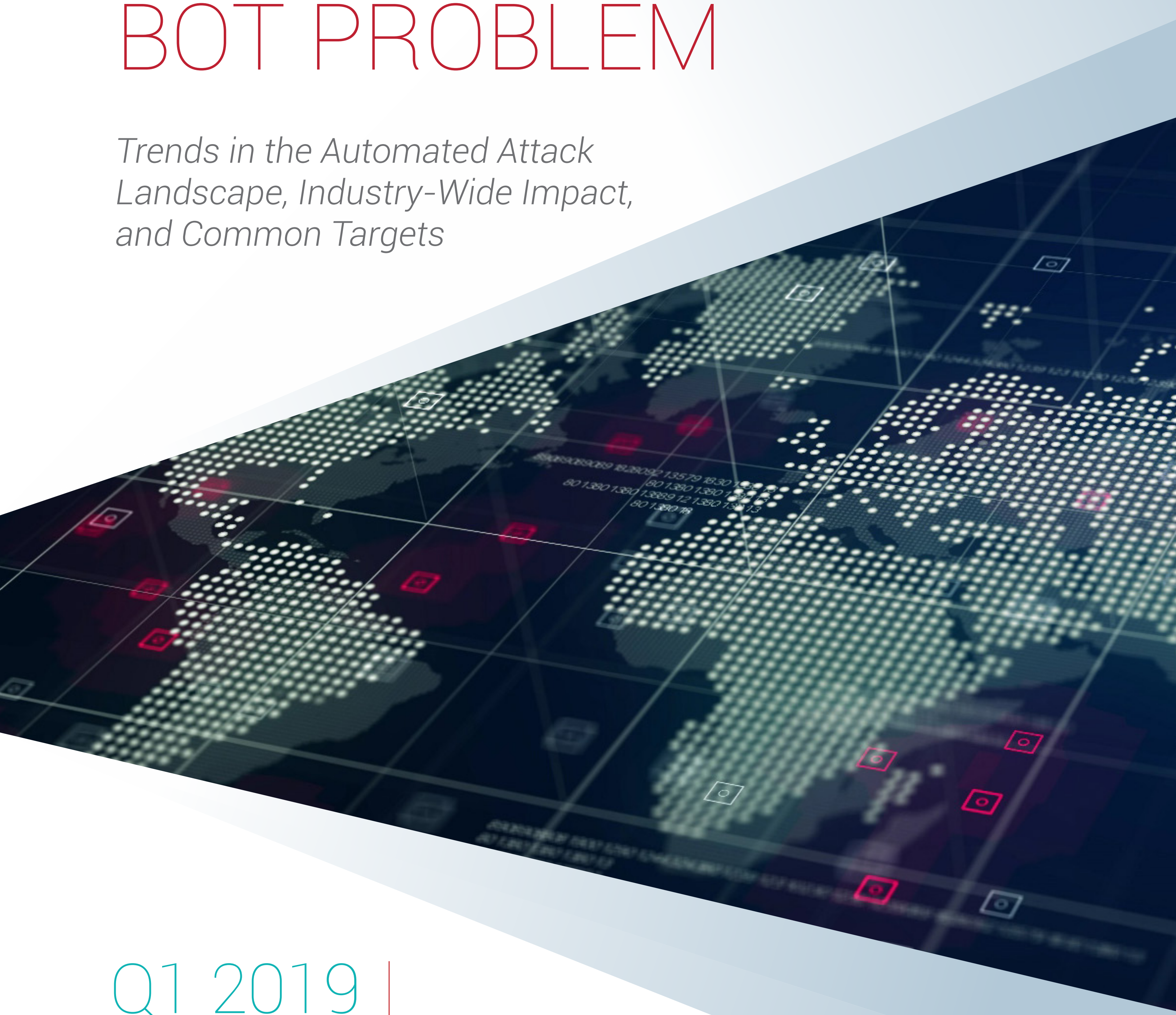
radware

THE BIG, BAD BOT PROBLEM

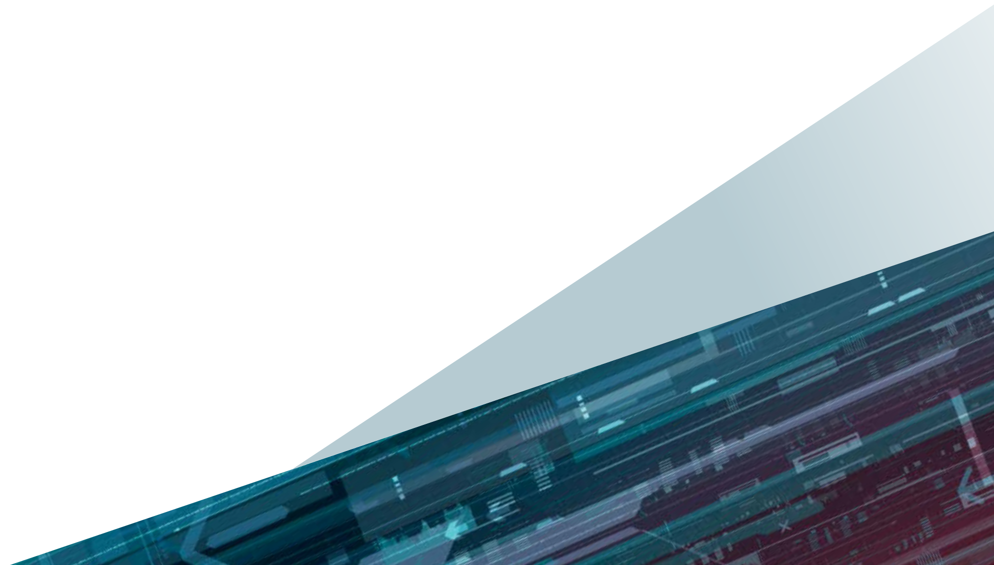
Trends in the Automated Attack Landscape, Industry-Wide Impact, and Common Targets



Q1 2019

MAY 2019

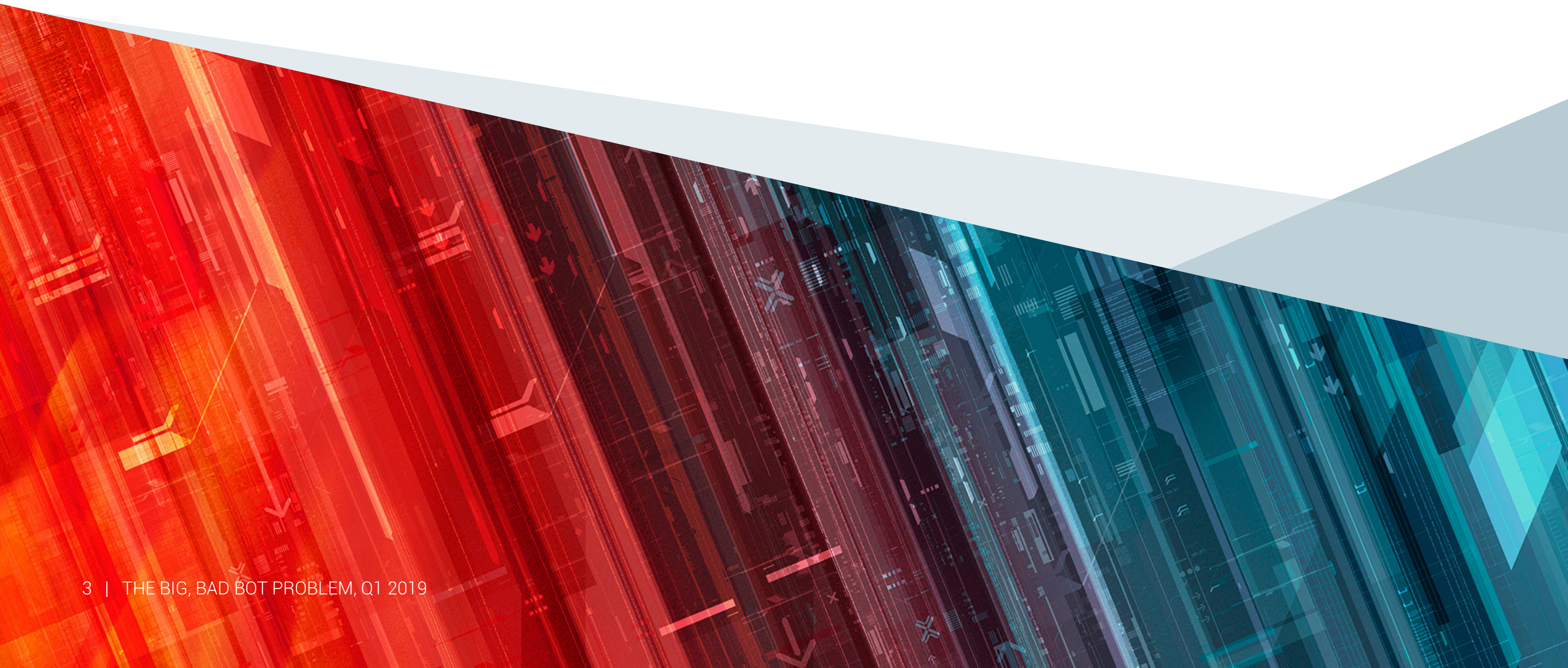
Table of Contents



Why Read This Report

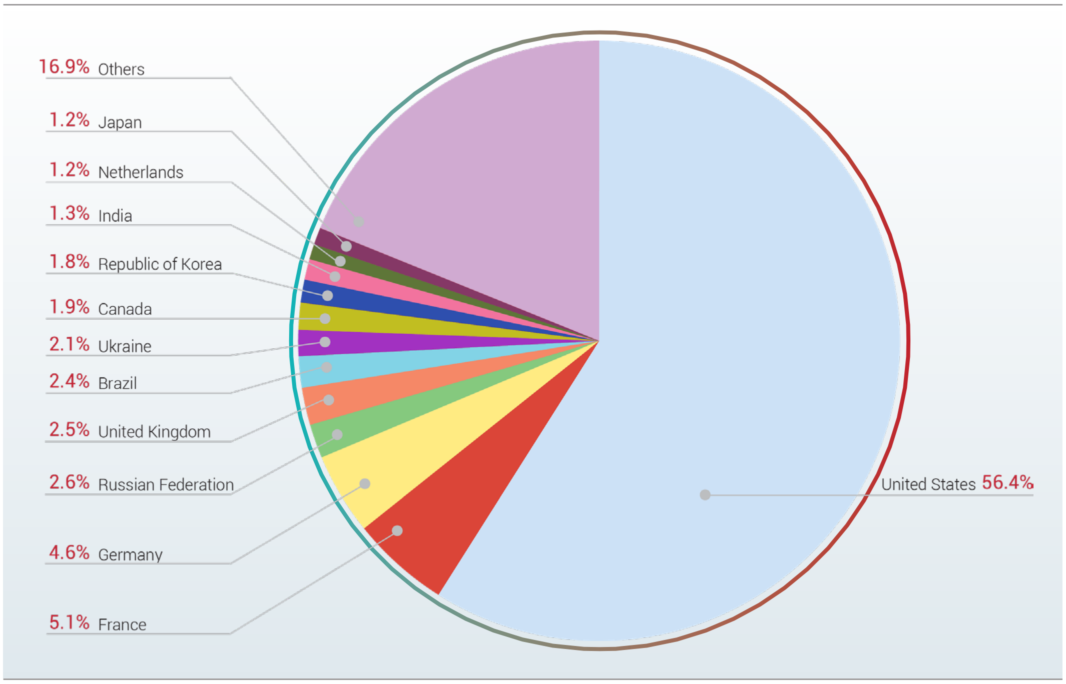
This report summarizes Q1, 2019 trends of automated threats (i.e., bot-generated attacks) as they emerge. It outlines the growing sophistication of bad bot attacks and provides detailed analysis of the impact of bad bots on various industries, including e-commerce, marketplaces and classifieds, media and publishing, and real estate. The report also sheds light on the prevention measures that organizations can implement to avert automated attacks.

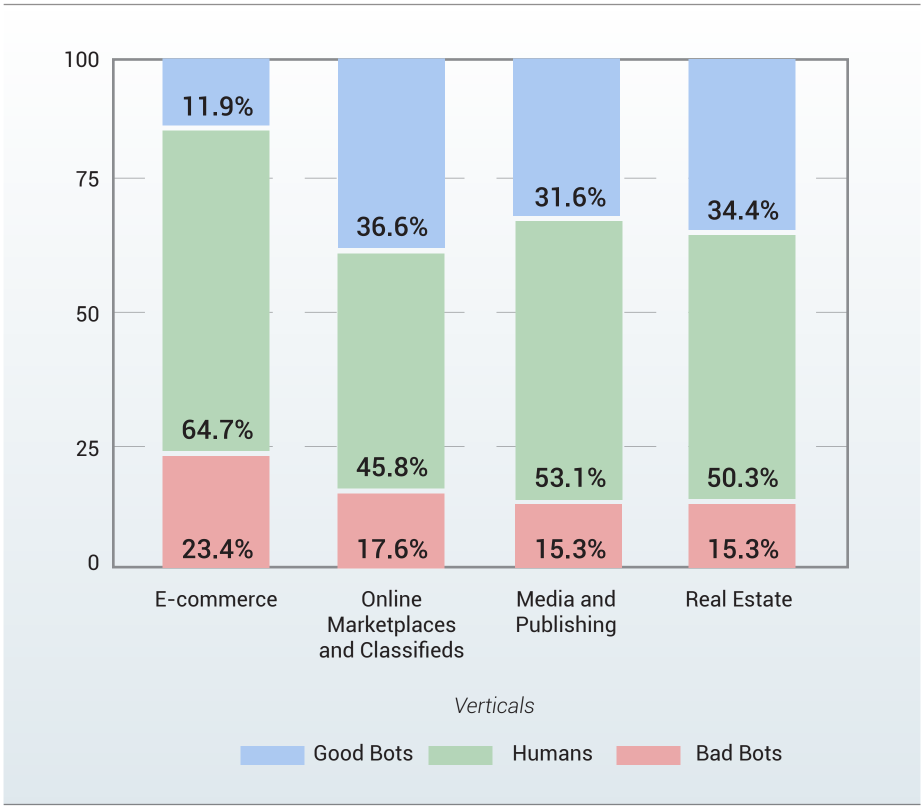
The analysis is based on the data collected from our global client base, encompassing all industries, all geographies, and all company sizes. This is an in-depth analysis of billions of transactions between applications and bots, including behavioral classification of millions of samples of bad bot requests from thousands of anonymized sources across the web.



Key Findings

In Q1 2019, most of bad bots originated from the US





E-commerce, marketplaces/ classifieds and media websites suffer the highest percentage of bad bot traffic.

Bot Trends

Distribution of Internet Traffic

In Q1 2019, 21.5% of the total traffic was bad bots.

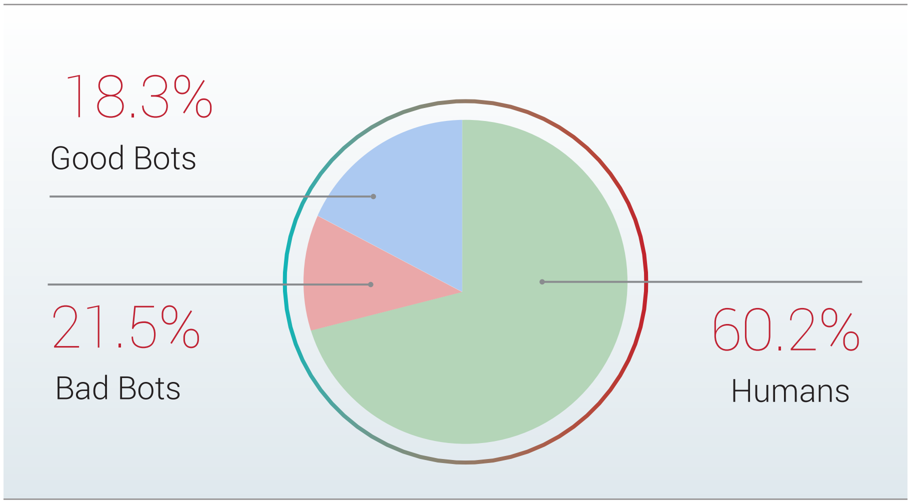


Figure 1: Distribution of Internet Traffic

Origins of Bad Bots

In Q1 2019, most of bad bots originated from the US. With the US being the preferred location for bad bots to identify themselves, the role of bot management solution becomes crucial. Many organizations tend to block countries of origin of bad bots but now with the US coming into the center, such action can lead to massive false positives.

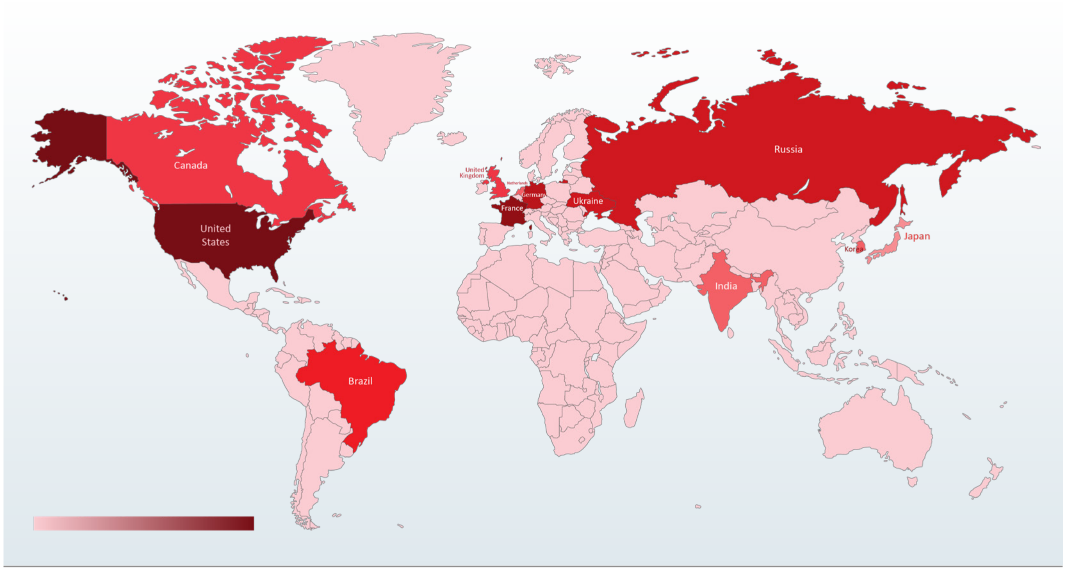


Figure 2: Origin of Bad Bots – World Map

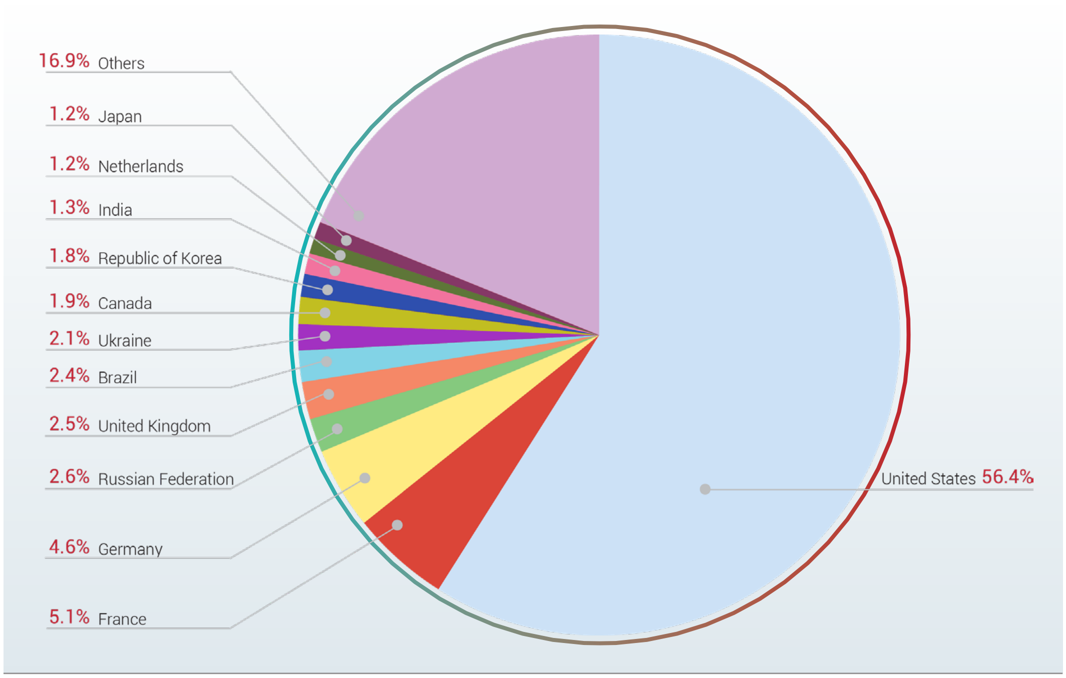


Figure 3: Origin of bad bots – Country-wise

Types of Bad Bots

Bad bots can be classified into four types based on their sophistication level and behavior. Each type is characterized by the level of automation and evasion mechanism. The first and second generations of bots generally use few IP addresses and make thousands of hits from each one. Third and fourth generation bots can rotate through thousands of IP addresses, but make only one or two hits from each address, using an evasion technique known as 'low and slow,' which allows them to sneak past basic security systems.

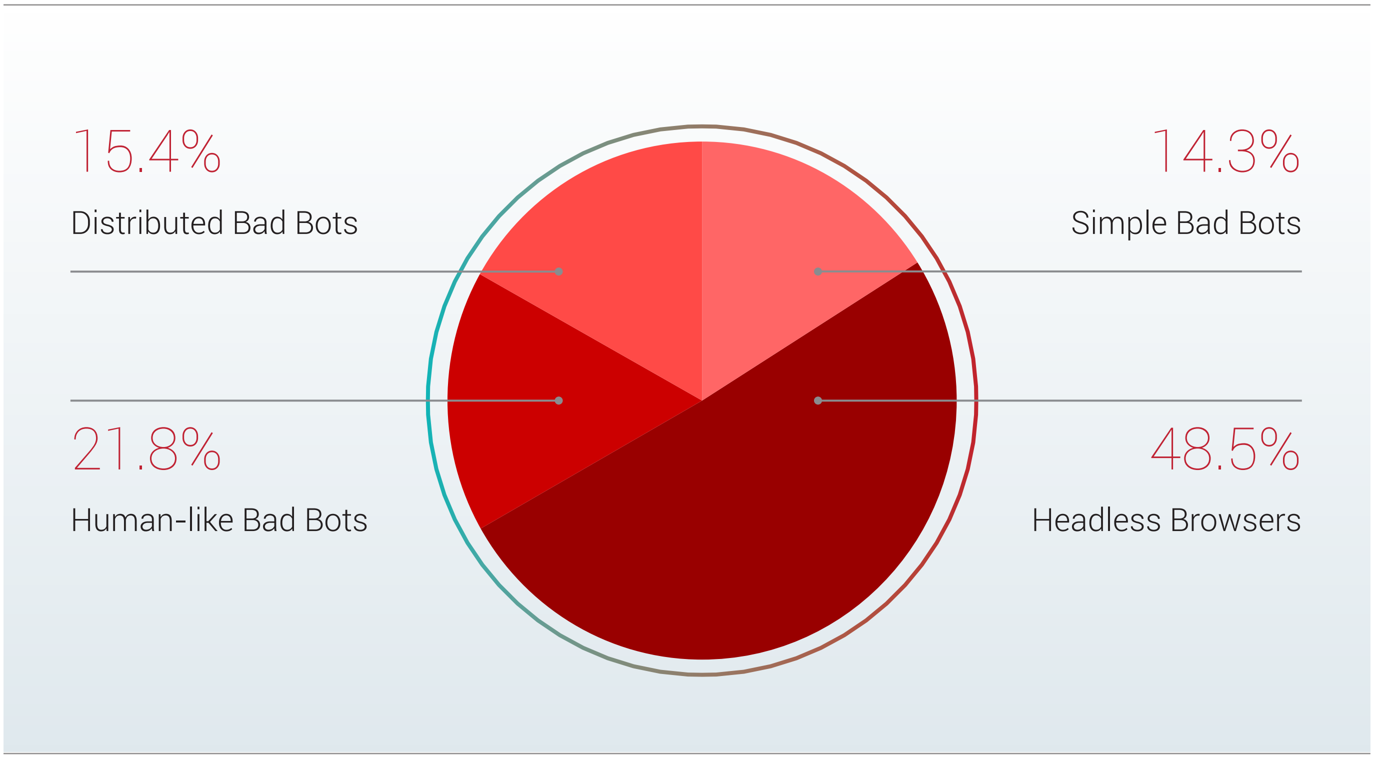


Figure 4: Generation-wise bad bot traffic categorization

SIMPLE BAD BOTS

14.3% of the bad bots were first generation bots in Q1 2019. These bots make cURL-like requests from a small number of IP addresses. Such bots can be identified through a blacklist of User Agent (UA) as well as by analyzing the traffic emanating from different IP addresses. These bots are generally used for basic scraping attacks and spam forms on various forums/ websites.

HEADLESS BROWSERS

Abusing open-source tools such as PhantomJS and others, these can bypass challenges by storing cookies and executing JavaScript. These bots can also dynamically spoof their IPs, making detection harder. It requires fingerprinting the bot's browser and device characteristics — such as the presence of specific JavaScript variables, iFrame tampering, sessions, cookies, etc. Second generation bad bots accounted for 48.5% of traffic in Q1 2019.

HUMAN-LIKE BAD BOTS

Third generation bad bots represented 21.8% of traffic. These are human-like bots that use dedicated or hijacked browsers to perform sophisticated automated attacks such as account takeover, API abuse, carding, ad fraud, and more. They can simulate basic human-like interactions (such as simple mouse movements and keystrokes) and are difficult to detect using traditional security solutions, such as web application firewalls (WAFs).

DISTRIBUTED BAD BOTS

Fourth generation bad bots accounted for 15.4% of traffic in Q1 2019. These bad bots use more advanced human-like interaction characteristics and are massively distributed across tens of thousands of IP addresses to perform large-scale distributed attacks, generating multiple transactions simultaneously. Detecting and mitigating fourth generation bots requires advanced bot management capabilities, such as collective intelligence, behavioral profiling, and contextual correlation.

Industry-Wide Traffic Distribution

Bad bots are present across nearly all industries and verticals. Some industries attract more bad bots than others. E-commerce, online marketplaces/classifieds, media and publishing, and real estate have the highest percentage of bad bot traffic.

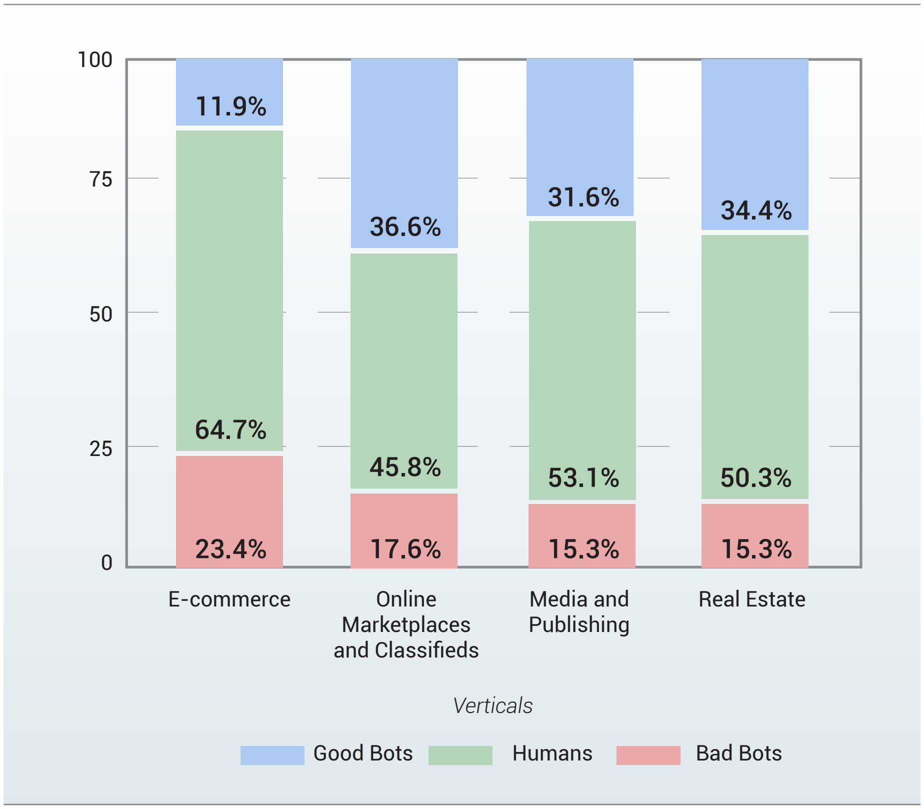


Figure 5: Industry-Wide traffic distribution

Bad Bot Targets by Vertical

Some businesses conduct mischievous activities against their competitors. These businesses deploy bad bots to scrape the content and aggregate data such as product names and pricing for tracking their competitors. However, competitors are not the only threat that online businesses face. Cybercriminals are a grave threat and deploy bad bots to take over user accounts, perform DDoS attacks, and exploit vulnerabilities in applications and APIs. Below is an industry analysis of the overall traffic of four different industries that have the highest percentage of bad bots.

MEDIA AND PUBLISHING

On digital publishing sites, the homepage is the most targeted section, followed by "all articles" and news. Competitors and third-party aggregators scrape unique content of digital publishers to post on their websites. Bad bots are also deployed to perform ad fraud on digital publishing sites and this is one of the reasons behind the presence of a relatively high number of bad bots on these websites.

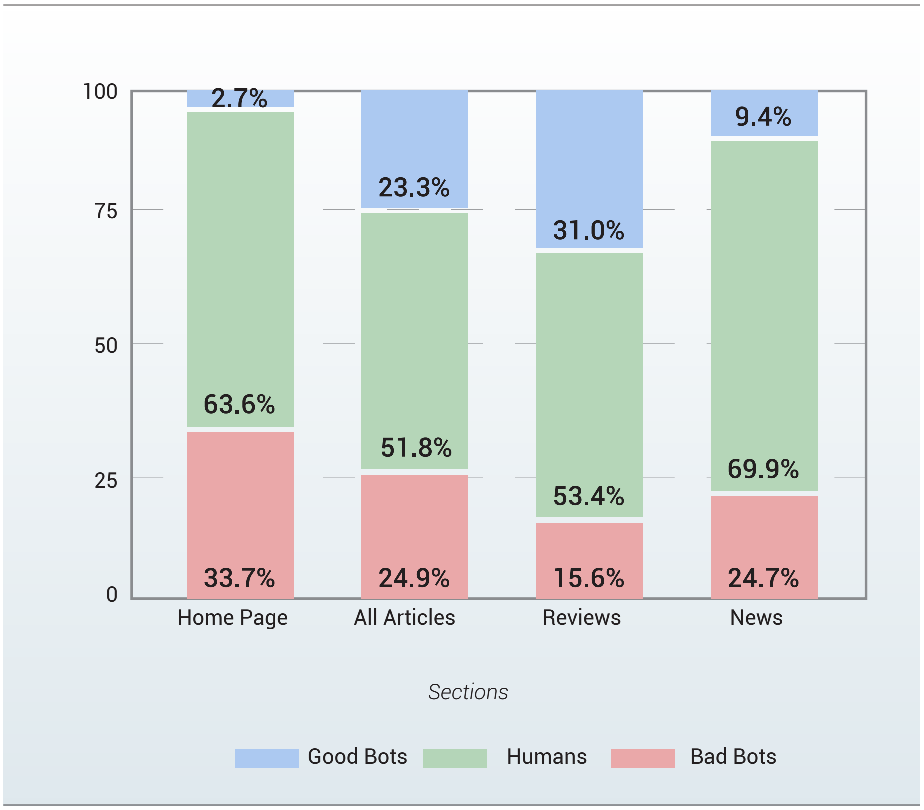


Figure 6: Traffic analysis of media and publishing sites

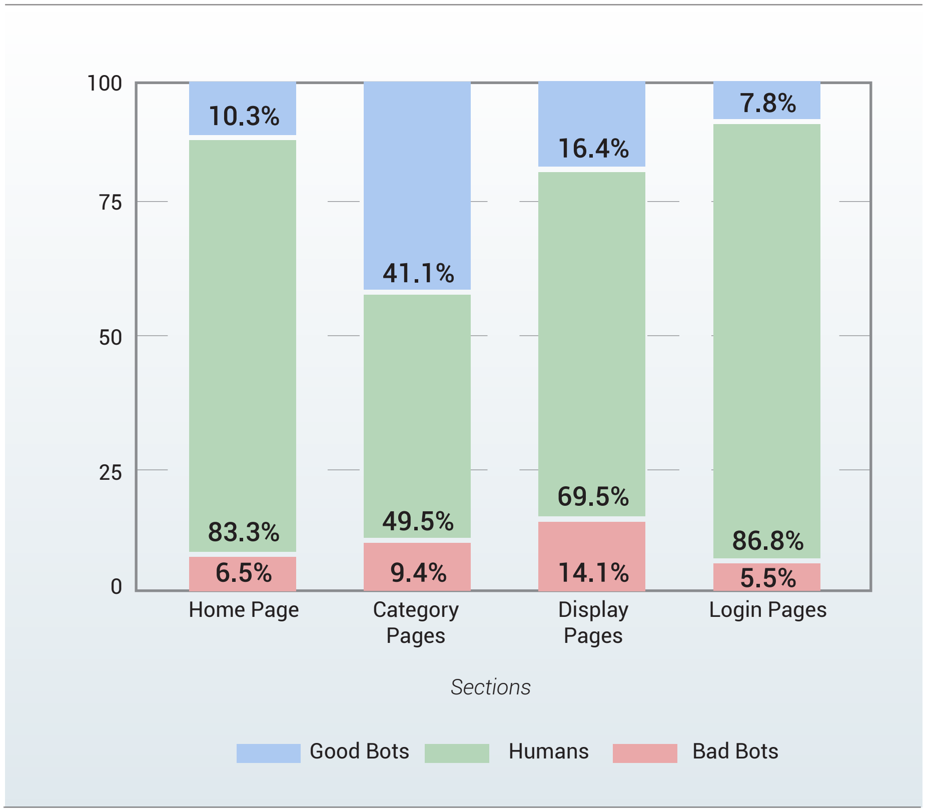


Figure 7: Traffic analysis of real estate portals

REAL ESTATE

In the real estate sector, display pages are the most targeted section, followed by category and login pages. Competitors and aggregators usually target real estate portals to scrape listings.

ONLINE MARKETPLACES AND CLASSIFIEDS

Product pages on marketplaces and classifieds portals are the most targeted section. We also observed a change of priority from other industries in the approach attackers used. 18.7% of total bad bots were directed at login pages to take over user accounts on marketplaces and classifieds portals.

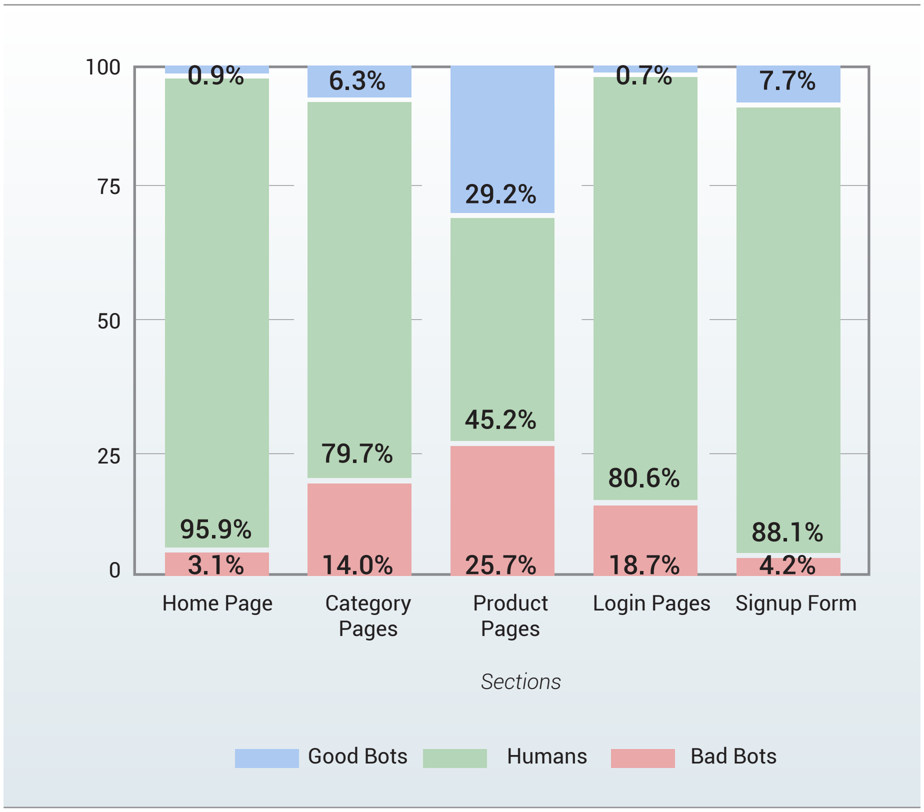


Figure 8: Traffic analysis of marketplaces and classifieds portals

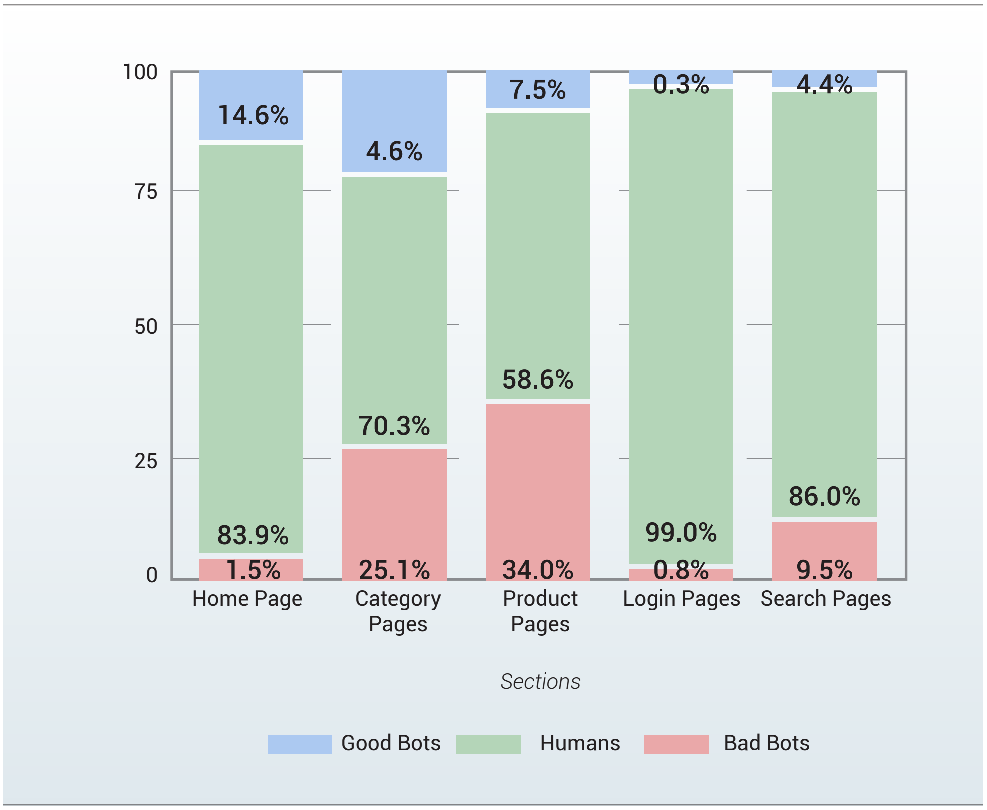


Figure 9: Traffic analysis of e-commerce portals

E-COMMERCE

Category pages on e-commerce firms were the most targeted by bad bots followed by product pages and search pages to scrape content. Less than one percent of the bad bots targeted login pages of the e-commerce firms to takeover user accounts in Q1 2019s.

Recommendations

1. DEPLOY CHALLENGE-RESPONSE AUTHENTICATION

Challenge-response authentication is one basic security protocol that can help you filter bad bots. There are different types of authentication using challenge-response authentication, CAPTCHAs being the most widely used one. Challenge-response authentication can help in filtering outdated user agents/browsers and basic automated scripts but won't assist in blocking sophisticated bots that mimic human behavior. Also, challenge-response authentication requires a risk scoring mechanism, as showing multiple CAPTCHAs to users disrupts the customer experience.

2. BLOCK BAD BOT HARBOR COUNTRIES, DATA CENTERS, AND ISPS

Geo-based blocking of traffic is a common practice, especially to local brands. Some countries are known harbors of bad bots. However, blocking a country will not solve the bad bot problem entirely as bad bots now shift through IPs and use advanced technologies to evade detection.

Similar to countries, data centers and ISPs also safe harbor bad bots. Organizations can block suspected data centers and ISPs. However, blocking all the traffic coming from data centers or ISPs without considering the user behavior can cause false positives. For example, a significant number of users on digital publishing sites come from commercial organizations that use secure web gateways (SWGs) located in data centers to filter user-initiated traffic.1 Blocking data center traffic without considering domain-specific user behavior can cause false positives for digital publishing sites. Below is a list of data centers with bad bot requests coming from them in Q1 2019 which can help your organization filter bots.

[1] *https://www.shieldsquare.com/how-invalid-traffic-misclassification-causes-loss-of-opportunities-for-publishers/*

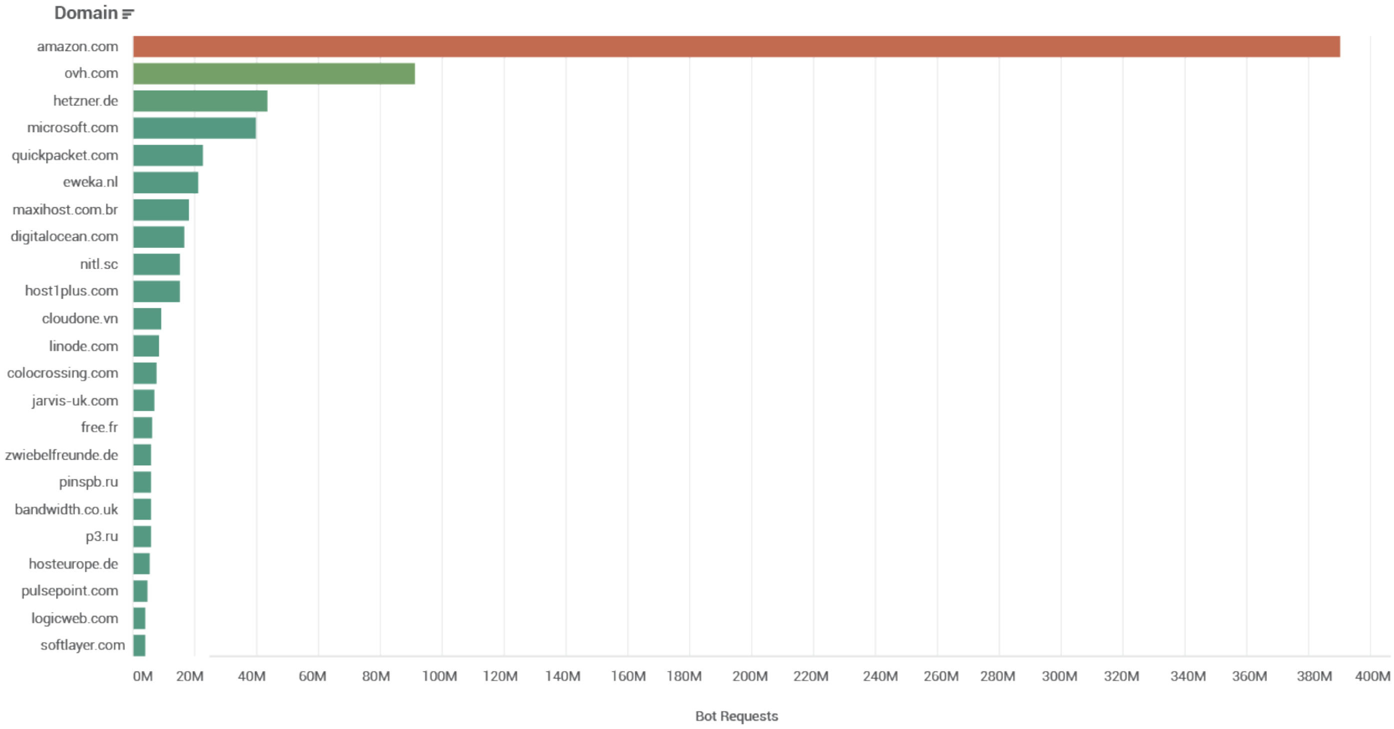


Figure 10: List of data centers that safe harbor bad bots

3. IMPLEMENT STRICT AUTHENTICATION MECHANISM ON APIS

With the widespread adoption of APIs to facilitate interoperability among web applications, automated attacks on poorly protected APIs are mounting. APIs typically only verify the authentication status, but not if the request is coming from a genuine user. Attackers exploit these flaws in various ways (including session hijacking and account aggregation) to imitate genuine API calls. Organizations must examine every API request to ensure that it's coming from a genuine user/device and not from a malicious bot.

4. MONITOR FAILED LOGIN ATTEMPTS AND SUDDEN SPIKES IN TRAFFIC

Cyber attackers deploy bad bots to perform credential stuffing and credential cracking attacks on login pages. Since such approaches involve trying different credentials or a different combination of user IDs and passwords, it increases the number of failed login attempts. The presence of bad bots on your website (to perform scraping, account takeover, or any other type of automated activity) suddenly increases the traffic. Monitoring failed login attempts and a sudden spike in traffic can help organizations take preemptive measures before bad bots cause any damage.

5. DEPLOY A DEDICATED BOT MANAGEMENT SOLUTION

In-house measures, such as the aforementioned practices, provide basic protection but do not ensure the safety of your business-critical content, user accounts, and other sensitive data. Sophisticated third and fourth generation bots can be distributed over thousands of IP addresses and can attack your business in multiple ways. They can execute low and slow attacks every hour or make large-scale distributed attacks which can result in downtime. A dedicated bot management solution facilitates real-time detection and mitigation of such sophisticated, automated activities. It works even better as part of an integrated attack mitigation system that extends the protection of networks and applications from automated attacks.

About Radware

Radware® (NASDAQ: RDWR), a leading provider of cyber security and application delivery solutions, acquired ShieldSquare in March 2019. ShieldSquare is now Radware Bot Manager.

Radware® (NASDAQ: RDWR) is a global leader of cybersecurity and application delivery solutions for physical, cloud and software-defined data centers. Its award-winning solutions portfolio secures the digital experience by providing infrastructure, application and corporate IT protection and availability services to enterprises globally. Radware's solutions empower more than 12,500 enterprise and carrier customers worldwide to adapt quickly to market challenges, maintain business continuity and achieve maximum productivity while keeping costs down. For more information, please visit www.radware.com

Radware encourages you to join our community and follow us on: Radware Blog, LinkedIn, Facebook, Twitter, SlideShare, YouTube, Radware Connect app for iPhone® and our security center DDoSWarriors.com that provides a comprehensive analysis of DDoS attack tools, trends and threats.

www.radware.com | www.radwarebotmanager.com

This document is provided for information purposes only. This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law. Radware specifically disclaims any liability with respect to this document and no contractual obligations are formed either directly or indirectly by this document. The technologies, functionalities, services or processes described herein are subject to change without notice.