IMPERVA®

WHITE PAPER

Attackers Use DDoS Pulses to Pin Down Multiple Targets, Send Shock Waves Through Hybrids

Introduction

A typical DDoS attack pattern can be characterized as a prolonged wave with a gradual ramp-up that leads to a peak and is followed by either a slow or sudden descent. When repeated, the pattern often resembles a triangle or a sawtooth waveform (Fig. 1).

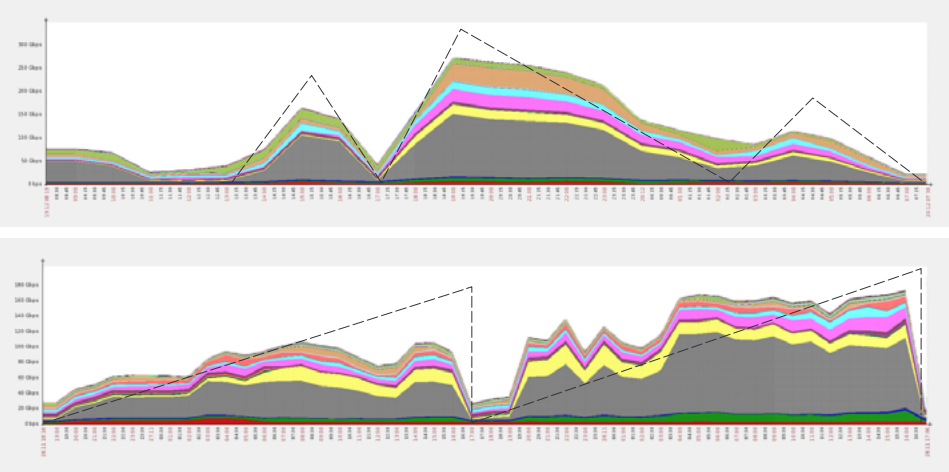


Fig. 1: Typical DDoS traffic patterns resembling a triangle or sawtooth waveform

Over the last few months, however, Imperva Incapsula has witnessed the emergence of a new assault pattern, which we refer to as a "pulse wave" DDoS attack.

Comprised of a series of short-lived pulses occurring in clockwork-like succession (Fig. 2), pulse wave assaults accounted for some of the most ferocious DDoS attacks we mitigated in the second quarter of 2017. In the most extreme cases, they lasted for days at a time and scaled as high as 350 Gbps.

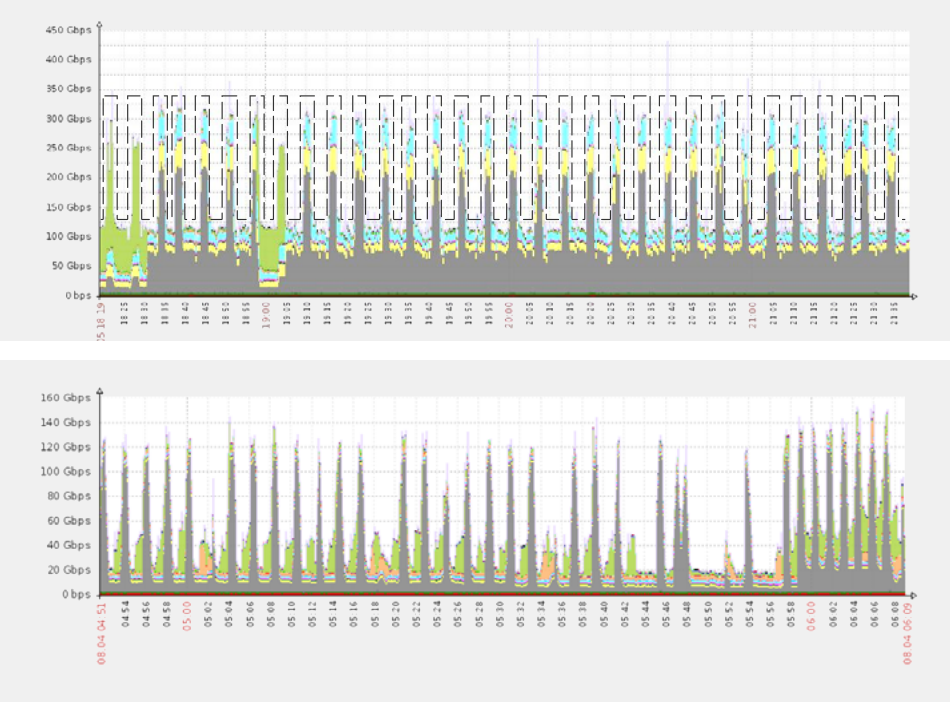


Fig. 2: Pulse wave DDoS attack examples Incapsula mitigated in Q2 2017

Following a review of these assaults, we believe that pulse wave DDoS represents a new attack methodology—one that is purposefully designed to increase the offenders' attack output and take advantage of soft spots in hybrid "appliance first, cloud second" mitigation solutions.

With this white paper, our goal is to bring attention to the existence of pulse wave DDoS attacks, the reasons for their existence and describe the threat they pose. It also takes a step back to discuss a general shift in the threat landscape—from continuous DDoS barrages to repeat short burst attacks—which serves as a backdrop for the emergence of pulse wave assaults.

Defining Pulse Wave Attacks

The modus operandi of a pulse wave attack starts with its definition, having a focus on the following criteria:

- **Immediacy** – The most distinguishable aspect of pulse wave assaults is the absence of a ramp-up period—all attack resources are committed at once, resulting in an event that, within the first few seconds, reaches a peak capacity that is maintained over its duration.
- **Frequency** – Assaults follow a highly repetitive pattern, consisting of one or more pulses every 10 minutes.
- **Persistence** – Assaults last at least one hour, but usually for several hours or even days at a time.
- **Size** – A single pulse is large enough to completely congest a network pipe, (i.e., 10 Gbps or more).

We estimate that pulse wave DDoS events most likely result from skilled bad actors portioning their attack resources to launch multiple assaults at the same time.

If this is so, than the interval between each pulse is being used to mount a secondary assault on a different target. With effective DDoSing it's likely even more simultaneous attacks can be launched—further boosting resource utilization and the offenders' bottom line.

Another advantage pulse wave attacks offer is an ability to disrupt appliance-first hybrid mitigation solutions. They prey on the devices' inability to deal with sudden attack traffic spikes as they attempt to divert it from the appliance to the cloud.

Targeting the Appliance-First Hybrid Bottleneck

To address the rapid escalation in DDoS attack sizes, mitigation hardware vendors spent the last few years looking for ways to add scalability to their on-premises appliances—in a way that wouldn't impact their core business.

In most cases the solution of choice to infuse on-premises appliances with off-premises cloud-based scrubbing, resulting in an appliance-first hybrid mitigation solution.

As the name implies, in such setups, the appliance still serves as a first line of defense. But when faced with an attack exceeding its capacity limits, it activates the cloud and failovers all traffic to it for the duration of the assault.

This model made business sense because it didn't undermine vendors' ability to continue selling mitigation appliances. Instead, introducing a secondary cloud service created an upsale proposition for both legacy customers and new buyers alike.

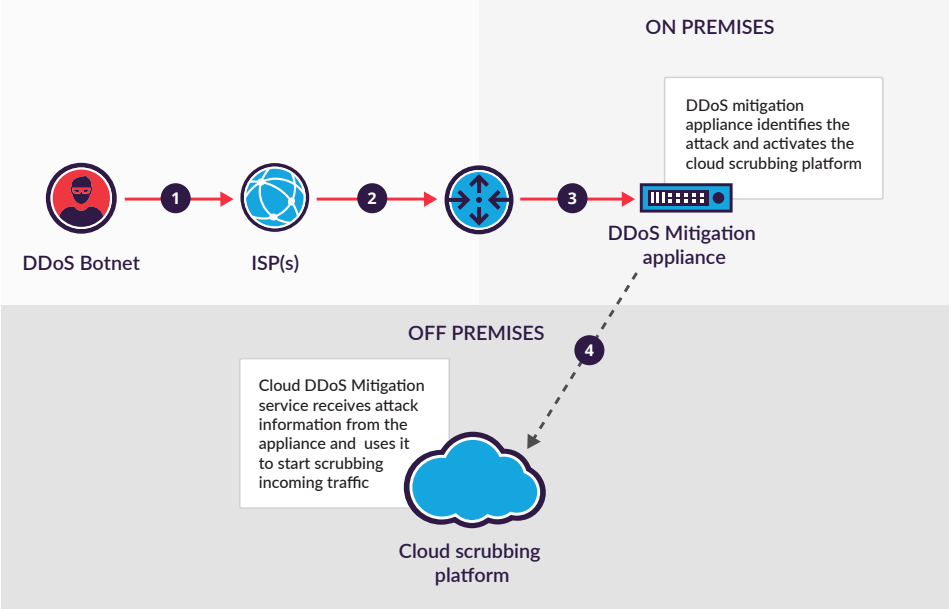


Fig. 3: Appliance-first hybrid DDoS mitigation topology

From a technical point of view, this topology is a suitable solution for scalability issues. However, its inherent weakness is a lack of immediate scalability, since the cloud activation and traffic failover take several minutes to complete.

Moreover, the appliance and cloud have to continually communicate with one another for the failover to properly take place,. While this isn't a problem during a typical DDoS assault (which ramps up slowly), the first burst of a pulse wave immediately congests the network pipe. This effectively cuts off all synching between on-prem and off-prem components.

This communication breakdown can manifest itself in several ways, depending on the specifics of the mitigation configuration. But it usually boils down to the following scenarios.

Flawed Failover

As we've seen, a lack of communication prevents the appliance from reaching out to the cloud and initiating the traffic diversion process. In extreme cases, this prevents the cloud pathway from being activated.

Even when the cloud is configured to activate automatically if the appliance becomes unreachable, it still takes time to verify its service unavailability. This results in the loss of precious seconds (or even minutes) during the most critical assault stage.

Moreover, the pulse wave attack pattern forces the hybrid solution to constantly alternate between its routing settings. By the time the cloud finally does kick in, the short-lived attack pulse is often already over.

This results in deactivation of the cloud pathway and the rerouting of traffic back to the appliance. However, the next pulse wave hits just as this process begins and the appliance-to-cloud communication once again goes down. This forces reactivation of the cloud pathway, only for the appliance to ascertain it's too late and is no longer needed.

The more persistent pulse wave DDoS attacks that we've seen can maintain this cycle for days at a time, which would potentially lock the target network into a state of partial unavailability.

Once that point is reached, the best an operator can hope for is to activate the cloud in an always-on mode, something for which appliance-first hybrids were never designed. The result is a harsh trade-off that exchanges unavailability for perpetual performance degradation.

Loss of Early Attack Information

As part of the failover process, in a regular attack scenario an appliance communicates all of its collected information to the cloud. This includes the traffic's exact signature, based on samples collected during the pre-failover phase of an attack. This information exchange crucial Because It enables more effective mitigation, permitting the cloud to rapidly generate a filtering policy.

But with pulse wave DDoS attacks, communication lines are cut early on and the exchange never takes place. All of the data from the first minutes of an assault are effectively lost; the cloud is forced to reconstruct the attack signature from scratch.

To make matters worse, we have seen perpetrators alternate attack vectors between pulses. Every change requires a new sampling period between the appliance and the cloud.

This deteriorates responsiveness of the mitigation solution. When combined with the previous flawed failover scenario, the damage potential of pulse wave attack is further amplified.

Depleting the Diversion Stock

It's very common for appliance-first hybrid services to limit the number of times their users can divert traffic to the cloud. This results in users having a predefined "stock" of diversions available for use over a given duration.

The frequency of pulse wave assaults has the side effect of rapidly depleting this stock. This can cause an attack target's expenses to pile up quickly, making mitigation an extremely costly proposition for the target.

Shift in The Landscape: DDoS Attacks are Getting Shorter

While pulse wave attacks constitute a new attack method and have a distinct purpose, they haven't emerged in a vacuum. Instead, they're a product of the times and should be viewed in the context of a broader shift toward shorter-duration DDoS attacks.

Multiple industry reports—including our own quarterly DDoS Threat Landscape report—point to an increased number of short-lived DDoS events over the past year. As a result, the majority of all DDoS attacks today, both at the network and application layers, consistently last less than one hour. Moreover, the percentage of such short burst attacks is growing each quarter.

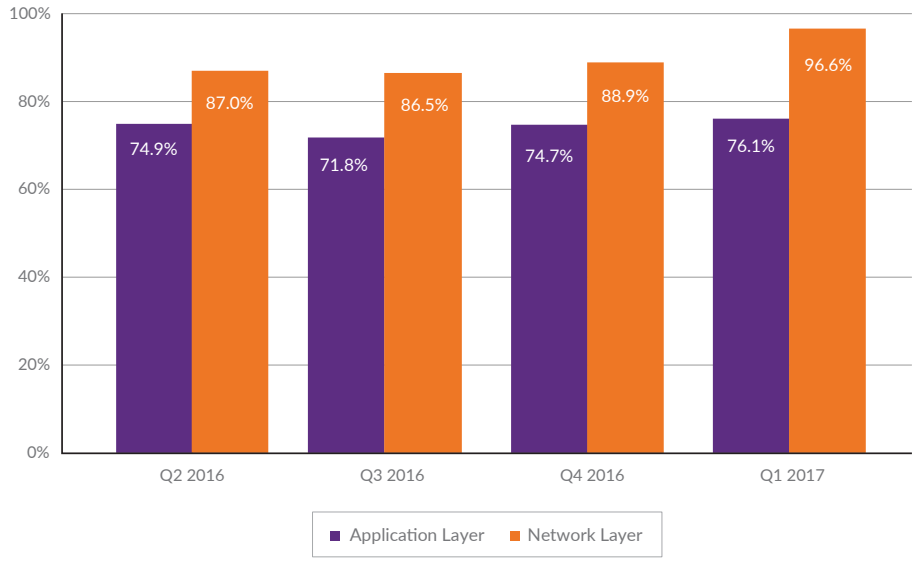


Fig. 4: Distribution of DDoS attacks by duration

The steady decrease in average attack duration reveals this trend impact.

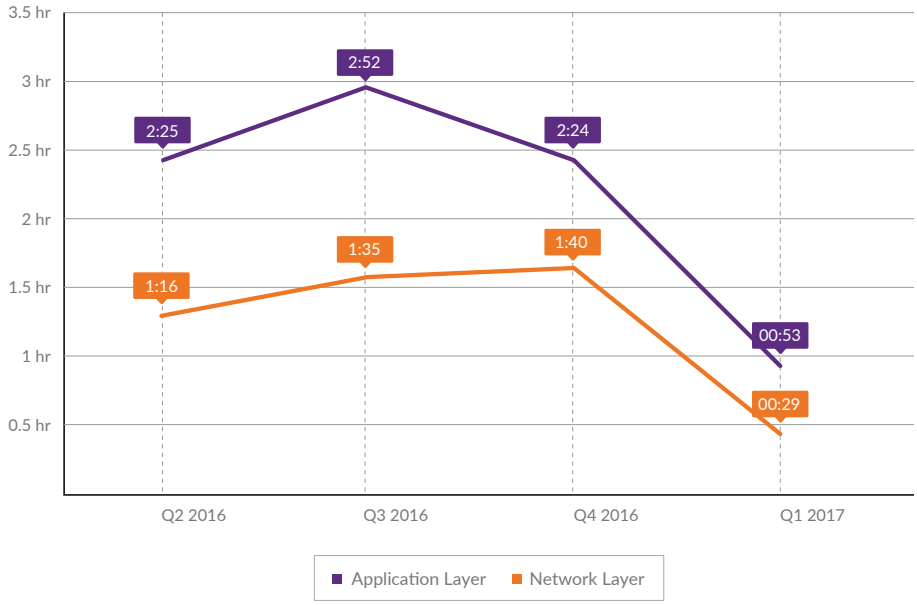
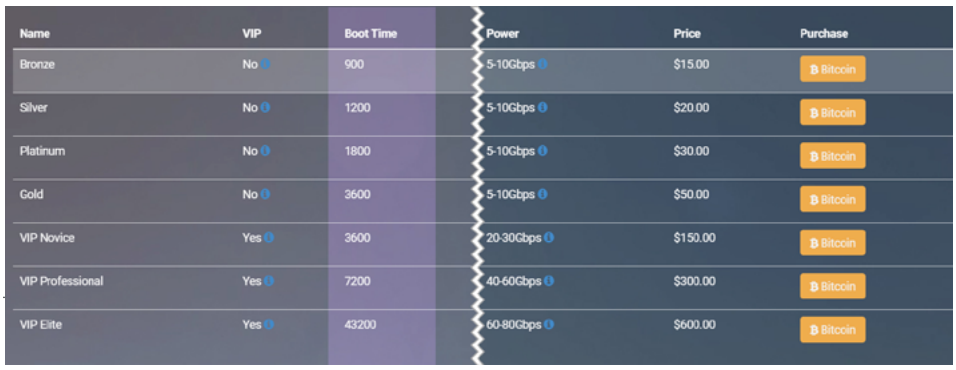


Fig. 5: Average attack duration per quarter

The increase in shorter attacks is attributed to the following three factors:

- **Probing** – A professional actor uses a short assault as a way to test a potential target's defenses and gauge response. Typically, one or two attack bursts is sufficient to gather the necessary intelligence.
- **DDoS-for-hire** – Such attacks are usually waged by non-pros using inexpensive booter or stresser services that are characterized by short boot time and limited capacity.



Name	VIP	Boot Time	Power	Price	Purchase
Bronze	No	900	5-10Gbps	$15.00	Bitcoin
Silver	No	1200	5-10Gbps	$20.00	Bitcoin
Platinum	No	1800	5-10Gbps	$30.00	Bitcoin
Gold	No	3600	5-10Gbps	$50.00	Bitcoin
VIP Novice	Yes	3600	20-30Gbps	$150.00	Bitcoin
VIP Professional	Yes	7200	40-60Gbps	$300.00	Bitcoin
VIP Elite	Yes	43200	60-80Gbps	$600.00	Bitcoin

Fig. 6: Example of boot time (in seconds) offered by a stresser service

- **Hit-and-run attacks** – These are repetitive, low volume, quick-strike assaults that exploit the slower time-to-mitigation of on-demand mitigation solutions. One goal is to exhaust the human resources of a targeted organization.

Pulse wave DDoS assaults are a fourth factor—providing yet another reason why attackers are gravitating toward using short-lived DDoS bursts.

Handiwork of Professional Offenders

Now that the context and reasons for pulse wave assaults has been established, the last point is to identity the offenders. While any examination of possible perpetrators is speculative by definition, the nature of pulse wave attacks offers some clues:

- **Technologically savvy** – The profile of a pulse wave assailant includes a good technical understanding of mitigation solutions coupled with creativity in using specially crafted attacks to exploit appliance weaknesses.
- **Firepower** – A non-amplified, multi-100 Gbps attack requires a well-developed and powerful botnet; this usually indicates the work of a professional. By comparison, the average DDoS-for-hire service typically offers attacks that generate up to 10 Gbps, and never more than 50 – 60 Gbps.
- **Precision** - The clockwork-like repetitiveness of pulse wave attacks—and their ability to reach peak traffic within seconds—highlights the level of control offenders have over their assault resources. Most likely, this points to the use of proprietary botnets created from a relatively small number of high-capacity, connected devices (e.g., servers).

Taken together, these telltales lead us to conclude that this new attacker MO is that of sophisticated bad actors.

Conclusion - Minutes to Go Down, Hours to Recover

Pulse wave attacks target the Achilles' heel of appliance-first mitigation solutions—the non-scalable, hardware piece serving as the hybrids' first layer of defense. By repeatedly striking this weak spot with massive and immediate force, pulse wave attacks send the entire system into disarray, persistently delaying engagement of an effective scrubbing process.

Multiple downtime instances result, each having painful repercussions for the target. Even a brief service interruption requires an hours-long recovery process. For a commercial organization, every such instance also translates into tens of thousands of dollars in direct and indirect damages. For professional offenders—already inclined to split up their attack resources for optimized utilization—this serves as another reason for them to launch pulse wave DDoS assaults.

Consequently, we expect to continue encountering such assaults. We also forecast them to grow larger and become more persistent, fueled by botnet resource evolution and the previously described macro trends we've observed in the DDoS landscape.

To counter these threats, the hybrid mitigation industry should move away from the appliance-first solution. It should instead adopt a new topology that deploys the cloud as the first line of defense. Doing so would eliminate the bottleneck that can be exploited by pulse wave DDoS attacks.

About Imperva Incapsula

Imperva Incapsula is a cloud-based application delivery service that protects websites and increases their performance, improving end user experiences and safeguarding web applications and their data from attack. Incapsula includes a web application firewall to thwart hacking attempts, DDoS mitigation to ensure DDoS attacks don't impact online business assets, a content delivery network to optimize web traffic, and a load balancer to maximize the potential of web environments.

Only Incapsula provides enterprise-grade website security and performance without the need for hardware, software, or specialized expertise. Unlike competitive solutions, Incapsula uses proprietary technologies such as client classification to identify bad bots, and big data analysis of security events to increase accuracy without creating false positives.