attacks targeting APIs increased by 58

This dual threat requires a multi-layered defensive strategy capable of both absorbing massive network-layer floods and intelligently distinguishing malicious application-layer requests from legitimate traffic.

2 A Survey of Recent High-Impact DDoS Incidents (2024-2025)

To illustrate the diversity and escalating power of modern DDoS threats, this section provides a technical summary of five distinct, high-impact incidents that occurred between late 2024 and late 2025. These examples showcase the various vectors, scales, and strategies employed by contemporary threat actors.

2.1 The 5.6 Tbps Volumetric Assault (Q4 2024)

During the week of Halloween in 2024, Cloudflare's systems autonomously mitigated a 5.6 Tbps DDoS attack, which at the time was the largest ever publicly reported.[4] The attack was launched by a botnet based on the Mirai malware variant. A key technical detail of this incident is that the immense volume was generated by a botnet comprising only around 13,000 compromised devices.[9] This demonstrates the incredible amplification factor that can be achieved from a relatively small number of insecure IoT endpoints, such as routers and cameras. The incident serves as a stark reminder of the persistent and potent threat posed by the vast ecosystem of unpatched and poorly configured IoT devices.

2.2 Coordinated Application-Layer Campaign (January 2025)

In January 2025, a series of highly sophisticated application-layer (L7) attacks targeted three disparate organizations in rapid succession: a French home supply store, an Indonesian government agency, and a major US beverage company.[10] These were not volumetric attacks but were measured in requests per second (rps), peaking at 6 million rps, 10 million rps, and an astonishing 13.5 million rps, respectively. The primary vector was the HTTP/2 Rapid Reset technique, which exploits a feature of the protocol to overwhelm web servers with a flood of requests that are immediately cancelled.

Forensic analysis of these three events revealed a crucial link: an average of 20

2.3 The Record-Breaking 7.3 Tbps Multi-Vector Attack (May 2025)

In mid-May 2025, the record for the largest DDoS attack was broken again. This incident, which is the focus of the detailed case study in Section 3, peaked at 7.3 Tbps.[2, 11] It is notable not only for its immense scale but also for its tactical execution. The attack was an extremely short "hit-and-run" burst, lasting only 45 seconds.[1, 12] It was also a multi-vector attack, combining a massive UDP flood with a complex mix of numerous reflection and amplification techniques. This incident represents the state-of-the-art in blended volumetric threats, designed to overwhelm defenses with both raw power and technical complexity.

2.4 The 11.5 Tbps UDP Flood (September 2025)

Over the Labor Day weekend in September 2025, another record was set with an 11.5 Tbps attack, representing a nearly 60

2.5 The Aisuru Botnet Onslaught (October 2025)

In late 2025, a powerful new botnet known as Aisuru demonstrated capabilities that dwarfed previous records, with operators flexing DDoS capacities exceeding 29 Tbps in tests.[6] An attack launched by Aisuru targeted multiple US Internet Service Providers (ISPs), highlighting a strategic focus on disrupting core internet infrastructure. What sets Aisuru apart is its propagation method. While older botnets like Mirai rely on scanning the internet for devices with weak, default credentials, Aisuru's operators employed a more advanced supply-chain attack. They compromised the firmware update server of Totolink, a manufacturer of networking equipment, and distributed malicious scripts to expand the botnet, which grew to an estimated 300,000 nodes.[6] This incident signals a dangerous evolution in the botnet arms race, moving from opportunistic exploitation to sophisticated compromises of the technology supply chain, which are far more efficient and difficult to defend against.

3 In-Depth Case Study: The 7.3 Tbps Attack of May 2025

The multi-vector DDoS attack that peaked at 7.3 Tbps in mid-May 2025 serves as a quintessential example of the modern hyper-volumetric threat. Its combination of immense scale, tactical execution, complex composition, and successful mitigation provides a rich dataset for understanding both the capabilities of modern attackers and the requirements for effective defense.

3.1 Target Profile and Context

The target of the attack was an unnamed hosting provider that utilizes Cloudflare's Magic Transit service for network protection.[14, 15] Magic Transit is a BGP-based service that advertises a customer's IP prefixes from Cloudflare's global network, effectively routing all traffic through its data centers for inspection and scrubbing before it reaches the customer's origin infrastructure.

The choice of target is strategically significant. An attack on a large hosting provider acts as a force multiplier for the attacker. A successful disruption would not impact a single organization but could potentially take thousands of the provider's downstream customers offline simultaneously. This elevates the incident from an attack on a single entity to an assault on a piece of critical internet infrastructure, magnifying its potential impact and underscoring the high-stakes nature of the event.

Table 1: Comparative Analysis of Five Major DDoS Incidents (2024-2025)

Incident Name / Target Type	**Date**	**Peak Magnitude**	**Primary Vector(s)**	**Duration**	**Key Characteristics & Suspected Motive**
5.6 Tbps Volumetric Assault	Q4 2024	5.6 Tbps	Mirai UDP Flood	Minutes	Mirai-based botnet of only 13k devices; demonstrates high amplification from IoT. Motive: Likely capability demonstration or disruption.
Coordinated L7 Campaign	Jan 2025	13.5M rps	HTTP/2 Rapid Reset	Hours	Coordinated attacks on 3 targets with 20% IP overlap; used reconnaissance. Motive: Testing of a new attack platform.
7.3 Tbps Multi-Vector Attack	May 2025	7.3 Tbps	UDP Flood, NTP/RIPv1/QOTD Reflection	45 sec	Record-breaking ”hit-and-run” burst; blended brute-force with complex vectors. Motive: Capability demo or state-sponsored probe.
11.5 Tbps UDP Flood	Sep 2025	11.5 Tbps	UDP Flood	35 sec	Pure brute-force saturation attack; demonstrates continued escalation in raw volumetric power. Motive: Disruption.
Aisuru Botnet Onslaught	Oct 2025	>29 Tbps (capability)	Not Specified	Varies	Next-gen botnet targeting ISPs; used supply-chain compromise for propagation. Motive: Strategic disruption of core infrastructure.

3.2 Attack Anatomy and Technology

The attack was a masterful execution of a short-burst, high-intensity volumetric campaign, blended with multiple vectors to maximize pressure on the defensive systems.

3.2.1 Volumetric Characteristics

The attack's defining feature was its sheer scale delivered with incredible velocity. It reached a peak traffic rate of 7.3 terabits per second (Tbps).[11] Over its brief duration of just 45 seconds, it delivered a total of 37.4 terabytes of malicious data to the target's network.[14, 16] To contextualize this volume, delivering 37.4 TB in 45 seconds is equivalent to flooding a network with the data of over 9,350 full-length HD movies or streaming nearly a year's worth of high-definition video content in less than a minute.[11, 16]

This "hit-and-run" or "burst" characteristic is a deliberate and increasingly common tactic.[1, 12] The goals are twofold: first, to cause maximum disruption by delivering a crippling blow before automated defenses can fully analyze the traffic and deploy fine-tuned mitigation rules; and second, to minimize the attacker's operational footprint and exposure, making forensic analysis and traceback more difficult.

3.2.2 Primary Vector Analysis: UDP Flood

The attack was overwhelmingly dominated by a single vector: a User Datagram Protocol (UDP) flood. This vector accounted for 99.996

1. **Bandwidth Saturation:** The primary goal is to consume all available bandwidth on the target's internet connection, creating a traffic jam that prevents legitimate packets from getting through.

2. **Resource Exhaustion:** Network infrastructure components like firewalls, load balancers, and servers must expend CPU and memory resources to process each incoming packet. When a target is flooded, these resources can be exhausted, causing the devices to slow down or fail. For packets arriving at closed ports, the system is often obligated to respond with an ICMP "Destination Unreachable" message, further consuming outbound bandwidth and processing power.[13]

The attackers in this incident employed a "carpet bombing" strategy, spraying packets across a vast range of destination ports on the single target IP. The attack hit an average of 21,925 destination ports per second, with a peak of 34,517 ports per second, indicating a well-engineered effort to maximize pressure on the target's packet processing capabilities.[14, 17]

3.2.3 Secondary Vector Composition

Despite the near-total dominance of the UDP flood, the attack was technically multi-vector. The remaining 0.004

The inclusion of these secondary vectors, even at low volume, is not accidental. It represents an attempt to complicate mitigation. A simple defense against a pure UDP flood might involve aggressive rate-limiting of all UDP traffic. However, such a blunt approach could inadvertently block legitimate UDP-based services like DNS, VoIP, or online gaming.[11] By blending multiple vectors, the attacker forces the defender to employ more intelligent and granular filtering. This tactic can be seen as a form of "denial of defense," aiming to probe for weaknesses in the logic of the mitigation system by presenting it with a diverse range of traffic types simultaneously.[18] The identified secondary vectors included:

- **NTP Reflection/Amplification:** Abuses the Network Time Protocol (UDP/123) on old servers.

- **RIPv1 Amplification:** Exploits the legacy Routing Information Protocol v1 (UDP/520).

- **QOTD Reflection:** Uses the Quote of the Day protocol (UDP/17).

- **Echo Reflection:** Uses the Echo Protocol (UDP/7).
- **Mirai UDP Floods:** Traffic patterns characteristic of the Mirai IoT botnet.
- **Portmap Flood:** Exploits the Portmapper service (UDP/111).

Table 2: Vector Composition of the 7.3 Tbps Attack

Attack Vector	**Type**	**Protocol/Port**	**Mechanism**
UDP Flood	Flood	UDP/Random	Overwhelms the target with a high volume of UDP packets, saturating bandwidth and exhausting network device resources. Constituted 99.996% of traffic.
NTP Reflection	Reflection & Amplification	UDP/123	Spoofed requests are sent to vulnerable NTP servers, which reply with a much larger response to the target's IP address.
RIPv1 Amplification	Reflection & Amplification	UDP/520	Exploits an old, unauthenticated routing protocol to send spoofed routing updates to flood or confuse the target network.
QOTD Reflection	Reflection	UDP/17	Spoofed requests to servers running the Quote of the Day service cause them to send a quote to the target IP.
Echo Reflection	Reflection	UDP/7	Spoofed requests to servers running the Echo Protocol cause them to reflect the received data back to the target IP.
Mirai UDP Flood	Flood (IoT-based)	UDP/Various	A specific pattern of UDP flood traffic generated by devices compromised by the Mirai malware.

3.2.4 The Global Botnet Infrastructure

A key element of any hyper-volumetric attack is a massively distributed source. This attack was launched from a botnet of over 122,145 unique source IP addresses, originating from 5,433 different Autonomous Systems (ASNs) across 161 countries.[11, 19] This extreme geographic and network distribution is a critical strategy to defeat localized or regional defenses. By sourcing the attack globally, the traffic arrives at the target from all directions, making it impossible to simply block a single country or network provider.

Analysis of the traffic origins revealed that nearly half of the attack volume came from just two countries: Brazil and Vietnam, each contributing approximately 25

3.3 Attacker Profile and Motive Analysis

Official attribution for such attacks is notoriously difficult and was not provided in this case. However, an analysis of the attack's characteristics allows for informed hypotheses regarding the attacker's profile and motive. There was no public claim of responsibility for the attack, nor were there any associated ransom demands reported.[20, 21] This makes a purely financial motive, such as extortion, unlikely.

The evidence points toward two primary possibilities:

1. **Capability Demonstration:** The attack may have been a showcase by a major DDoS-for-hire service or a botnet operator. In the competitive underground market, demonstrating the ability to launch a record-breaking attack that can bypass even a top-tier mitigation provider is the ultimate advertisement. Such an event would serve to attract high-paying customers seeking the most powerful "stresser" services available.

2. **State-Sponsored Probe:** The scale, technical sophistication, and targeting of critical internet infrastructure are all hallmarks of a potential state-sponsored operation. In this scenario, the attack would not be intended to cause a sustained outage but rather to act as a probe. A nation-state actor could use such a short, intense burst to test the defensive capabilities and response times of a major infrastructure provider like Cloudflare. The data gathered—such as detection thresholds, mitigation latency, and defensive capacity—would be invaluable intelligence for planning future cyberwarfare campaigns.[21] The complex choreography and immense resources required lend significant weight to this hypothesis.

3.4 Impact Assessment

A critical aspect of this case study is the distinction between the attack's potential impact and its actual impact.

- **Potential Impact:** An unmitigated 7.3 Tbps attack would have been catastrophic for the targeted hosting provider. It would have completely saturated their network links, resulting in a total service outage for them and all of their downstream customers. The financial losses from downtime, customer churn, and emergency remediation, combined with the severe reputational damage, would have been immense.

- **Actual Impact:** Due to the successful and fully automated mitigation, the actual impact was minimal.[14, 22] The attack was detected and blocked at Cloudflare's network edge, and legitimate traffic continued to flow to the hosting provider's infrastructure without significant degradation. The entire event was handled without the need for human intervention, alerting, or incident escalation, demonstrating the efficacy of a mature, automated defense architecture.

3.5 Defensive Strategies and Mitigation Analysis

The successful defense against this record-breaking attack was not accidental; it was the result of a specific, modern architectural approach to DDoS mitigation. The defense rested on two core pillars: a globally distributed architecture and intelligent, real-time automation.

3.5.1 The Role of Anycast Architecture

The foundational technology that made this defense possible is a global Anycast network. With Anycast, the target's IP prefix is announced from every one of Cloudflare's data centers simultaneously. This means that attack traffic from the 122,000+ source IPs was not directed to a single location. Instead, it was routed to the nearest of 477 data centers across 293 cities worldwide.[11, 14]

This geographically distributed absorption is the only viable method for handling a multi-terabit attack. No single data center or scrubbing appliance, regardless of its capacity, could withstand the full 7.3 Tbps onslaught. By distributing the attack across hundreds of locations, the load on any single point was reduced to a manageable level. For example, a data center in Brazil absorbed the traffic from the Brazilian portion of the botnet, while a data center in Vietnam handled the Vietnamese traffic. This strategy uses the distributed nature of the attack against itself.

3.5.2 Automated Detection and Response

The second critical pillar was the complete automation of the detection and mitigation process. For a 45-second burst attack, human-in-the-loop response times are wholly inadequate. The defense was handled by Cloudflare's autonomous systems, primarily a software daemon named 'dosd' (denial of service daemon) that runs in every data center.[14] The process unfolds in milliseconds:

1. **Sampling:** As packets arrive at the network edge, a statistical sample is diverted for analysis without impacting the flow of traffic.

2. **Analysis:** The 'dosd' engine, a sophisticated heuristic system, analyzes the headers of the sampled packets in real-time. It looks for anomalies and patterns indicative of a DDoS attack, such as malformed packets, unusual protocol usage, or a sudden surge of traffic from disparate sources to a single destination.

3. **Fingerprinting and Rule Generation:** Once 'dosd' identifies an attack, it generates a "fingerprint"—a unique signature that accurately describes the malicious packets while excluding legitimate traffic. Based on this fingerprint, it compiles a mitigation rule.

4. **Deployment:** This mitigation rule is deployed as an extended Berkeley Packet Filter (eBPF) program. eBPF allows for custom programs to be run directly within the Linux kernel's networking stack. This is extremely efficient, as it allows the malicious packets to be identified and dropped at line-rate, as soon as they are received by the network card, before they can consume valuable CPU, memory, or application resources.[14, 22]

5. **De-provisioning:** Once the attack subsides, the system automatically removes the eBPF rule to ensure no legitimate traffic is inadvertently blocked.

This entire cycle occurred autonomously in hundreds of data centers around the world, effectively neutralizing the threat in real-time without human intervention. This highlights a fundamental economic asymmetry in modern cyber warfare: the attack was launched using a vast number of cheap, insecure, and low-power IoT devices, yet it required a multi-billion dollar, globally distributed, and highly sophisticated automated defense system to defeat it.[11, 14] This disparity underscores the collective cost the internet bears for the lack of baseline security in connected devices.

4 Conclusion: Strategic Imperatives for DDoS Resilience in the Hyper-Volumetric Era

The analysis of the 2024-2025 DDoS threat landscape, and the 7.3 Tbps attack in particular, yields several critical conclusions that should inform the strategic posture of any organization reliant on internet availability.

First, the nature of the DDoS threat has fundamentally changed. Hyper-volumetric attacks are no longer a theoretical risk but an operational reality. The industrialization of the cybercrime ecosystem through DDoS-for-hire services and the escalating arms race in botnet technology mean that multi-terabit attacks are accessible to a wide range of actors. Consequently, DDoS must be elevated from a mere operational nuisance to a strategic threat to business continuity, customer trust, and, when directed at core services, national infrastructure.[1, 8]

Second, the success of the defense against the 7.3 Tbps attack provides a clear blueprint for modern DDoS resilience. Legacy, on-premises defense strategies, which rely on physical appliances with finite capacity, are obsolete in this new era. They are easily overwhelmed by multi-terabit floods. Resilience is not achieved by simply purchasing more bandwidth or a bigger "black box." It requires a fundamental architectural shift.

The strategic imperatives for achieving this resilience are clear:

- **Embrace Global Scale:** Defense must be as distributed as the threat. Leveraging a cloud-based, globally distributed scrubbing architecture with an Anycast network is essential to absorb and mitigate attacks close to their source, preventing massive traffic volumes from ever reaching the target's infrastructure.
- **Prioritize Intelligent Automation:** The velocity of modern attacks, especially short-burst campaigns, has rendered manual, human-in-the-loop responses ineffective. Defense must be predicated on real-time, autonomous systems that use heuristic and behavioral analysis to detect and mitigate threats in seconds, without human intervention.
- **Invest in Proactive Intelligence:** Defensive postures must shift from reactive to proactive. This involves participating in threat intelligence sharing ecosystems to gain visibility into emerging botnets and attack infrastructure. Services that identify and help purge malicious infrastructure before it can be weaponized are a critical component of reducing the overall threat surface.[23]
- **Implement Defense-in-Depth:** A resilient strategy must account for the full spectrum of threats. It requires a multi-layered defense capable of handling both brute-force volumetric floods at the network layer and sophisticated, low-and-slow attacks targeting the application layer.

The 7.3 Tbps event was not an endpoint but a milestone on an escalating trajectory. As attackers continue to innovate and scale their capabilities, organizations that fail to adopt a modern, automated, and globally-scaled defensive architecture will find themselves increasingly vulnerable in an era where the digital siege is a daily reality.