AI/ML in Nokia Deepfield Defender

Harnessing machine learning for adaptive, extensible and automated protection against DDoS attacks

Application note

Abstract

The Nokia Deepfield DDoS security solution uses the latest advances in big data network analytics, artificial intelligence (AI) and machine learning (ML), internet-scale mapping and programmable router silicon to provide communications service providers (CSPs), network operators and cloud builders with unparalleled DDoS protection and security automation capabilities.

This application note describes how Nokia Deepfield Defender, a cornerstone of the Deepfield DDoS security solution, uses AI/ML technology to help service providers and network operators improve and automate their DDoS security, ensure uninterrupted continuity of services and operations, and better protect their infrastructure and customers against complex and growing network security threats.

Contents

Introduction

The Nokia Deepfield DDoS security solution leverages the latest advances in big data network analytics, artificial intelligence (AI), machine learning (ML), internet-scale mapping and advanced router technology with programmable router silicon (e.g., Nokia SR/SXR routers and Nokia Deepfield 7750 DMS-1-24D and Juniper MX series) to deliver unparalleled DDoS detection, mitigation, automation and reporting capabilities.

This application note describes how Nokia Deepfield Defender, a cornerstone of the Deepfield DDoS security solution, uses AI and ML technology to help service providers and network operators better protect their infrastructure and customers against DDoS attacks that are growing in scale, frequency and sophistication. It includes information on:

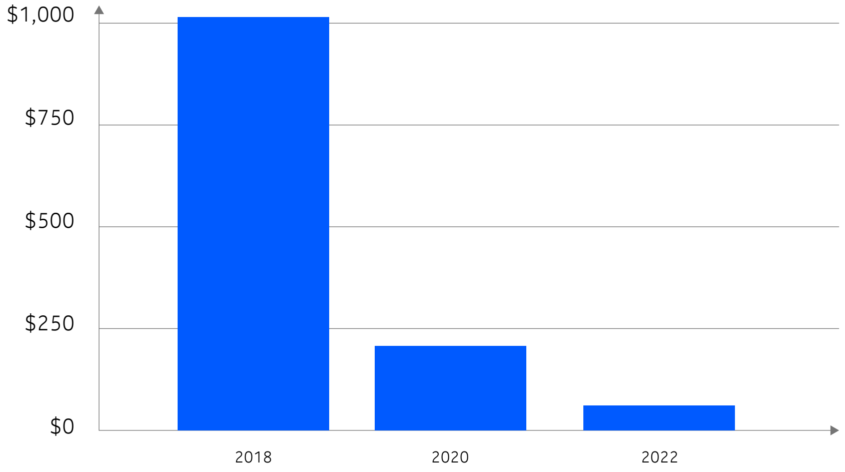
- How the DDoS threat landscape is changing and the new challenges it presents
- Why a new approach to DDoS security—one that includes network- and internet-based intelligence—is required
- Why some of the emerging AI and ML technology solutions fall short of addressing these needs
- How the Nokia Deepfield DDoS solution's unique approach to AI and ML optimizes the response to DDoS attacks and improves the overall efficiency of protection
- A real-world performance evaluation that shows how the Nokia DDoS security solution outperforms other commercially available solutions.

For a more detailed look at the Nokia Deepfield DDoS security solution and its components, please visit the Deepfield Defender web page and check out our product-related collateral.

Evolving DDoS threats pose greater network dangers

The combination of explosive growth in IoT and cloud computing, an increasingly lucrative extortion market and evolving nation-state threats represents an inflection point in the DDoS landscape.

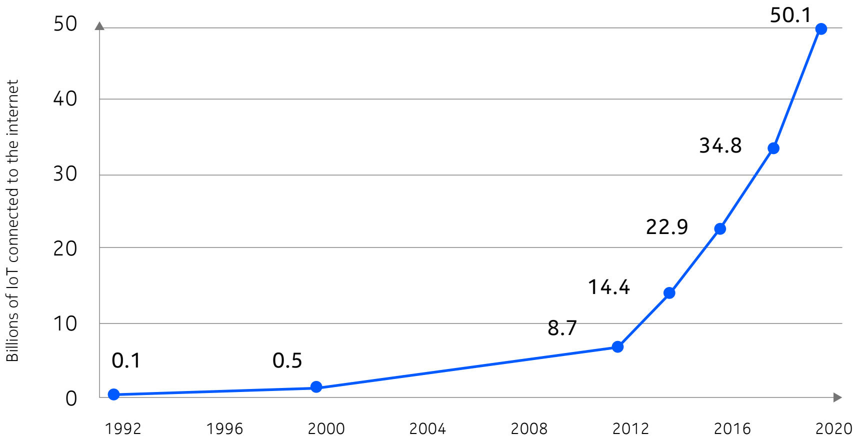
The days of DDoS attacks launched using poorly written shell scripts running on compromised home computers with limited bandwidth are over. Most DDoS traffic now originates from industrial-scale infrastructure operated by nation-state actors or criminals or from sophisticated commercial booter web applications that offer a menu of attacks at competitive prices ranging from $US50–$US500, most often paid in cryptocurrency. As shown in Figure 1, the prices for launching DDoS attacks that can cause significant impacts are dropping. Some of the sites offering these services now even have a free tier.



Source: www.zero.bs blog

Figure 1: Evolution of the daily average price for launching DDoS, 2018–2022

The influx of insecure IoT devices into the market is exacerbating the DDoS threat. Consumers and businesses are introducing vulnerable devices at an alarming rate, effectively doubling or tripling the pool of exploitable devices annually. Many of these devices are equipped with high-speed internet connections and operate on full-stack Linux, making them potent tools for malicious actors seeking to launch DDoS attacks.



Source: https://www.comptia.org/content/research/sizing-up-the-internet-of-things

Figure 2: Exponential growth of new IoT devices, 1992–2020

Also, DDoS attacks are becoming larger and significantly more challenging to detect and mitigate. In the past, most DDoS attacks utilized some form of synthetic traffic generation (i.e., "spoofing") to reflect/amplify or generate direct-path floods to victims from a relatively small number of IP header modification (IPHM)-friendly hosting providers or compromised home computers. Scrubbers were used to effectively mitigate these attacks because synthetic traffic usually contained distinguishable header or payload features (e.g., poorly randomized headers, patterns in attack payload) or otherwise failed basic scrubber protocol authentication.

In marked contrast to the pre-IoT era, most of today's largest DDoS attacks exclusively leverage large-scale botnets. Unlike their predecessors, which used synthetic amplification and flooding techniques, these botnets use valid (non-spoofed) IP addresses, full TCP/IP stacks, legitimate operating system (OS)-generated protocol headers, correct checksums and payloads carefully crafted to match the statistical distributions that can be seen in normal application traffic (e.g., web agent, form fields). Many of the botnets can even pass CAPTCHA challenges.

Most DDoS solutions deployed by service providers and network operators today use xFlow (NetFlow, sFlow, IPFIX, etc.) telemetry with one or more common detection algorithms, such as static bits-per-second (bps) or packets-per-second (pps) thresholds, traffic ratios, time series analysis or variance from pre-calculated traffic baselines. While many academics and vendors have proposed more advanced techniques, such as entropy (analysis of the randomness of certain attributes of malicious traffic), high false-positive rates and a lack of explainability often limit their operational applicability.

The challenge for earlier generations of ML solutions based on xFlow or payload samples is that ML requires meaningful features for extraction (i.e., distinguishable header or payload characteristics). These features are not present in most of today's large-scale botnet DDoS attacks.

The Nokia Deepfield DDoS security solution

Nokia brings together petabyte-scale big data IP analytics (provided by Deepfield Defender) with the power of advanced network routers (such as Nokia 7750 Service Routers and Nokia 7730 Service Interconnect Routers) and next-generation DDoS mitigation systems (such as the 7750 Defender Mitigation System) to fight DDoS with unprecedented scale, speed and efficiency.

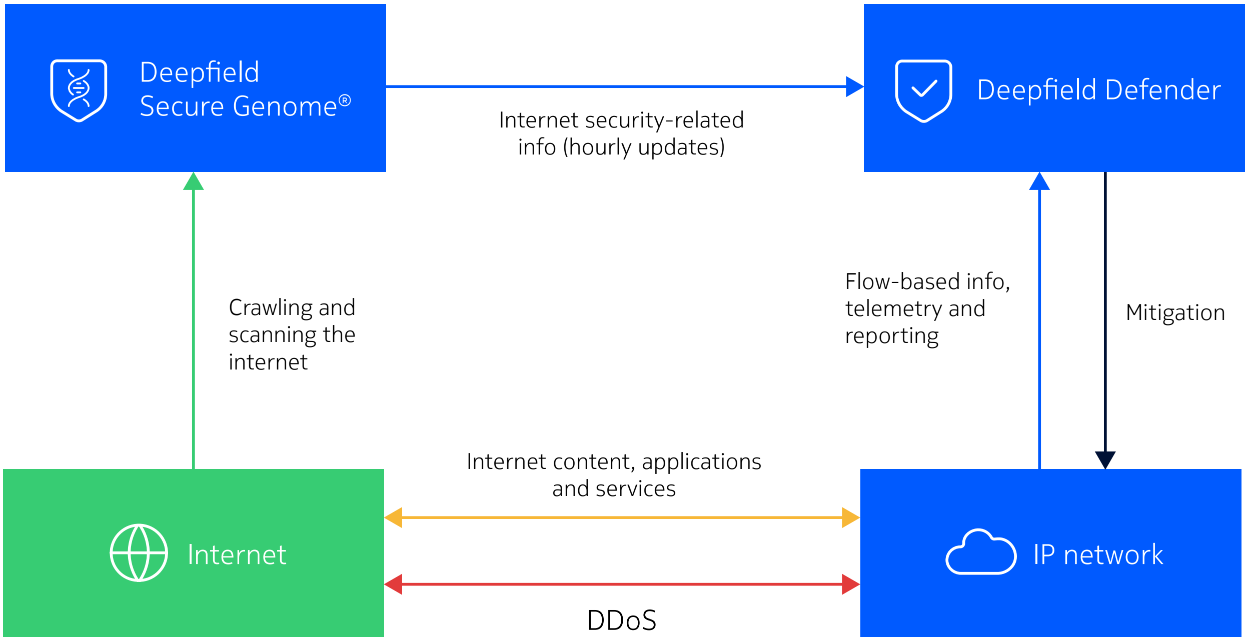

Deepfield Secure Genome®
Deepfield Defender
Internet security-related info (hourly updates)
Crawling and scanning the internet
Flow-based info, telemetry and reporting
Mitigation
Internet content, applications and services
Internet
IP network
DDoS

Figure 3: Nokia Deepfield DDoS security solution

Deepfield Defender is a software application that combines network data (telemetry, DNS, BGP, etc.) with the patented Nokia Deepfield Secure Genome®, a continuously updated cloud-based data feed that tracks the security context of the internet. Secure Genome has detailed visibility into more than 5 billion IPv4 and IPv6 addresses. It tracks all internet traffic globally, arranges it into 30-plus categories, and uses more than 100 ML rules to automatically classify and precisely allocate applications and flows into security-related traffic types and categories. As a result, Secure Genome “knows” the intricate security details of the internet, including specific information about prior attacks, insecure servers and compromised IoT devices that can be used for DDoS attacks.

Defender correlates the internet security knowledge from Secure Genome with the telemetry information obtained from the network to detect DDoS attacks faster and more accurately and drive agile network-based mitigation using advanced IP routers such as Nokia FP4/FP5/FPcx-based IP routers or a dedicated DDoS mitigation system such as the Nokia 7750 Defender Mitigation System (7750 DMS-1).

Figure 4 shows a graphical representation of how Deepfield Defender uses AI/ML.

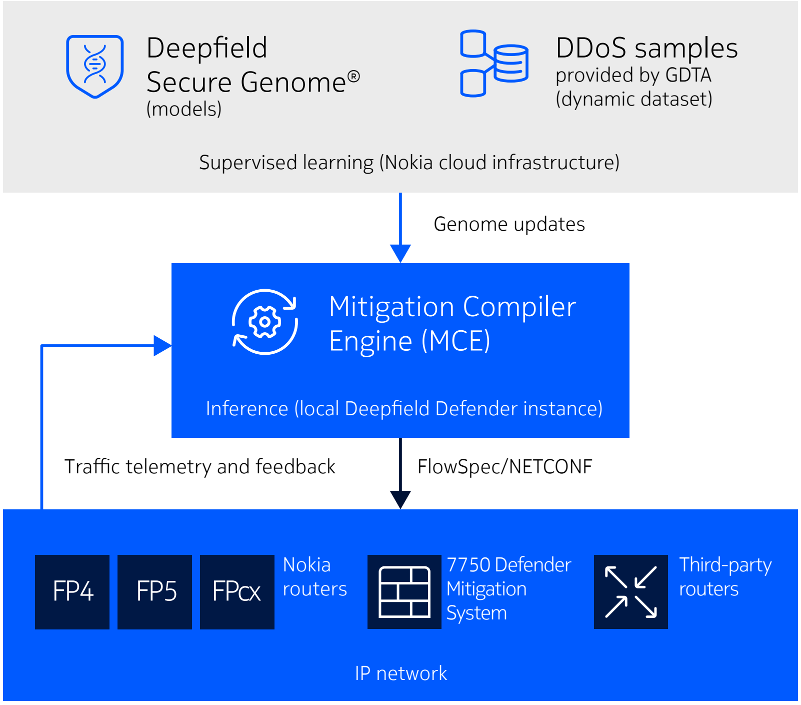


Figure 4: Use of AI/ML in Deepfield Defender

Using advanced AI/ML algorithms, Defender calculates the optimal mitigation strategy for a particular DDoS attack or multiple concurrent attacks. Working in real time, it instructs the routers or 7750 DMS-1 to apply security-related filters or access control lists (ACLs) and neutralize the attack.

Defender provides the foundation for Nokia Deepfield's next-generation DDoS detection and mitigation solution. Leveraging rich telemetry and IP network programmability, this DDoS solution offers significant advantages over legacy appliance- or DPI-based approaches. These include better scalability, more accurate DDoS detection with fewer false positives, and faster, more efficient, and more cost-effective DDoS mitigation. The result is holistic, 360-degree DDoS security that meets the demands of the 5G, cloud and IoT era.

The sections that follow describe how the Deepfield DDoS security solution uses ML to detect and mitigate DDoS attacks, covering the main aspects of supervised learning and model training versus feature extraction and inference. They also describe how the solution uses the large-scale Nokia Deepfield DDoS Library as its foundational dynamic dataset and how it uses the Deepfield Cloud Genome® and Secure Genome to significantly expand the feature set beyond previously available xFlow fields. In addition, the sections outline how the dynamically interpreted Deepfield Model Language (DML) can be used in detection and mitigation language models to ensure extensibility and protection against future attacks.

The Nokia Deepfield approach to AI: Data sets, algorithms and models

This section provides an overview of how Nokia Deepfield addresses three important aspects of AI implementation: high-quality data sets, learning and models.

Bigger, better data sets

Enhancing the speed and accuracy of DDoS detection and increasing the agility and precision of DDoS mitigation are crucial goals in network security. AI technology plays a pivotal role in achieving these goals.

The combination of rising volumes of DDoS traffic, novel attack techniques and the exponentially growing universe of insecure endpoints and systems that can be co-opted into botnets and used to launch DDoS attacks creates a vast threat surface that needs to be constantly monitored. The threat landscape, which is increasing in size and complexity, calls for better security solutions that can correlate vast amounts of internet security-related data with network-related data to obtain near-real-time security information to drive real-time decisions about whether certain network flows are threats, attacks or "good traffic."

Data is the lifeblood of AI, especially when it comes to AI for DDoS security. Access to high-quality DDoS security data is critical for training AI models to do their jobs confidently. To avoid relying on insufficient or misrepresented data sets that can lead to invalid results and "house-of-cards" failures (based on third-party research or data), we knew we needed to start with a super-large, highly relevant and highly confident data set. So, we built one of the world's largest DDoS attack libraries: we collected, preprocessed and analyzed more than 10,000 real-world DDoS attacks collected from "the wild," stored them in our Deepfield DDoS Library, and then used the library to train our detection models. We added features such as xFlow details, Transmission Control Protocol (TCP) flags, time-to-live (TTL) and time series analysis to help us achieve faster and more accurate detection.

Combined with our internal testing tools, our DDoS library also helps us replay DDoS attacks in a controlled environment, where we can test the efficiency of AI inference–DDoS mitigation strategy and execution. We constantly update the library with new attacks based on new DDoS vectors and new ranges of source and target IP addresses. This library is the cornerstone of our supervised training.

Supervised and unsupervised learning

We use our high-quality labeled data to supervise the training of our AI models. This training is greatly enhanced by the Genome data set, which we have expanded and maintained for more than ten years. Secure Genome provides a detailed, global internet security context. We use it as an input for training our models and as a data feed that provides up-to-date security details to Deepfield Defender deployments.

For Defender deployments, Secure Genome facilitates unsupervised learning, helping Defender make better decisions about what is good traffic and what are malicious threats and DDoS attacks. These decisions have a significant impact on the network infrastructure and customers, so it's important to consider the wider internet and network context rather than solely focusing on traffic thresholds or anomalies. Secure Genome provides this internet-wide security context, while flow-based network telemetry provides information about a particular network context (deployment).

Explainability of our AI algorithms and models

Selecting the right AI algorithms and models is essential. While large language models (LLMs) such as generative pre-trained transformers (GPTs) have gained popularity and are being implemented in many places, they can be complex and difficult to understand. Often, their “intelligence” comes with logic and rationale weights that are hidden from end users.

For a critical area such as DDoS security, where decisions must be traceable, explainability is a key factor to consider for any implementation of AI. We understood the challenges and limitations of applying LLMs, neural networks and fuzzy logic and the need for outcomes to be fully traceable and explainable. Our approach to creating, optimizing and using AI algorithms and models has focused on meeting a desired outcome—faster and more accurate detection and more agile and precise mitigation—rather than an intrinsic property of the data.

To overcome the limitations of some machine learning models, including a lack of reasoning and cognitive grouping, we embraced multilayered decision tree-based and deep learning models. These models allow us to constantly evaluate and improve their performance and understand why a particular detection or mitigation decision was made.

Rigorous evaluation and validation are crucial for assessing the performance of any AI solution. The “Evaluating the performance and efficiency of a DDoS detection and mitigation” section of this application note details the metrics and benchmarks we used to measure our models’ performance based on real-world DDoS attacks.

Legal considerations

AI relies on big data, and applying AI involves dealing with large amounts of sensitive (and sometimes confidential) data. We paid the utmost attention to important issues such as bias and fairness in the way internet security context and network data are collected and used and how AI algorithms process these data and make inference decisions.

In this process, we observe and comply with major privacy and regulatory requirements, just as we do with our DDoS threat alliance program.

The threat landscape and AI technology are constantly evolving. We aim to keep our solution up-to-date with the latest advancements, addressing network security domain-specific considerations as well as additional operational and legal considerations relevant to applying AI in this area.

Dynamic dataset: Deepfield DDoS Library

The first challenge in any supervised machine learning approach is to get high-quality, relevant and confident training data. In the context of DDoS protection, the key questions relating to training data are:

- **How** do you obtain DDoS attack (and peacetime) samples?
- What **features** do you select in these samples?
- How do you **evaluate the accuracy** of your detection model(s) against the samples?

Motivated by this challenge, we created one of the industry's largest datasets of real-world DDoS attacks to serve as the foundation for developing our ML models. The Deepfield DDoS Library includes more than 10,000 (and growing) geographically, topologically and commercially diverse attacks representing every known DDoS vector, including amplification/reflection, flooding, and botnet and application-layer attacks.

We continuously update this dataset with samples from our internal Deepfield sources (including honeypots), traces from "dark web" commercial booters and real-time contributions from members of the Nokia Global DDoS Threat Alliance (GDTA) program.

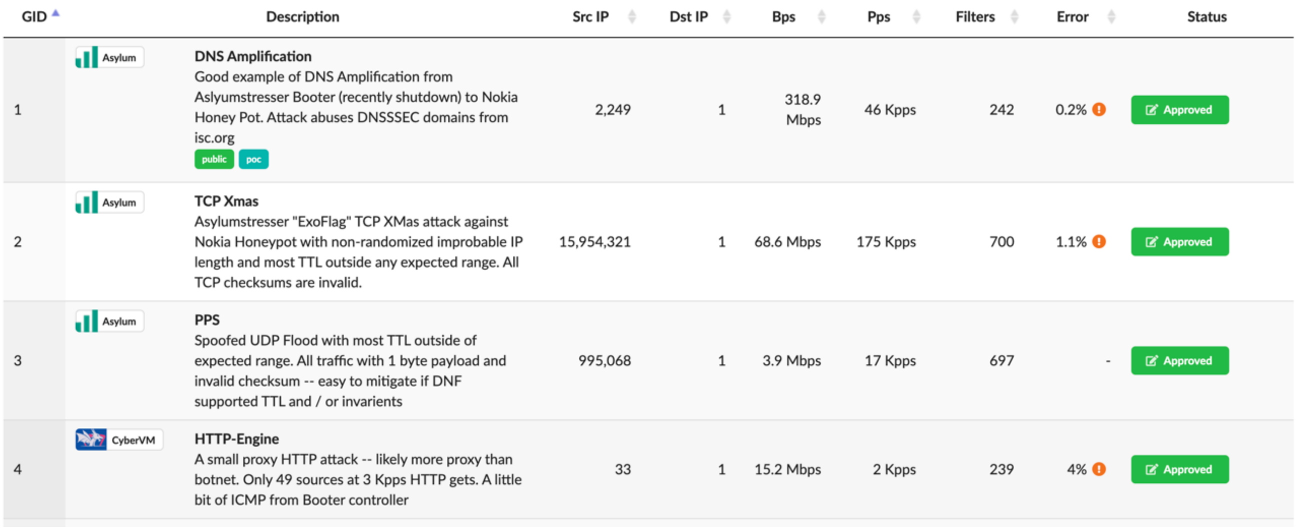
The GDTA plays a key role in maintaining the Deepfield DDoS Library's completeness and timeliness. This opt-in, membership-based organization enables participating Deepfield customers to share information about DDoS threats with us. By doing so, they help improve the Secure Genome data feed, which provides additional internet-related security context.

The Deepfield DDoS Library contains samples ranging from low-volume HTTPS application attacks to multi-terabit amplification and botnet floods. The underlying sample formats vary based on the sample's origin and include xFlow packet capture (PCAP), raw PCAP and Deepfield-augmented flow Apache Parquet files. All samples in the library are automatically classified and manually verified for DDoS detection accuracy.

We maintain a large, dedicated security analyst team that continuously reviews models and augments sample classification. In addition, many samples submitted to the GDTA come with analysis and classification performed by CSP security teams.

Figure 5 shows some samples in this dataset, including several now-defunct commercial booters [1]. Note that each sample is assigned a relative error or accuracy rate based on the validation of ML classified flows by the CSP security team or a Nokia security analyst.

The error rate represents a per-flow, upper bound on DDoS classification. The error bound in the "HTTP-Engine," for example, includes individual flows where the model lacked sufficient features to accurately discriminate between malicious botnet flows and otherwise legitimate HTTP flows.



GID ▲		Description	Src IP	Dst IP	Bps	Pps	Filters	Error	Status
1	Asylum	**DNS Amplification** Good example of DNS Amplification from Aslyumstresser Booter (recently shutdown) to Nokia Honey Pot. Attack abuses DNSSSEC domains from isc.org public poc	2,249	1	318.9 Mbps	46 Kpps	242	0.2% ❗	Approved
2	Asylum	**TCP Xmas** Asylumstresser "ExoFlag" TCP XMas attack against Nokia Honeypot with non-randomized improbable IP length and most TTL outside any expected range. All TCP checksums are invalid.	15,954,321	1	68.6 Mbps	175 Kpps	700	1.1% ❗	Approved
3	Asylum	**PPS** Spoofed UDP Flood with most TTL outside of expected range. All traffic with 1 byte payload and invalid checksum -- easy to mitigate if DNF supported TTL and / or invarients	995,068	1	3.9 Mbps	17 Kpps	697	-	Approved
4	CyberVM	**HTTP-Engine** A small proxy HTTP attack -- likely more proxy than botnet. Only 49 sources at 3 Kpps HTTP gets. A little bit of ICMP from Booter controller	33	1	15.2 Mbps	2 Kpps	239	4% ❗	Approved

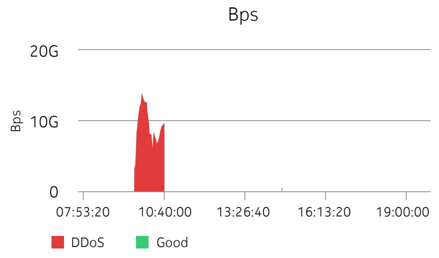
Figure 5: Deepfield DDoS Library samples

Feature engineering: Enriching xFlow with Deepfield Secure Genome

As mentioned in previous sections, conventional DDoS detection methodologies primarily leverage xFlow telemetry attributes such as bytes, packets, IP addresses, ports, protocols and TCP flags. Historically, these xFlow fields in the IP header have provided sufficient features for classifying simple attack patterns. For example, Domain Name System (DNS) amplification attacks can often be classified using time series analysis and atypical rates of large DNS packet sizes. At the same time, SYN floods are often detectable through an atypical ratio of SYN to other TCP flags.

Figure 6 shows an example of a recent DNS amplification attack against a subscriber using a 100 Mbps internet service from one of our CSP customers. At the peak of the attack, the subscriber received more than 14 Gbps of traffic from 65,000 unique source IP addresses. Coincident with the attack, the subscriber also exchanged “legitimate” traffic (i.e., non-attack traffic) with 700 source IP addresses, including content delivery networks (CDNs), DNS servers and gaming traffic.

The attack is readily apparent in the graph on the left. Similarly, sample flows on the right show atypically large DNS packet sizes and flow repetition. Any one of several features in xFlow, time series or baseline trends, or even static thresholds might be sufficient for manual and model- or ML-based detection strategies.

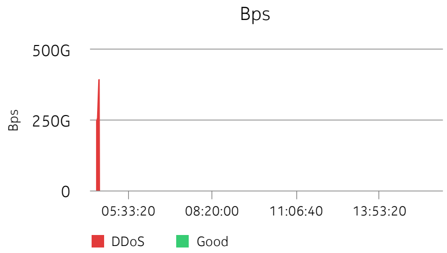


timestamp	protocol	topflags	addr.src	port.src	addr.dst	port.dst	max_ttl	bytes
1701589400	udp		8.28.0.9	53	xx.xx.53.4	33897	254	1500
1701589430	udp		8.28.0.9	53	xx.xx.53.4	53086	254	1500
1701589440	udp		8.28.0.9	53	xx.xx.53.4	52280	254	1500

Figure 6: A 14 Gbps DNS amplification attack against a CSP’s subscriber. Flow includes atypical DNS packet sizes and repetition in flow source–destination tuples.

Models that depend exclusively on xFlow features tend to exhibit elevated error rates, both in false positives and false negatives, particularly when confronting modern IoT botnet attacks or sophisticated synthetic traffic floods.

Figure 7 depicts a recent 400 Gbps quick UDP internet connection (QUIC) attack targeting a popular web server within a CSP’s network. This attack peaked with over 5,000 distinct source IP addresses encompassing a diverse array of networks (BGP Autonomous System Numbers, ASNs) and geographic origins. A closer examination of the QUIC packets directed at the victim servers revealed that all packets contained valid, well-structured QUIC requests characterized by statistically typical payloads, including aspects such as requested content and user agents. Such nuances in attack patterns underscore the limitations of relying solely on traditional xFlow-based detection models.



timestamp	protocol	topflags	addr.src	port.src	addr.dst	port.dst	max_ttl	bytes
1701589400	udp		82.79.29.233	38725	xx.xx.23.108	443	57	1350
1701589430	udp		61.221.247.77	45001	xx.xx.23.108	443	60	1350
1701589440	udp		81.196.92.19	36818	xx.xx.53.4	443	55	1350

Figure 7: Botnet QUIC attack

Traditional time series or baseline detection methods are adept at identifying statistical anomalies in data streams (e.g., the significant red spike on the left in Figure 7). However, they often struggle to differentiate between normal fluctuations, such as those resulting from a new software or game release, and genuinely malicious activities. A key limitation is their inability to discern individual legitimate and malicious flows within a mixed sample.

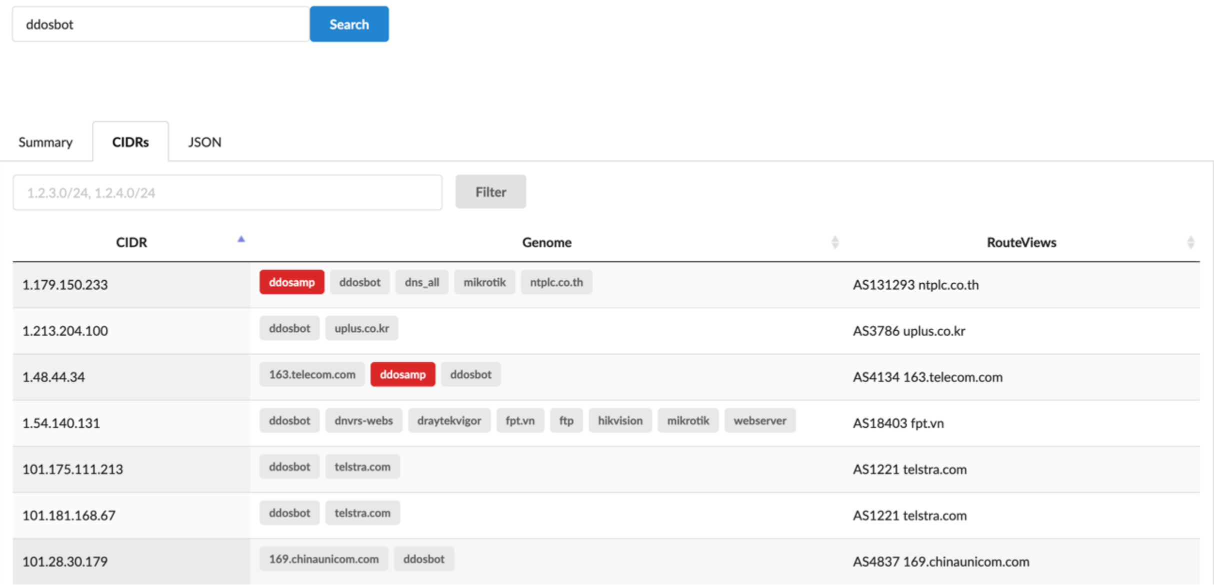
We will return to this QUIC attack example after first describing how we use the Cloud Genome and Secure Genome to significantly expand the xFlow feature set.

Deepfield Genome

Nokia Deepfield Genome technology represents more than a decade of research and development for Cloud Genome mapping technology. See the Wired Magazine article for a non-technical overview of Genome or more detailed information in our US patent [2] [3].

Deepfield Genome combines real-time crawling of the entire active IPv4/IPv6 internet address space with third-party data feeds and our ML technology to map internet applications, infrastructure and topology.

Secure Genome includes information about millions of IoT devices, misconfigured and vulnerable servers, malicious hosting sites and DDoS botnets. Deepfield Defender uses Secure Genome definitions (currently updated hourly) with custom-built ML algorithms focused on traffic behavior to provide real-time tracking of botnets and other emerging threats in CSP networks (Figure 8).



CIDR	Genome	RouteViews
1.179.150.233	ddosamp ddosbot dns_all mikrotik ntplc.co.th	AS131293 ntplc.co.th
1.213.204.100	ddosbot uplus.co.kr	AS3786 uplus.co.kr
1.48.44.34	163.telecom.com ddosamp ddosbot	AS4134 163.telecom.com
1.54.140.131	ddosbot dnvrs-webs draytekvigor fpt.vn ftp hikvision mikrotik webserver	AS18403 fpt.vn
101.175.111.213	ddosbot telstra.com	AS1221 telstra.com
101.181.168.67	ddosbot telstra.com	AS1221 telstra.com
101.28.30.179	169.chinaunicom.com ddosbot	AS4837 169.chinaunicom.com

Figure 8: Secure Genome tracking DDoS bots

Figure 9 revisits the QUIC attack example discussed earlier. Genome identified the 'dahuatech' label for a subset of the source IP addresses, offering multiple additional insights such as IoT device types, operating systems, subscriber categories, customer premises equipment (CPE) identification, active web technologies or services, and historical data.

As we will explore in subsequent sections, Defender's decision tree and ML models use these Genome-derived features to enhance accuracy and reduce false negatives and positives.

timestamp	protocol	topflags	addr.src	port.src	addr.dst	port.dst	max_ttl	bytes	Genome
1701589400	udp		82.79.29.233	38725	xx.129.23.108	443	57	1350	dahuatech
1701589430	udp		61.221.247.77	45001	xx.129.23.108	443	60	1350	dahuatech
1701589440	udp		81.196.92.19	36818	xx.129.23.108	443	55	1350	dahuatech

Figure 9: QUIC attack flow records with Secure Genome and “dahuatech” feature

While a comprehensive exploration of Deepfield Genome-enriched features is outside the scope of this application note, it is instructive to highlight some key features commonly used for detection. These include:

- **‘cve’**, which identifies endpoints running software with known critical exploits
- **‘ddosbot’**, which marks endpoints observed in confirmed DDoS attacks within the past 48 hours
- **‘ddosamp’**, which identifies misconfigured servers used in amplification DDoS within the past 48 hours
- **‘cpe_type’**, which categorizes several hundred types of CPE, including as Wi-Fi and Passive Optical Network (PON) residential gateways
- **‘proxy’**, which denotes endpoints functioning as known proxies
- **‘tor’**, which indicates connections through the Tor network, among many others.

This expanded feature set enables a more nuanced and robust analysis of network security threats.

Extensibility with the Deepfield Model Language

One crucial aspect that makes Deepfield Defender a powerful tool for efficiently addressing new types of attacks is the set of models it uses for detection, mitigation and data manipulation.

The Deepfield Model Language, or DML, draws inspiration from popular data science and analytics languages such as Pandas, R and NumPy—with significant ML and DDoS domain-specific additions.

DML can naturally parse and operate on xFlow fields, as well as those enriched from Genome variables, and supports time series, clustering and statistical operators. It includes models for well-known DDoS vectors (e.g., amplification, spoofing, botnet) and novel attack vectors.

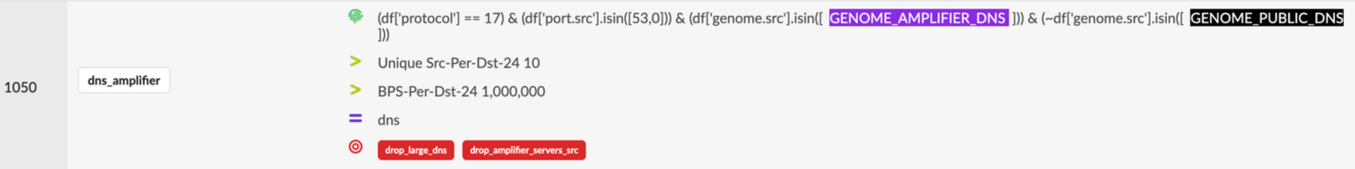
We leverage a continuous stream of data samples in the Deepfield DDoS Library to facilitate daily supervised learning within our cloud infrastructure. In contrast with isolated learning and detection in individual CSP environments, our approach provides a global perspective on DDoS threats. It captures a wider array of attack vectors, including some that may be novel to any given individual CSP.

The Genome database and models from the cloud component described above are downloaded to Defender instances within CSP data centers on an hourly basis. Defender combines the offline DML-based models with xFlow, Genome lookups, and run-time DML rules and statistical operators.

DML includes six key components:

1. **Deepfield Cloud Genome and Deepfield Secure Genome definitions:** Genome maintains more than 30,000 definitions matching all major commercial internet content, services and application types (e.g., Netflix, Salesforce, LINE), as well as IoT device types, CDN servers, cloud and hosting companies, carrier networks, compromised endpoints, known Common Vulnerabilities and Exposures (CVE) and more.
2. **Deepfield Secure Genome groups:** In manual processing and during supervised learning, Defender groups sets of Genome tags and classless interdomain routing (CIDR) ranges into blocks with common features. For example, most CDN and cloud providers use one of several common TTL defaults (namely, Microsoft Windows or Linux settings). Amazon CloudFront, Cisco devices and many firewall endpoints use less common TTL defaults. Collectively, supervised and unsupervised learning help to group and dynamically assign Genome group membership based on peacetime and DDoS samples from our dynamic dataset.
3. **Time series and anomaly analyses:** Models encompass standard time series calculations such as bps, pps, rate of change, baselines and various thresholds. Additionally, metrics such as cardinality (the count of unique addresses, ports, TTLs, etc.) and entropy are used as input to the models as part of the time series analysis.
4. **Clustering and association algorithms:** During supervised learning and dynamic run-time (inference), Defender uses standard clustering and association algorithms to identify patterns in the sample data. For example, a remote Netgear AC1200 sending a few SYN packets to a protected server is not an atypical occurrence, but a 10 Mpps SYN flood from 300 endpoints (99 percent of which are Netgear AC1200 with known CVE) might trigger the Defender model for a SYN IoT attack. We observe that while many DDoS attacks involve thousands of attacking sources, botnet elements within a given attack frequently share common CVE and device type attributes.
5. **Models:** Defender maintains models for every known DDoS attack vector and mitigation strategy, as well as novel attack vectors. Each model includes a match component that covers xFlow fields, Genome features and dynamic time series calculations (e.g., bps, pps, cardinality, entropy). As described above, the models and the Genome signature database are downloaded to CSP Defender deployments and evaluated against xFlow telemetry while running inference.

6. **Decision tree:** Models serve as input to detection and mitigation decision trees. Deepfield uses the tree structure to prioritize detection and mitigation accuracy, as well as to evaluate DDoS protection strategies when satisfying constraints. For example, a model that consumes xFlow and Deepfield Secure Genome data may detect 100,000 well-known DDoS amplifiers and ten legitimate DNS servers in an attack. While the most accurate mitigation strategy is to surgically block the 100,000 sources and permit the ten valid servers, edge and scrubber filter size constraints may make this strategy untenable. Based on filter table constraints (total number of filters, filters consumed by other attacks or operator policy), the decision tree performs constraint satisfaction to choose rate-limiting large DNS packets over more expensive strategies. Figure 10 illustrates an example of a DNS decision tree for DNS amplification mitigation.



```
1050    dns_amplifier
        (df['protocol'] == 17) & (df['port.src'].isin([53,0])) & (df['genome.src'].isin([ GENOME_AMPLIFIER_DNS ])) & (~df['genome.src'].isin([ GENOME_PUBLIC_DNS
        ]))
        > Unique Src-Per-Dst-24 10
        > BPS-Per-Dst-24 1,000,000
        = dns
        drop_large_dns  drop_amplifier_servers_src
```

Figure 10: Example of a DNS amplifier decision tree

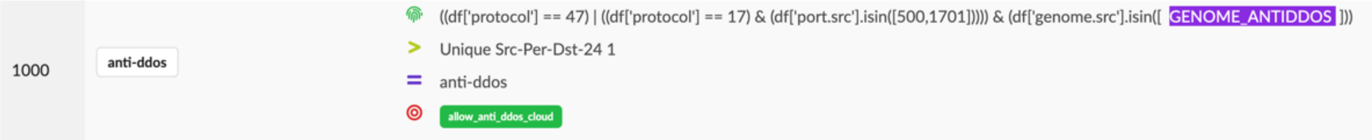
While a full explanation of DML is beyond the scope of this document, we provide a few examples below of how Deepfield Defender can use DML's unique capabilities to improve DDoS protection.

Example 1: Cloud DDoS protection awareness

Some enterprises opt for DDoS protection from CSPs and third-party cloud providers such as Akamai (Prolexic solution) and Cloudflare. These providers often use Generic Routing Encapsulation (GRE) tunnels, which can lead to significant, statistically anomalous traffic spikes during attack periods. These spikes could trigger false positives in detection systems based on thresholds or baselines. To avoid false positives, the Defender decision tree includes a specific cloud DDoS provider (**'anti-ddos'**) model.

Key aspects of the 'anti-ddos' model include static elements, such as matching protocol 47/GRE or IPsec traffic, and dynamic components, such as membership in the purple 'GENOME_ANTIDDOS' group. This group is dynamically evaluated during offline, supervised learning and real-time inference (operation) within Defender's CSP deployments. Supervised learning occasionally identifies unexpected GRE traffic patterns during attacks. These are atypical in that they don't match known botnet or spoofed DDoS GRE patterns and instead originate from cloud infrastructure, suggesting a commercial service rather than a malicious source.

Figure 11 presents the internal Deepfield security analyst's view of this rule. In the broader context of the Defender decision tree, flows identified as **'anti-ddos'** are, by default, exempt from further mitigation processes. As with all model behaviors, operators can override these defaults.



```
1000    anti-ddos
        ((df['protocol'] == 47) | ((df['protocol'] == 17) & (df['port.src'].isin([500,1701])))) & (df['genome.src'].isin([ GENOME_ANTIDDOS ]))
        > Unique Src-Per-Dst-24 1
        = anti-ddos
        allow_anti_ddos_cloud
```

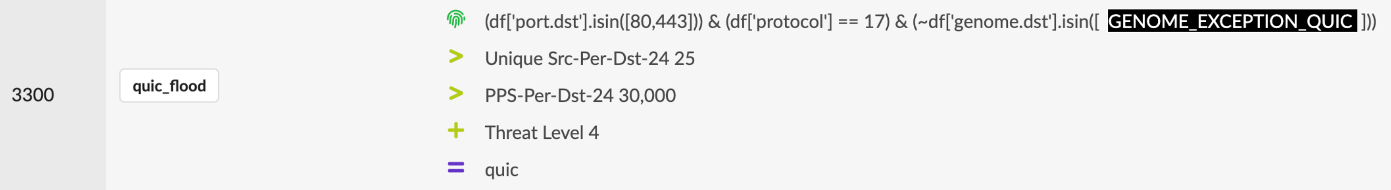
Figure 11: Defender decision tree model to avoid false positives for common cloud DDoS traffic during attacks

Example 2: QUIC support awareness

As the adoption of QUIC has increased among major webscale providers, so has its use by DDoS threat actors. This is particularly challenging for CSPs because restricting access to port 443, commonly used by QUIC, negatively impacts legitimate services. The Nokia Deepfield Emergency Response Team (ERT) has noted a significant proportion of DDoS traffic in recent years attributed to QUIC floods. CSPs need to mitigate these QUIC floods effectively while ensuring uninterrupted access to legitimate QUIC-based services such as TikTok, Google and Uber.

One of our approaches to detecting QUIC flood traffic involves a model that includes static and dynamic elements. The static aspect characterizes typical QUIC traffic, identifying it through features such as UDP and destination ports 80 and 443. We also incorporate dynamic evaluation through the **‘GENOME_EXCEPTION_QUIC’** group membership. This evaluation considers a variety of factors, including whether the destination address is a known QUIC server or a Cisco Adaptive Security Appliance (ASA) host.

Figure 12 presents the internal Deepfield security analyst’s view of this rule. In the broader context of the Defender decision tree, flows identified as **‘quic’** are blocked by default because they are directed to a destination Genome knows does not support that protocol.



3300 quic_flood

(df['port.dst'].isin([80,443])) & (df['protocol'] == 17) & (~df['genome.dst'].isin([GENOME_EXCEPTION_QUIC]))
> Unique Src-Per-Dst-24 25
> PPS-Per-Dst-24 30,000
+ Threat Level 4
= quic

Figure 12: Defender decision tree model for QUIC flooding attacks

Evaluating the performance and efficiency of DDoS detection and mitigation

Evaluating the performance and effectiveness of DDoS solutions can be challenging for many network operators. All DDoS vendors claim unique ML algorithms, novel advanced countermeasures, specialized hardware and other capabilities that are technically challenging for their customers to evaluate.

In an industry filled with claims and counterclaims, the best way to determine the value of a DDoS mitigation solution is to consider three key metrics:

1. Mitigation performance (false positive/false negative)
2. Scale
3. Cost.

Of course, the central metric for any DDoS solution is its ability to filter DDoS traffic. The key concern, however, is not whether mitigation will drop 100 percent of DDoS traffic. Most network operators can BGP blackhole all traffic destined to a customer under attack to achieve a 100 percent drop without a scrubber.

We believe the important metric is the false-positive rate for DDoS traffic.

In other words, assuming all DDoS traffic is blocked, how much legitimate (non-DDoS) traffic will a scrubber drop during a DDoS event? No DDoS mitigation is perfect, and all mitigations will introduce some (hopefully small) percentage of collateral damage.

We measure the false-negative and false-positive rates using the Deepfield DDoS Library and the Deepfield Peacetime Library. The peacetime dataset, akin to the DDoS library, encompasses a diverse range of customer profiles, geographies, network topologies and traffic distributions. For example, customer profiles include cable operators, fixed and mobile network operators, data center clients, large financial institutions, small and midsize businesses (SMBs), enterprises and retail broadband customers.

Unless specifically indicated otherwise, the Deepfield DDoS solution matches or surpasses the false-positive rates of competing appliance-based scrubbing solutions, as assessed using the combination of DDoS and peacetime samples. Most vendors test their false-positive and false-negative rates internally, but not all share these performance metrics externally.

For some attacks and customer types, Defender's false-positive rates will depend on the complexity of the attack and the router filter capacity. In the sections below, we annotate each attack with a discussion of the countermeasure, its relationship to router filter capacity and corresponding false-positive considerations. We also cover deployment considerations in a dedicated section.

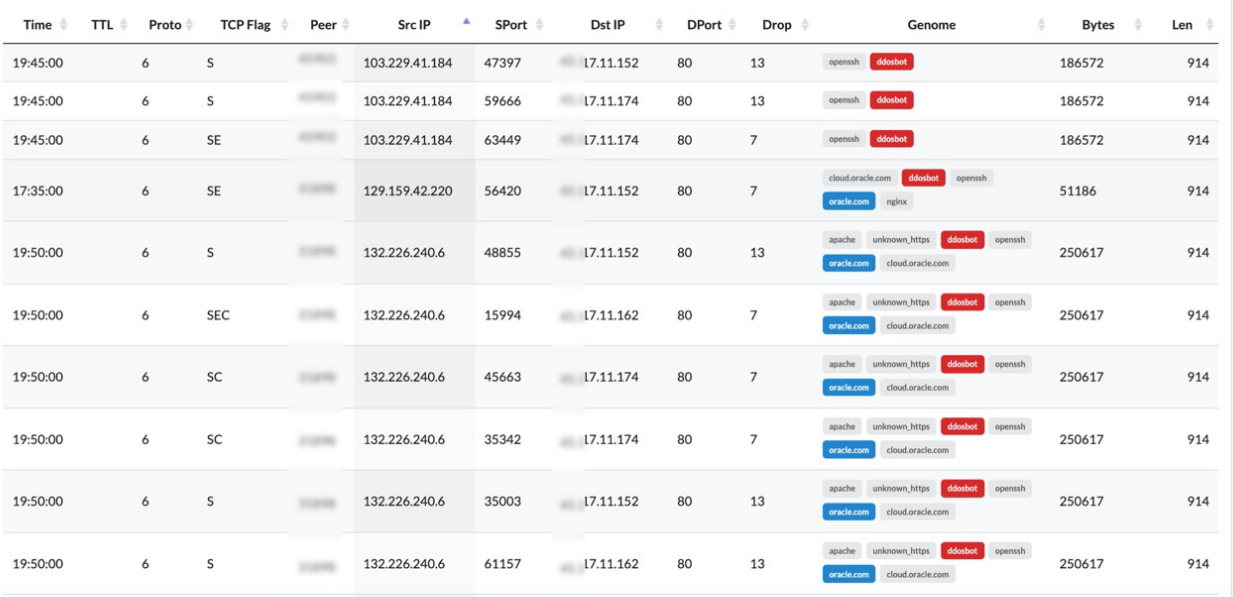
As noted earlier, application-layer DDoS attacks are almost exclusively a botnet problem. Based on real-world data from the Deepfield DDoS Library, Defender can block more than 90 percent of application-layer DDoS attacks, maintaining a false-positive rate below 5 percent and requiring fewer than 2,000 router filter entries. For lower-bandwidth attacks (below 10 Mbps), the false-positive rate may increase to around 10 percent because detection accuracy in these scenarios can depend on having adequate IPFIX sampling rates.

The number of router filter entries required will increase to 2,000 for some particularly sophisticated application vectors, such as botnets exclusively resident within the CSP or cloud provider's own infrastructure. We recommend that end-application customers always deploy a firewall, web application firewall (WAF) or a similar system to provide low-bandwidth application attack protection that complements the upstream volumetric DDoS attack protection provided by Defender.

Table 1: Nokia Deepfield Defender protects against all known botnet and application DDoS attacks

Attack	Average peak	Countermeasure	False negative/false positive
HTTP(S) GET/POST	1 Gbps	ABM	0% / < 5%
Rudy	100 Mbps	ABM	0% / < 10%
Slowloris	50 Mbps	ABM	0% / < 10%
UDP flood	50 Gbps	ABM, ASM, TBM	0% / < 5%
TCP flood	50 Gbps	ABM, ASM, TBM	0% / < 5%
Protocol flood	50 Gbps	ABM, ASM, TBM	0% / < 1%
Gaming	10 Gbps	ABM, ASM, TBM	0% / < 5%
SIP	500 Mbps	ABM, ASM	0% / < 5%
SQL	100 Mbps	ABM, ASM	0% / < 5%
DNS	10 Gbps	ABM, ASM	0% / < 10%
QUIC	10 Gbps	ABM, ASM	0% / < 5%

In addition to application DDoS, booters increasingly use botnets for protocol-based floods (UDP, TCP, GRE, QUIC, Gaming etc.). Figure 13 shows a 2 Mpps TCP Syn/Xmas attack on a large US financial website on 20 December 2021. This TCP flood used around 2,000 bots, including a mix of Mikrotik, Cisco VoIP phones, Hikvision cameras and compromised IBM and Oracle Cloud server accounts.



Time	TTL	Proto	TCP Flag	Peer	Src IP	SPort	Dst IP	DPort	Drop	Genome	Bytes	Len
19:45:00		6	S	[redacted]	103.229.41.184	47397	[redacted]17.11.152	80	13	openssh ddosbot	186572	914
19:45:00		6	S	[redacted]	103.229.41.184	59666	[redacted]17.11.174	80	13	openssh ddosbot	186572	914
19:45:00		6	SE	[redacted]	103.229.41.184	63449	[redacted]17.11.174	80	7	openssh ddosbot	186572	914
17:35:00		6	SE	[redacted]	129.159.42.220	56420	[redacted]17.11.152	80	7	cloud.oracle.com ddosbot openssh oracle.com nginx	51186	914
19:50:00		6	S	[redacted]	132.226.240.6	48855	[redacted]17.11.152	80	13	apache unknown_https ddosbot openssh oracle.com cloud.oracle.com	250617	914
19:50:00		6	SEC	[redacted]	132.226.240.6	15994	[redacted]17.11.162	80	7	apache unknown_https ddosbot openssh oracle.com cloud.oracle.com	250617	914
19:50:00		6	SC	[redacted]	132.226.240.6	45663	[redacted]17.11.174	80	7	apache unknown_https ddosbot openssh oracle.com cloud.oracle.com	250617	914
19:50:00		6	SC	[redacted]	132.226.240.6	35342	[redacted]17.11.174	80	7	apache unknown_https ddosbot openssh oracle.com cloud.oracle.com	250617	914
19:50:00		6	S	[redacted]	132.226.240.6	35003	[redacted]17.11.152	80	13	apache unknown_https ddosbot openssh oracle.com cloud.oracle.com	250617	914
19:50:00		6	S	[redacted]	132.226.240.6	61157	[redacted]17.11.162	80	13	apache unknown_https ddosbot openssh oracle.com cloud.oracle.com	250617	914

Figure 13: Botnet TCP S/Xmas flood to large US financial institution on 20 December 2021

The analysis of the attack PCAPs reveals that most of the botnet application payloads include valid requests (HTTP(S), SIP, DNS, etc.). While some attacks (e.g., on the Valve game server) use specially crafted payloads intended to crash the server, most attacks use valid content and headers to bypass firewall protections. Magic application "byte strings of death" attacks are rare and now largely anachronistic. Most application-layer DDoS attacks instead focus on generating high workloads using valid queries. Typically, the only distinguishing characteristic of application DDoS attacks is that the traffic originates from botnets. Our position is that the efficiency of DDoS mitigation solutions should be tested using real-world application-layer DDoS attack samples.

Deepfield Defender will always choose the optimal strategy for mitigating DDoS, making trade-offs between the desired false-positive rate and the available router filter capacity. In the example in Figure 13, Defender has multiple options for filtering the botnet flood, including type of IoT, previous botnet membership or, for the simplest filter strategy, the booter has inexplicably generated invalid TCP SYN payload lengths (914). Except for rare, special-case applications that involve TCP option headers, internet TCP SYN will range from 40–60 bytes depending on the OS and padding considerations.

Deployment considerations

For the first twenty years of the internet's existence, network operators had limited DDoS mitigation options — namely, to purchase and deploy large numbers of hardware scrubbers, usually from a single dominant vendor.

Recent advances in networking hardware and ML have provided CSPs and data center operators with multiple options for DDoS mitigation, including line-speed programmable router filters, a dozen or more third-party scrubber vendors, and a new generation of edge-located application-level next-generation firewalls.

Almost all vendor solutions will mitigate DDoS attacks. The primary mitigation design trade-off usually involves the scale, cost and granularity of mitigations.

While scrubbers continue to effectively mitigate many types of DDoS attacks (particularly amplification and aberrant spoofed packet floods), scrubber appliance-based DDoS protection comes at a significant cost. Maintaining scrubbing capacity at the same rate as the total network bandwidth growth is prohibitively expensive for most operators.

The alternative approach is to leverage existing programmable routers. Since network operators already deploy and dimension routers to handle bandwidth growth, the cost relationship between traffic growth and DDoS mitigation is decoupled, making routers a dramatically more economical solution.

Here are key design considerations for using Deepfield Defender to upgrade, replace or augment existing scrubbers for DDoS mitigation:

- **Number of filter entries:** Most amplification and reflection attacks require fewer than 20 filter entries per mitigation. The worst-case botnet or IPHM flood attacks may require up to 2,000 entries per mitigation to maintain a false-positive rate below 5 percent. We recommend provisioning up to 10,000 filter entries per Juniper MX and 200,000 filter entries on Nokia Service Routers based on FP4/FP5 processors. The number of filter entries available for mitigation directly impacts the number of simultaneously supported mitigations and expected false-positive rates.

- **Programming API:** FlowSpec is a less expressive API than NETCONF. Specifically, FlowSpec lacks the ability to configure parameters such as TTL, CIDR grouping and access to other ACL constructs otherwise supported on Nokia SR and Juniper MX routers. As with any computer programming language, less expressivity means more code, or, in the case of DDoS attacks, more ACLs. NETCONF provides an advantage over FlowSpec in that a larger number of attacks (especially some botnet and spoofed flood attacks) can be blocked with fewer than 100 filter entries. (FlowSpec may require up to 2,000 for these same attacks.) Nokia is currently advancing a draft for FlowSpec v2 in the IETF that addresses many of these FlowSpec v1 shortcomings.
- **Router vendor and OS version:** Deepfield can use Nokia SR and Juniper MX routers to block all DDoS traffic (or a Nokia 7750 DMS-1 if a dedicated appliance is preferred). Nokia SR routers based on FP4 and FP5 provide additional scale and advanced filter mechanisms, including line-speed payload filters, advanced filter commands and real-time packet samples. The additional SR filter capacity and specialized filters reduce Defender false-positive rates below 5 percent for some more sophisticated application botnet and IPHM attacks.
- **Commercial model:** Some CSPs do not provide any DDoS protection to customers. They readily BGP blackhole enterprises or subscribers when attacks jeopardize backbone or router links. Other CSPs provide DDoS protection to a select few high-end enterprises (usually financial institutions) or managed security service providers (MSSPs) that can support its cost. Defender provides an alternative DDoS defense approach with the precision (low false-positive rate) and scale that CSPs need to protect all customers.
- **Threat model:** Scrubbers successfully mitigate reflection and amplification attacks, the most common forms of DDoS attacks today. However, our research and similar work by academic and industry groups show that botnets are rapidly overtaking amplification as the dominant form of volumetric DDoS attacks. The challenge is that botnets can generate valid traffic, including legitimate, statistically representative payloads and IP headers. Many botnet payloads (as examined from the Deepfield DDoS Library attack samples) will pass CAPTCHA challenges as well as most scrubber authentication protections and ML filters. CSPs and data center operators need to evaluate DDoS mitigation architectures for current and emerging threat models.

Based on these considerations, we recommend that customers pursue a Defender-only DDoS mitigation architecture with Nokia 7750 SR, Nokia 7750 DMS-1 or Juniper MX routers or use Defender with network-based volumetric protection in conjunction with edge firewalls, WAFs or scrubbers for additional, granular application-layer DDoS protection.

Conclusion

Nokia Deepfield Defender draws on more than 25 years of research, development and experience building internet- and cloud-scale DDoS solutions. It addresses the need for a fundamental re-evaluation of DDoS protection strategies to counter new threats emerging in the era of the cloud and IoT [4] [5].

Early Deepfield data highlighted the potential for exponential growth in DDoS attack frequency and volumes. It also correctly predicted the rapid evolution of DDoS from localized IP spoofed traffic (using IP header modification, IPHM) to large-scale botnets—a transition that has made many existing countermeasures and technologies obsolete.

We started Deepfield when cloud and webscale demands began driving dramatic innovations in the router hardware market. These included next-generation line cards based on more capable merchant silicon (e.g., FP4/FP5) and powerful new programming interfaces (e.g., FlowSpec, NETCONF, gRPC). These and other innovations have brought major improvements in router filter scalability and flexibility over the past five years.

Ultimately, we created Defender because programmable routers afforded one of the only DDoS mitigation technologies capable of scaling at the same rate as internet traffic and DDoS attacks. Integrating Defender with programmable routers and our ML models and real-time inference capabilities has enabled us to present a high-precision, line-speed alternative to traditional hardware scrubbers. Defender dramatically reduces the number of scrubber appliances required for carrier and large data center networks. The result is superior DDoS protection at a fraction of the hardware cost and footprint.

When you adopt programmable router-based DDoS mitigation, your main consideration is the trade-off between filter capacity and the acceptable rate of false positives. Defender's ML-driven models are trained to identify and block 100 percent of DDoS attacks. But sophisticated IPHM or botnet-based application-layer attacks may require complex filter configurations in excess of 2,000 entries, or the combined use of Defender with Nokia 7750 DMS-1 or third-party systems such as scrubbers, firewalls or WAFs.

Despite the frightening myths surrounding DDoS, most attacks involve one of three basic delivery vectors: amplification or reflection, IPHM flood or botnet application. Extensive testing with the 10,000-plus real-world attacks in the Deepfield DDoS Library shows that Defender provides significant protection for all three of these DDoS attack vectors.

CSPs and data center operators should evaluate DDoS mitigation solutions on scale, cost and efficacy. They should also consider DDoS mitigation false-positive tolerances against the cost and complexity of different solutions.

We are confident that Deepfield Defender will become an essential element in many networks of service providers and network operators that are seeking a DDoS security solution that is not only accurate and scalable but also cost-effective.

Abbreviations

ACL	access control list
AI	artificial intelligence
API	application programming interface
BGP	Border Gateway Protocol
bps	bits per second
CDN	content delivery network
CIDR	classless interdomain routing
CPE	customer premises equipment
CSP	communications service provider
CVE	Common Vulnerabilities and Exposures
DDoS	distributed denial of service
DML	Deepfield Model Language
DMS	Defender Mitigation System
DNS	Domain Name System
ERT	Emergency Response Team
GDTA	Global DDoS Threat Alliance
GPT	generative pre-trained transformers
GRE	Generic Routing Encapsulation
gRPC	gRPC Remote Procedure Calls
HTTPS	Hypertext Transfer Protocol Secure
IETF	Internet Engineering Task Force
IP	Internet Protocol
IPFIX	IP Flow Information Export
IPHM	IP header modification
LLM	large language model
ML	machine learning
MSSP	managed security service provider
NETCONF	Network Configuration Protocol
OS	operating system
PCAP	packet capture
PON	Passive Optical Network
pps	packets per second

QUIC	quick UDP internet connection
SIP	Session Initiation Protocol
SMB	small and midsize businesses
SYN	synchronize
TCP	Transmission Control Protocol
TTL	time-to-live
UDP	User Datagram Protocol
VoIP	voice over IP
WAF	web application firewall

References

[1] P. Paganini, "FBI seized 48 domains linked to DDoS-for-hire service platforms," Security Affairs, 15 December 2022. [Online]. Available: https://securityaffairs.com/139670/cyber-crime/fbi-ddos-for-hire-service-platforms.html.

[2] R. McMillan, "Your Website Comes From 1,000 Places. Here's How to Map Them," Wired Magazine, 31 July 2012. [Online]. Available: https://www.wired.com/2012/07/deepfield/.

[3] C. Labovitz, "System and method for management of cloud-based systems". United States of America Patent 10374961, 2017.

[4] C. Labovitz, "Tracing DDoS End-to-End 2021," NANOG, 2021. [Online]. Available: https://www.nanog.org/news-stories/nanog-tv/nanog-82-webcast/tracing-ddos-end-to-end-in-2021/.

[5] D. Menscher, "Exponential growth in DDoS attack volume," Google, 16 October 2020. [Online]. Available: https://cloud.google.com/blog/products/identity-security/identifying-and-protecting-against-the-largest-ddos-attacks.

[6] C. Labovitz and R. Malan, The First Twenty Years of DDoS Countermeasures, Internal Nokia Presentation (available upon request), 2021.

[7] CloudFlare, "Famous DDoS attacks | The largest DDoS attacks of all time," [Online]. Available: https://www.cloudflare.com/learning/ddos/famous-ddos-attacks.

[8] M. Nawrocki, "The far side of DNS amplification: tracing the DDoS attack ecosystem from the internet core," in IMC '21, 2021.

[9] S.-J. Moon et al., "Accurately Measuring Global Risk of Amplification Attacks using AmpMap," USENIX, 2021. [Online]. Available: https://www.usenix.org/conference/usenixsecurity21/presentation/moon.

[10] C. Labovitz and J. Nazario, "Bots, DDoS and Ground Truth: One Year and 5,000 Operator Classified Attacks," NANOG, 2018. [Online]. Available: https://archive.nanog.org/meetings/nanog50/presentations/Tuesday/NANOG50.Talk58.groundtruth.pdf.

[11] D. Gassen, D. McPherson and C. Labovitz, "BGP Flow Specification Deployment Experience," [Online]. Available: https://archive.nanog.org/meetings/nanog38/presentations/labovitz-bgp-flowspec.pdf.

About Nokia

At Nokia, we create technology that helps the world act together.

As a B2B technology innovation leader, we are pioneering networks that sense, think and act by leveraging our work across mobile, fixed and cloud networks. In addition, we create value with intellectual property and long-term research, led by the award-winning Nokia Bell Labs.

Service providers, enterprises and partners worldwide trust Nokia to deliver secure, reliable and sustainable networks today – and work with us to create the digital services and applications of the future.

Nokia OYJ
Karakaari 7
02610 Espoo
Finland
Tel. +358 (0) 10 44 88 000

Document code: CID213968 (March)