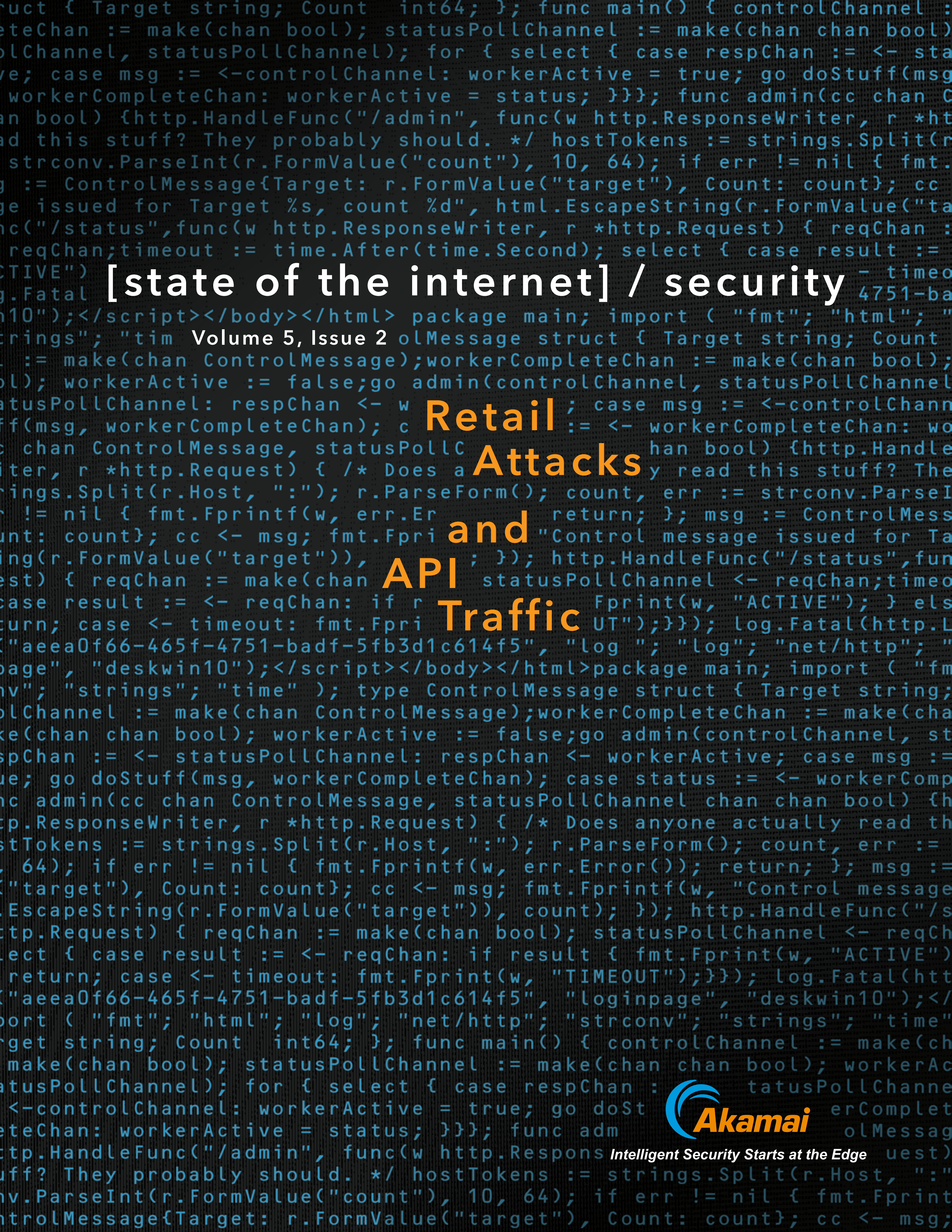


[state of the internet] / security

Volume 5, Issue 2

Retail Attacks and API Traffic

Akamai

Intelligent Security Starts at the Edge

Letter from the Editor

In the past decade, one of the biggest changes in security has been the understanding that we're not simply technologists divorced from the business. Our work is an integral part of every aspect of business. To security practitioners who are relatively new to the industry, this might sound like an obvious statement. However, at some organizations, this type of integrated thought process is still a work in progress. There's a definite continuum between organizations that see the security team as an integrated part of the corporation and those that see it as separate from other business concerns.

An important part of a security team being seen as a legitimate business partner is their ability to identify the risks that the business faces. In the Dark Ages of the 1990s, this resulted in security teams being viewed as the "Department of No." Since then, security leaders have learned to better quantify and communicate risks. The best security teams have clear definitions of risks and risk assessment processes that enable them to explain our perceptions in a way that's more nuanced than the simple binary of *yes or no*. More importantly, they can often attach a theoretical monetary cost to those risks, thereby giving business leaders the ability to better weigh potential costs.

But identifying risks is hard. *Really* hard. Understanding the variations and nuances that might have a significant effect on a business decision is a difficult process, even for the topics we know intimately. We might overestimate the impact, which lessens our standing within the business, or we might underestimate the risk, which leads to finger pointing when things go wrong. Like so many aspects of security, there is not one singular path – it's a balancing act unique to the individuals and the organization.

This issue becomes exponentially harder when we're facing unknowns and issues we have little or no visibility into. All three stories in this issue of the *State of the Internet / Security* report cover aspects of security that we feel numerous organizations are not as cognizant of as they should be. Our survey of API traffic surprised us by revealing that 83% of the hits we see there are API driven. Research into DNS traffic revealed that IPv6 traffic may be underreported, as many systems capable of IPv6 still show a preference for IPv4. Finally, our look at credential abuse and the botnets abusing retailer inventories shows that this is a rising problem that needs attention.

Part of risk assessment is constantly wondering about what problems we *should* be looking at but aren't. There are things that we don't (and can't) know, simply because we lack visibility. Hopefully we can help chip away at this and move one or two more topics into the knowable category.

Table of Contents

TL;DR

- Akamai detected nearly 28 billion credential stuffing attempts between May and December 2018. Tools like the All-in-One botnet are responsible for a large number of the attempts against retail organizations.
- A recent analysis of Akamai's ESSL network revealed an 83% to 17% split between API and HTML traffic on our secure content delivery network. This is a significant increase since the same survey was performed in 2014.
- The reporting of IPv6 usage might be underreported based on Akamai's analysis. This leads to a dangerous assumption that IPv6 isn't worth monitoring.

Overview

All three of our stories in this issue of the *State of the Internet / Security* report are about things most organizations aren't examining. Whether the cause is that organizations don't perceive some issues as important to their environment, if they don't have tooling to monitor these issues, or if the resources to monitor this traffic are not available, this traffic is often being overlooked.

Although organizations examine the traffic generated by botnets, without specialized tools that traffic is often treated the same as any other type of network activity. There are very few places where this is more dangerous than in the retail sector, where botnet creators and retail defenders are playing a multidimensional game, with real money on the line. Our team looked at All-In-One (AIO) bot tools and considered them in the context of the billions of credential abuse attempts we see on a monthly basis.

> "
>
> **Although organizations examine the traffic generated by botnets, without specialized tools that traffic is often treated the same as any other type of network activity.**

Another type of traffic that a lot of organizations have limited visibility into is API traffic. In 2014, Akamai did an internal audit of the JSON and XML traffic on our Enhanced SSL (ESSL) network and found that 47% of the traffic was driven by these two protocols. A similar survey of our traffic in October 2018 showed that 69% of the traffic is now JSON, 14% is XML, and only 17% is HTML.

The Internet has been slowly moving to IPv6, and according to the Internet Society, 28% of the top 1,000 sites are IPv6 capable, while only 17% of the top 1 million sites can say the same. But our research suggests that this might be an underreporting of the numbers, because so many systems show a preference for IPv4, even when they're capable of handling IPv6 traffic. Because IPv6 is still seen as a minority of traffic, it's not a major selling point for a number of security tools. Not all organizations consider the IPv6 space worth monitoring, even when the capability is present.

Akamai Research

TOOLS OF MASS (RETAIL) DESTRUCTION

Between May 1 and December 31, 2018, there were 10,000,585,772 credential stuffing attempts in the retail industry detected on Akamai's network. When that's expanded to all other customer industries, Akamai detected 27,985,920,324 credential abuse attempts over eight months. That works out to more than 115 million attempts to compromise or log in to user accounts every day.

The reason for these attempts isn't complex. The malicious actors responsible for them are looking for data – such as personal information, account balances, and assets – or they're looking for opportunities to cash in on the online retail market that's expected to hit $4.88 trillion by 2021.

The credential stuffing attempts logged by Akamai are automated, thanks to bots. Bots can represent up to 60% of overall web traffic, but less than half of them are actually declared as bots – making tracking and blocking difficult. This dilemma is compounded by the fact that not all bots are malicious, as we discussed in Issue 1 of this year's *State of the Internet / Security* report.

> "
>
> **Bots can represent up to 60% of overall web traffic, but less than half of them are actually declared as bots – making tracking and blocking difficult.**

Play the Numbers

For criminals, credential stuffing attacks are a numbers game. They're counting on the fact that people recycle their passwords across different accounts. When this happens, a compromised set of credentials from one website quickly translates into dozens of others.

It's a two-step process; stuff the login page with the maximum amount of credential pairs to verify their validity, and once verified, take control of the compromised account. This second stage is commonly known as account takeover, or ATO.

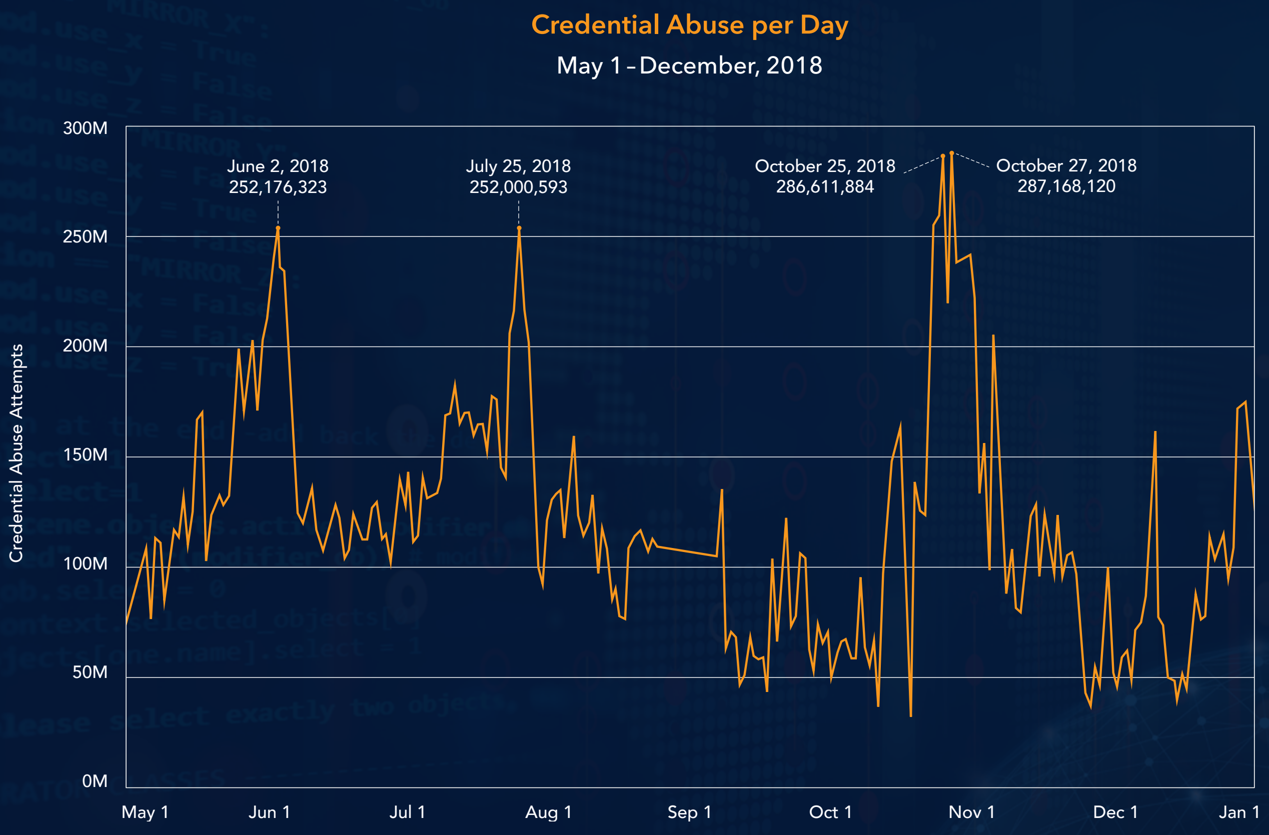


Fig. 1

Four of the top days for credential stuffing are highlighted between May 1 and December 31, 2018

Consider the 116 million accounts compromised during the LinkedIn data breach. Using this list of email address and password combinations, criminals targeted dozens of other websites in hopes that people were using their LinkedIn credentials elsewhere. These credential stuffing attempts led to several secondary account takeovers. This is why security professionals stress the use of password managers, as well as the use of long and unique password strings for each website.

Fighting Credential Stuffing Attacks Is an Uphill Battle

The battle against credential stuffing isn't an easy one to fight. When asked, 71% of the respondents to an Akamai survey conducted by Ponemon Institute said that preventing credential stuffing attacks is difficult because fixes that prevent such action might diminish the web experience for legitimate users.

On average, organizations report experiencing 12.7 credential stuffing attempts each month, with each attempt targeting 1,252 accounts. The reflexive action to just block the bots responsible for these attempts outright makes sense at first, but such a move might cause serious harm to the business if legitimate customers are impacted.

The same survey revealed 32% of respondents lacked visibility into credential stuffing attacks, and 30% said they were unable to detect and mitigate them. When asked if their organization had sufficient solutions and technologies for containing or preventing credential stuffing attacks, 70% of those responding said their organization was lacking when it came to such defenses.

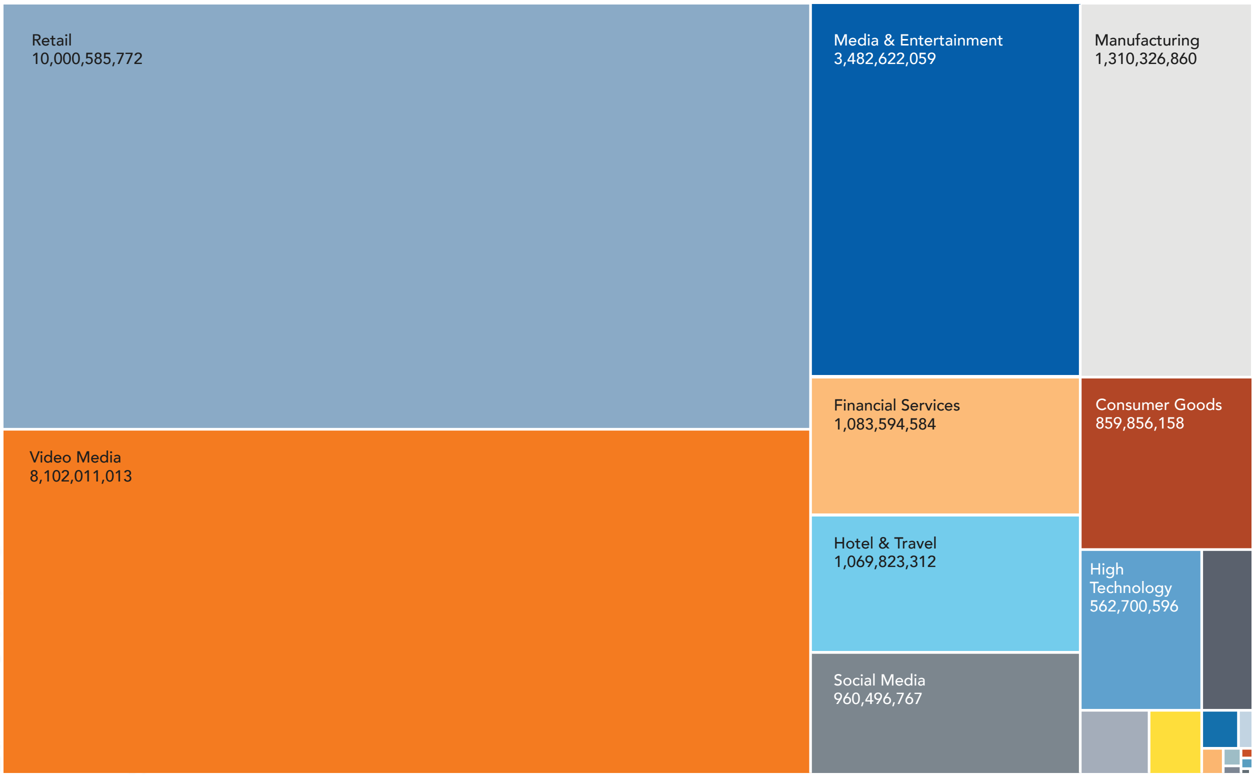
Credential stuffing attacks are a costly battle to fight as well. The survey determined that the baseline costs associated with such attacks, when considering application downtime, loss of customers, and IT overhead, amounted to annual totals of $1.7 million, $2.7 million, and $1.6 million, respectively.

Fig. 2

The combination of Video Media and Media & Entertainment saw 11.6 billion attempts

Credential Abuse Attempts by Vertical

May 1 – December 31, 2018



Top market segments

Overall

Over eight months in 2018, Akamai detected 27,985,920,324 credential abuse attempts. In Figure 2, we break this down into industries, using area to show the proportion to the whole. Each box represents a single vertical as recorded in our data. Not labeled in the bottom right corner are industries that only saw fewer than 250 million credential abuse attempts during this time, such as the automotive and public sector industries.

While retailers are the most popular targets of credential abuse attacks, the organizations that provide streaming media services came in a close second, suffering 8,102,011,013 attempts. Other media and entertainment organizations accounted for 3,482,622,059 attempts, while manufacturing sites received 1,310,326,860 attempts. We plan on exploring the attacks against media sites as a whole in future reports.

Retail

With 10 billion credential stuffing attempts during the eight-month reporting period, the retail industry is the largest targeted segment in our data. Within this industry, the apparel vertical experienced 3.7 billion attempts on its own, making it the most-targeted industry during the same time frame.

So why is retail, especially apparel, such a hot target? Short answer? Money.

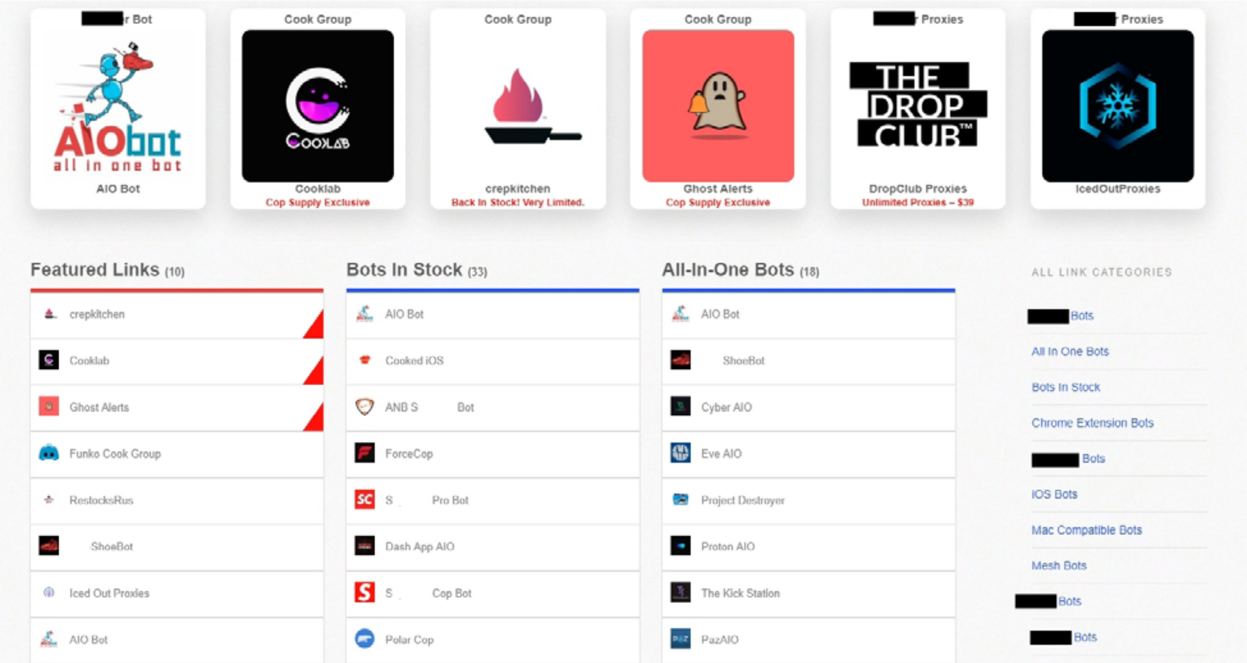


Fig. 3

An example of an AIO Marketplace. Offers include bots, as well as proxy and hosting services

[redacted] Bot
AIO Bot

Cook Group
Cooklab
Cop Supply Exclusive

Cook Group
crepkitchen
Back In Stock! Very Limited.

Cook Group
Ghost Alerts
Cop Supply Exclusive

[redacted] Proxies
DropClub Proxies
Unlimited Proxies – $39

[redacted] Proxies
IcedOutProxies

Featured Links (10)

- crepkitchen
- Cooklab
- Ghost Alerts
- Funko Cook Group
- RestocksRus
- ShoeBot
- Iced Out Proxies
- AIO Bot

Bots In Stock (33)

- AIO Bot
- Cooked iOS
- ANB S Bot
- ForceCop
- S Pro Bot
- Dash App AIO
- S Cop Bot
- Polar Cop

All-In-One Bots (18)

- AIO Bot
- ShoeBot
- Cyber AIO
- Eve AIO
- Project Destroyer
- Proton AIO
- The Kick Station
- PazAIO

ALL LINK CATEGORIES

- [redacted] Bots
- All In One Bots
- Bots In Stock
- Chrome Extension Bots
- [redacted] Bots
- IOS Bots
- Mac Compatible Bots
- Mesh Bots
- [redacted] Bots
- [redacted] Bots

Combined, the top 10 apparel brands are worth $111.3 billion, according to the BrandZ Top 100. However, the resale market for apparel alone is worth nearly 1 billion dollars. While the resale market is small by comparison, that's still a significant amount of money. The aftermarket economy is so lively, in fact, that reseller platform StockX says it facilitates about $2 million in transactions daily.

Within the retail industry, and particularly within the apparel vertical, the bots often associated with attempting credential stuffing and purchasing are called All-In-One bots, or AIOs. These bots are multifunctional tools that enable quick purchases by leveraging various evasion techniques and can target more than 120 retailers online. It isn't uncommon to see an AIO sold and designed with a specific retail outlet in mind, either.

The primary reason AIOs exist is to enter the resale market, but they can also be used to quickly check account and profile access on a given website using lists of usernames and passwords.

The criminal element targeting the retail industry has numerous options when it comes to accounts that have been successfully taken over. The personal information associated with the compromised account has worth, as does any unique status tied to it.

Retailers often offer discount codes or limited-edition items to known customers – it's part of their brand engagement strategy. Criminals who are successful with their credential stuffing attempts can hoard all of these perks and resell or trade them later. Sometimes the targeted account is used just to get a place in line during a special promotion, as existing customers are frequently placed higher in queue.

> "
>
> **Criminals who are successful with their credential stuffing attempts can hoard all of these perks and resell or trade them later.**

A successful AIO campaign may go completely undetected by a retailer, which might see the online sales and record-setting transactions as proof its product is in demand. They'll have little to no indication that its inventory clearing was automated and used to fuel a secondary-market or scrape information from its customers.

The problem with bots in the retail sales cycle is a systemic one. This isn't about a product that sells out quickly, creating a false sense of success. The use of AIOs deny the retailer the chance for engagement and value-add sales, inhibiting growth and brand enhancement. They create artificial scarcity, skew sales metrics and stock tracking, and hurt the retailer's customers and investors by placing information and the retailer's reputation at risk.

Deeper into retail

Credential Abuse

Retail Organizations by Type

May – December, 2018

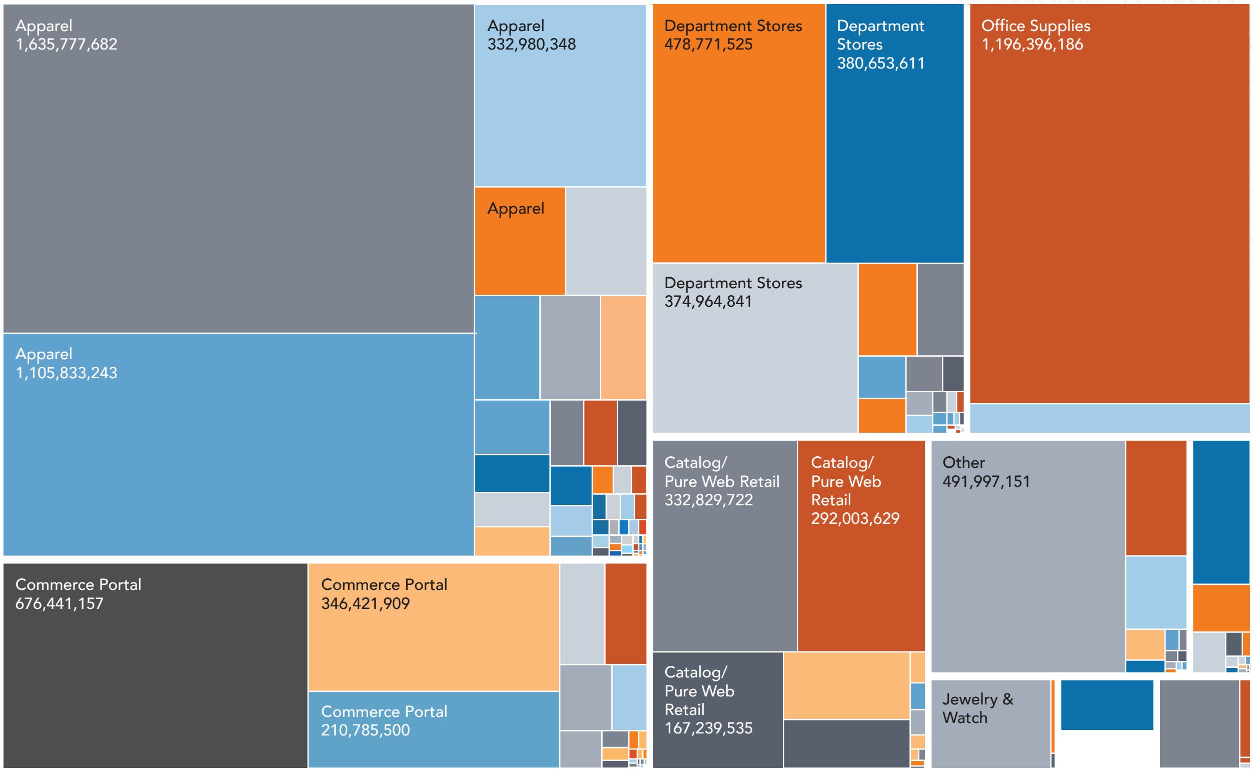


Fig. 4

A single organization was the target of 6% of all credential abuse attacks during the reporting period

Within the retail industry outside of the apparel vertical, Akamai tracked credential stuffing attempts against direct commerce (1.427 billion); department stores (1.426 billion); office supply stores (1.3 billion); and fashion, such as jewelry and watches (129,725,233). Each colored box in Figure 4 represents an individual organization, with businesses grouped by type and bounded by thicker white lines. For example, the upper left box represents a single organization that experienced 1.636 billion attacks.

As was the case with retail, the bots and bad actors are conducting these credential stuffing attacks and attempts with multiple goals in mind. When it comes to direct commerce – retailers that offer a single item or brand – the goal centers on accounts that have existing history, personal information that can be harvested, and unique deals and promotions. The same can be said for department stores, but there is an added bonus from criminals who can easily trade in compromised department store credit lines.

When credential stuffing leads to a successful ATO against an office supply chain or retailer, the attacker gets access to business and personal details, as well as other related information including purchase order (PO) contacts, which places the account holder at risk of Business Email Compromise (BEC) attacks or spear phishing.

Jewelry and accessories are viable targets as well, due to the fact that customers in these markets have a high net worth, so their personal details and account balances are a hot commodity.

In the financial services industry, where Akamai tracked 1.08 billion credential stuffing attempts during the reporting period, the objective is purely criminal. The actors responsible for these attacks are attempting to match compromised credentials with financial accounts. If the credential stuffing attempt leads to a successful ATO, the victim is exposed to having their finances siphoned off, or having their entire financial profile packaged and sold.

In the hotel and travel industry, where Akamai tracked 1.069 billion credential stuffing attempts during the reporting period, the actors conducting these attacks again have mixed goals. This industry includes retail as well as financial targets for criminals, as accounts can be scraped for personal information alongside reward and promotional information. There's a good deal of money to be made by compromising a rewards account for an airline or hotel, as well as reselling a room booked at a discount at full price or at a markup.

> **In the financial services industry, where Akamai tracked 1.08 billion credential stuffing attempts during the reporting period, the objective is purely criminal.**

Location

Finally, we come to location. As shown in Figure 5, the United States led the pack when it comes to credential stuffing source traffic, followed by Russia, Canada, Brazil, and India. Many of the AIO bots used are developed in the United States, so it isn't shocking to see it listed as the top source.

When it comes to targets, the United States is also at the top of the list with 22.47 billion credential stuffing attacks tracked, followed by China (2.01 billion), India (1.16 billion), Germany (792 million), and Canada (400 million).

Final thoughts

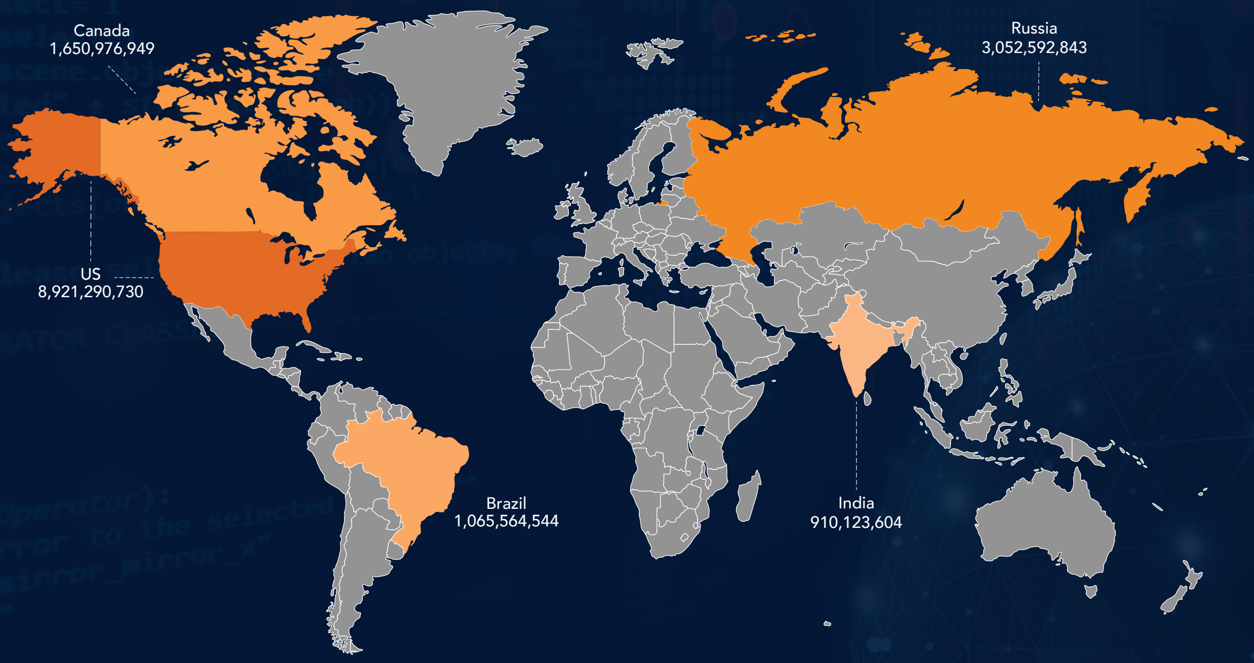
While credential stuffing attacks and ATOs are sometimes used for those playing the resale market game, the majority of these attacks are centered on business compromise, or collecting personal and financial information and either selling or trading it on the underground marketplace.

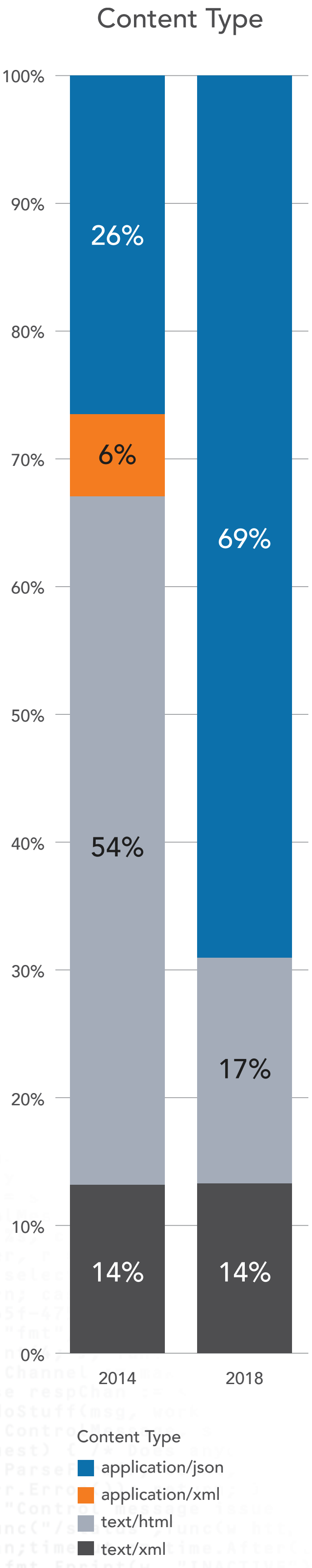
The only way to stop these types of attacks is to get better at detection and mitigation when it comes to the bots themselves, and to focus on keeping users from sharing credentials between websites. As long as passwords are recycled, credential stuffing and ATOs will continue to be a steady criminal enterprise.

Fig. 5

Source countries for credential stuffing attacks as tracked between May 1 and December 31, 2018

Top 5 Credential Abuse Source Countries





RISE OF API TRAFFIC

In 2014, Akamai researchers asked a relatively simple question: How much of the HTTPS traffic on our network is API compared to HTML? In other words, we wanted to know what portion of the traffic we see, and by extension the Internet as a whole, is content formatted for machines – some of which is triggered by human activity, and some of which is automated data exchanged behind the scenes without direct human interaction. The assumption had been that API traffic was a small portion of our traffic, but an informal analysis of our statistics revealed that API traffic accounted for 47% of all layout and data traffic we saw.

This was a major revelation – one that fueled a multitude of conversations in the past four years. We recently decided it was time to look at API traffic again, and the results were once again surprising: The traffic classified as APIs currently accounts for 83% of all hits, while HTML traffic has fallen to just 17%.

This shift in traffic patterns has significant ramifications in the security industry. Many, if not most, controls that have been historically used to protect the servers and systems that are the origin of traffic are focused on monitoring browser traffic. The mechanisms necessary to apply the same controls to API traffic may be less robust, harder to configure, or nonexistent in certain environments.

The traffic examined was drawn from the ESSL network, Akamai's secure content delivery network, designed primarily for secure transactions such as banking and retail. For more information about the collection of data, terms, and definitions used in this section, please see the extended description in the Appendix.

It's All About JSON

Our definition of API traffic, for the purposes of this report, is all HTTPS responses that contain a content type of JSON or XML on our ESSL network. While this may include some browser-based traffic, our examination determined that this was a very small minority of the traffic and would not significantly impact the measurements we're highlighting. We limited the organizations in our reporting to those that received more than 1 million hits during the 10 days we are examining.

Fig. 6

XML traffic from applications has almost disappeared since 2014

API traffic, and especially JSON, is generally based on solitary, atomic requests. In other words, each request is a stand-alone datagram, rather than being part of a stream or a multipart request. In contrast, while the HTML needed to construct a page might be relatively small, the dependent images and other code are generally quite extensive and consist of hundreds of objects.

When we looked at the content types in HTTPS traffic, we found that JSON is the single most popular content type. In the past, image files such as JPEG and GIF were the lion's share of the traffic on the ESSL network. Our current measurements show that even when counted together, these two file types don't account for as much traffic as JSON. On the other hand, if we count all image types as one, they do account for more traffic than JSON.

> **"**
>
> **Our current measurements show that even when counted together, these two file types don't account for as much traffic as JSON.**

We believe it is fair to say that our ESSL network is quickly evolving to serve API traffic as a primary use. This is because of some very large consumers of the traffic in the Media and High Tech verticals. These are the businesses that are serving our news, weather, streaming media, and games, in most cases.

As a whole, JSON traffic currently accounts for four times as much traffic as HTML does – but even if the top five organizations are removed from the analysis, JSON is responsible for twice as much traffic as HTML. There's no debating that this API traffic has become a major part of what organizations of all sizes are both consuming and producing.

API Hits

Vertical and Organization

(Millions)

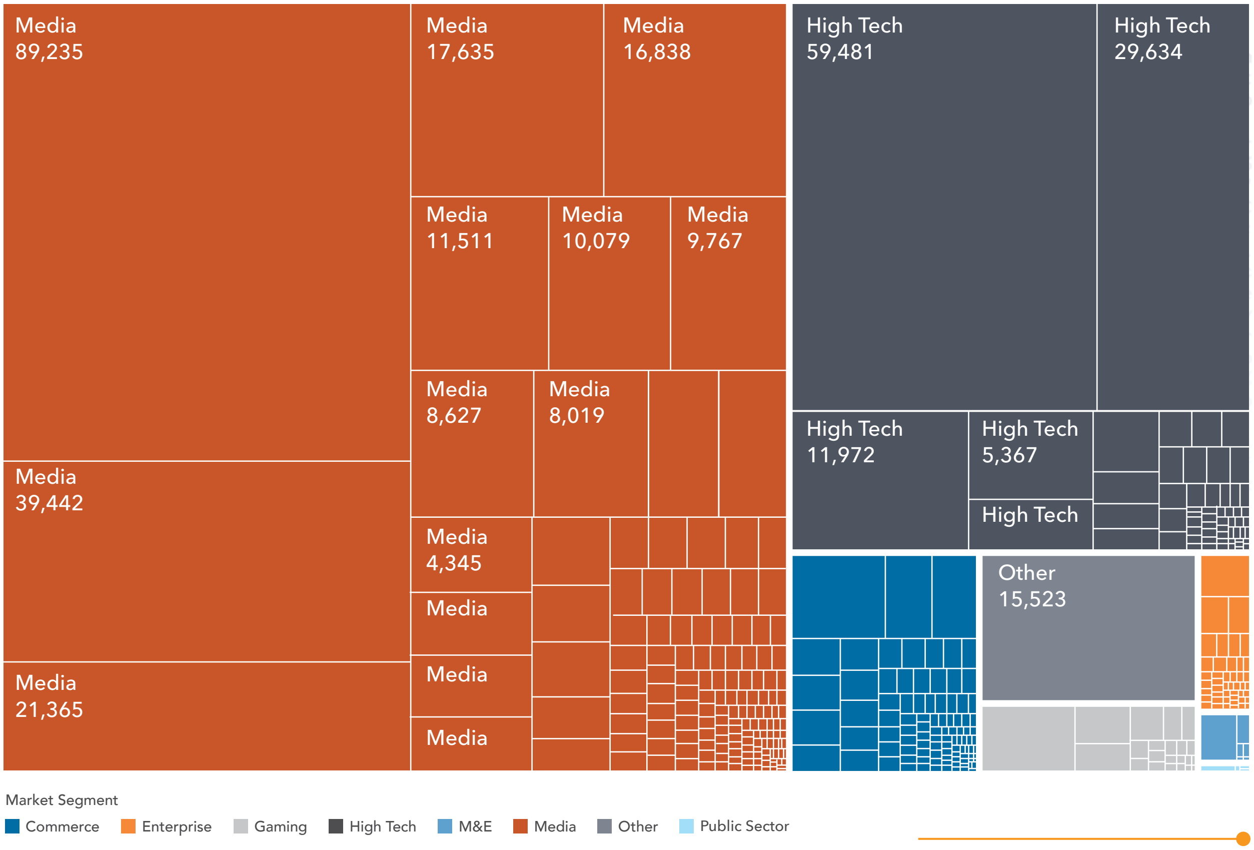


Fig. 7

Media organizations are the largest users of APIs by a significant margin

In Figure 7, we highlight the distribution of API traffic; at a glance, it's apparent that Media as a whole is responsible for almost two-thirds (63%) of hits. This is driven not only by a single large consumer, but also by the high number of media sites relying on API traffic. The second-largest producer of API traffic is in the High Tech vertical. Our categorization of verticals is based on Akamai classifications and traffic, and may not align completely with industry standards.

Aside from the security concerns, the shift towards API traffic is important from a performance perspective. Cacheability, a measurement of how much of the traffic can be saved on the servers used by content delivery networks such as Akamai, is comparable between HTML hits and API hits. While one-third of hits recorded were marked as "no-store," and therefore preventing caching, the cache hit rate for API traffic was actually slightly higher than that of HTML traffic. This means that a significant amount of the API traffic is being offloaded from the origin servers of the customer and is being served from edge servers near the end user. This significantly reduces the load on both the origin server and the Internet backbone as a whole.

The majority of JSON documents are not being consumed by browsers. Smartphones, applications, and embedded devices (such as gaming consoles, streaming devices, and smart TVs) are responsible for at least 66% of API traffic. In contrast, all browsers combined are responsible for only 27% of API traffic, and no other contributors are responsible for more than single-digit percentages.

Smartphones utilize a lot of APIs and require a cellular network to connect to the Internet. Because this communication is largely program driven, a higher fraction of cell access is from APIs, rather than HTML. The fraction of requests coming from cellular networks is twice as high as for HTML, at 37% vs. 18%.

Most of our examination of API and HTML traffic to this point has concentrated on the number of hits we recorded, but another important part of the discussion is the bits and bytes required to support each type of communication. Both API and HTML have a median object size of slightly under 1 KB.

> **"**
>
> **The majority of JSON documents are not being consumed by browsers. Smartphones, applications, and embedded devices (such as gaming consoles, streaming devices, and smart TVs) are responsible for at least 66% of API traffic.**

API Traffic by User Agent

TYPE	UA	%
Browser	Chrome	13%
	Mobile Safari	8%
	Firefox	2%
	IE	2%
	Edge	1%
	Safari	1%
	IE Mobile	0%
Non Browser	Other	66%
	CFNetwork	3%
	Apache HttpClient	2%

Fig. 8

The majority of API traffic is for custom applications and not easily categorized

However, the high end of traffic is where they diverge significantly; the 95th percentile of API traffic is 18 KB for JSON traffic, while HTML traffic is 105 KB. This is important, because a typical TCP segment is 1,400 – 1,500 bytes, so large objects won't fit in a single segment. Moreover, a small number of segments can be sent together in a single round-trip time (RTT) congestion window used by Akamai servers. Because of this, an 18 KB message can fit within the initial window of a single RTT, while a 105 KB message cannot. The result is that larger HTML traffic will take multiple RTTs to be transferred, while JSON traffic will mostly be transferred within a single RTT, even on a newly established connection.

Takeaways

In the past four years, we've seen API hits on the Akamai network grow from under half of all HTTP traffic to being 83% of the traffic, crowding out HTML hits. For security practitioners, this is vitally important – not all tools are capable of handling the shift, and you may be missing a major source of malicious traffic in your defenses. For our teammates whose responsibility is the performance side of the servers, they may also be missing a huge part of the equation if they're not taking JSON and XML traffic into account.

> **"**
>
> **For our teammates whose responsibility is the performance side of the servers, they may also be missing a huge part of the equation if they're not taking JSON and XML traffic into account.**

Applications are different from traditional web pages; they don't need the information on layout and style since that information is already included at build time. Instead, applications are requesting updated data and images, driven by news cycles, the weather, and sports. Or maybe it's your gaming and streaming systems getting updates, so you know about the latest releases. The data they are receiving is much smaller when you look at individual requests, but the volume of these requests is only going to grow with time.

IS IPV6 BEING UNDERREPORTED?

IPv6 Adoption in DNS Recursive Resolvers

In this section, we look at IPv6 adoption by the Domain Name System recursive resolvers. Recursive resolvers are a critical component of the Internet ecosystem because of their use by clients in resolving hostnames to IP address and their caching of responses to both reduce resolution delay and the load on authoritative servers.

We previously reported statistics on IPv6 adoption on the Akamai platform in June 2018. The average percentage of content requests on dual-stack enabled hostnames seen by the Akamai platform was approximately 45% – significantly higher than it was two years ago. The network traffic under analysis is client (end user)-to-edge and is informative of edge network adoption of IPv6. We explore data from DNS recursive resolver-to-authoritative-server to learn about adoption among the core components of the Internet.

High-level statistics on DNS traffic at Akamai's authoritative nameservers show that the vast majority of traffic is still IPv4. Only 11% of traffic was IPv6 as of July 2017 – much lower than the 45% of content delivery traffic that is IPv6. To understand why IPv6 adoption appears much lower in these core components, we take an in-depth look at Akamai's DNS traffic.

> "
>
> **Only 11% of traffic was IPv6 as of July 2017 – much lower than the 45% of content delivery traffic that is IPv6.**

First, we note the importance of analyzing dual-stack resolvers. This is non-trivial because there is no obvious way to correlate DNS requests over IPv4 with DNS requests over IPv6 and identify the single recursive resolver that generates both sets of traffic. So, here we leverage the fact that recursive resolver software will, under some circumstances, use multiple interfaces in the resolution of a single hostname, if multiple interfaces are available (e.g., dual stacking). Using DNS traffic logs from July 2017 containing 429K unique recursive resolver IP addresses, we first cluster the IP addresses based upon the technique described in the "Characterization of Collaborative Resolution in Recursive DNS Resolvers" white paper.

The IP addresses within a cluster are related to one another by being observed within the same DNS resolution. Clusters with both IPv4 and IPv6 addresses, therefore, likely include dual-stack resolvers. Six percent of the clusters appear to be made up of dual-stack resolvers. The dual-stack clusters tend to be a bit larger than other clusters; accounting for that, we estimated that 7% of all recursive resolvers are dual stacked as of July 2017. We found that one of the primary reasons for slow growth in IPv6 traffic among recursive resolvers is operators not configuring the resolvers with IPv6 addresses.

However, that is not the only reason. When we focused on the dual-stack resolvers, we examined their choice of interface to use for DNS resolution. While only 7% of recursive resolvers are dual stacked, they also account for the highest-traffic-volume resolvers. In all, dual-stack resolvers sent 37% of the DNS queries in the DNS traffic logs. So why is less than 11% of DNS traffic over IPv6? Clearly, dual-stack resolvers are preferring IPv4 over IPv6. For each cluster, we calculated the ratio of DNS queries using the IPv4 interfaces versus the IPv6 interfaces, and found the median ratio is 11:1. Less than 10% of the clusters favor IPv6 over IPv4. Our results show that a second reason for low IPv6 share in recursive resolvers is due to the vast majority of dual-stack resolvers preferring to send DNS queries over IPv4 than IPv6.

There are several forces that could be impacting the decision process. First, many domains have inconsistent numbers of IPv4 and IPv6 delegations. If a recursive resolver has more delegation options of IPv4 than IPv6, then it may seem to prefer IPv4 to IPv6 even when uniformly distributing queries among the delegations.

Second, many recursive resolver software packages represent known weight delegations according to inverse latency. If IPv6 connectivity is poor, resulting in high latency, recursive resolvers may prefer IPv4 for that reason. Finally, operators of recursive resolvers may enforce policy decisions to use IPv4 over IPv6 as a more familiar technology.

> "
>
> **IPv6 adoption by DNS recursive resolvers appears stunted in comparison to adoption at the edge.**

IPv6 adoption by DNS recursive resolvers appears stunted in comparison to adoption at the edge. There are a variety of causes, and plenty of work that may be done. Authoritative DNS server operators should ensure that they are providing as many IPv6 delegation options as IPv4 and that the performance of the IPv6 delegations is on par with the IPv4 delegations. Recursive resolver operators should check whether they have any policy in place preferencing IPv4 and make sure that deployments are all dual stacked.

Patterns in IPv4/IPv6 Address Assignment

In our study of dual-stack recursive resolver clusters, we observed patterns in IP address assignments that could be used to more rapidly associate IPv4 and IPv6 addresses in the future.

In the data, we observed a variety of ways to relate the IPv4 and IPv6 addresses of a dual-stack resolver: (i) the IPv4 octets embedded as the final 4 hextets (e.g., 1.2.3.4 and 89ab::1:2:3:4), (ii) the final IPv4 octet equal to the final IPv6 hextet (e.g., 1.2.3.4 and 89ab::4), and (iii) the full IPv4 address embedded within the IPv6 address, but not in the final 4 hextets (e.g., 1.2.3.4 and 89ab::1:2:3:4:5678).

We also observe incremental IP assignment patterns among the dual-stack resolvers within the same autonomous system (AS) that also aid in positively identifying dual-stack resolvers. For example, IP assignment can appear incremental in both IPv4 and IPv6, but shifted: w.x.y.z forms a cluster with a:b:c::${z+C} where C is a constant. Clearly, operators use a variety of patterns in IPv4/IPv6 address assignment. The patterns, if known a priori for an autonomous system, may be useful for matching IPv4 and IPv6 sides of dual-stack resolvers, simplifying the process of discovery greatly.

Clusters of Recursive Resolver IPs Defend Against Cache Poisoning

Going back to the clusters of recursive resolver IP addresses mentioned previously, we found that many clusters (38%) contain more than two IP addresses and are therefore more than just dual-stack resolvers. Instead, recursive resolver operators are using multiple IPv4 and/or multiple IPv6 addresses within the same DNS resolution. These larger clusters account for nearly 72% of the DNS traffic we observe.

> "
>
> **While many clusters contain two IP addresses, some clusters are very large, with greater than 10 IP addresses being common.**

There may be several reasons for this behavior, including load balancing across physical hardware, but there is also a security advantage. DNS cache poisoning attacks, e.g., the Kaminsky attack, may require spoofing DNS responses that match the DNS queries sent by the recursive resolver. The fields of the DNS response that must match the DNS query are (i) source IP/port, (ii) destination IP/port, (iii) DNS question, and (iv) DNS transaction ID. Within a DNS recursive resolver cluster, any given DNS query may be sent from any of the IP addresses in the cluster.

Thus, spoofing a matching DNS response must – in addition to the other fields – also guess the correct destination IP address from the cluster. While many clusters contain two IP addresses, some clusters are very large, with greater than 10 IP addresses being common. This significantly increases the difficulty of DNS cache poisoning attacks. As such, recursive resolver clusters are a viable method of mitigating the risk of cache poisoning.

Looking forward

Our hope is that security teams and security professionals will continue to grow more integrated with business units and their concerns in coming years. We have evolved from the "Department of No" mentality that existed in the past. The security industry as a whole has grown, but we have a lot more growing to do. Our profession touches everything now, and security has taken center stage when it comes to business planning and growth.

Each of the stories in this issue of the *State of the Internet / Security* report looked at aspects of security that are often overlooked by the mainstream but are nevertheless important to day-to-day operations. These stories create a backdrop for what we expect to see in the upcoming quarters and years.

One of the largest stories to watch is the API story. Since API traffic has eclipsed HTML traffic, security teams and businesses need to address this new reality. As mentioned, many tools currently in use are unable to deal with this shift in traffic types, which results in blind spots. With the proliferation of mobile devices and IoT technology, this trend will not be going away any time soon. How businesses choose to meet this challenge and shift will have a large impact.

> **"**
>
> **Since API traffic has eclipsed HTML traffic, security teams and businesses need to address this new reality.**

Related to the API story, the credential stuffing report demonstrates the power and heightened risk recycled credentials can have on various markets. This isn't the first time that security professionals have recommended not using the same credentials across different accounts, but sometimes old habits die hard. While we focused on retail this time around, retail isn't the only industry affected by these attacks. The impact credential stuffing has on other markets and industries – and the devastating impact of account takeover attacks on businesses – can't be understated or ignored.

The Internet is a quickly changing landscape, and these trends are just the beginning of a large shift. Security teams and professionals must constantly think outside the box to develop new ways to keep users and businesses secure and safe.

Appendix: Methodologies

GENERAL NOTES

We make a conscious choice to keep the mention of products to a minimum in the *State of the Internet / Security* report, only mentioning the products as much as needed to explain where and how the data was collected. In recent reports, we've been working to expand upon the description of the data collection so that readers can put our narrative and statistical data in context.

With the topics we're reporting on in this issue, especially our work on API traffic, we felt it was necessary in the discussion below to go deeper into both our solutions and the way we're using terminology in our content. There are enough vagaries in the terminology used by different segments of the Internet tech community that it's valuable to clarify our specific usage.

TOOLS OF MASS (RETAIL) DESTRUCTION

The data used to highlight the issues with credential abuse in this text was drawn from an internal tool called Cloud Security Intelligence (CSI). This tool is a repository of data from multiple product lines. The data is transitory, being deleted in 90 days or fewer. The full data set is currently in excess of nine petabytes and growing.

Credential abuse attempts were identified as unsuccessful login attempts for accounts using an email address as a username. In order to identify abuse attempts, as opposed to real users who can't type, two different algorithms are used. The first is a simple volumetric rule that counts the number of login errors to a specific address. This differs from what a single organization might be able to detect because Akamai is correlating data across hundreds of organizations.

The second algorithm uses data from our bot detection services to identify credential abuse from known botnets and tools. A well-configured botnet can avoid volumetric detection by spreading its traffic amongst many targets, by using a large number of systems in its scan, or spreading the traffic out over time, just to mention a few countermeasures.

"

A well-configured botnet can avoid volumetric detection by spreading its traffic amongst many targets, by using a large number of systems in its scan, or spreading the traffic out over time, just to mention a few countermeasures.

RISE OF API TRAFFIC

Similar to other large organizations, Akamai's Intelligent Edge Platform is not a single monolithic entity. Akamai's roots are as a Content Delivery Network, and many of the solutions we provide today have grown from those roots.

The Enhanced SSL (ESSL) service was originally created to be a separate network segment that satisfied the requirements of the Payment Card Industry Data Security Standards (PCI-DSS). Designed for merchants, retail organizations, and financial services institutions, the ESSL network has additional physical, logical, and digital controls required by PCI and other regulatory and compliance regimes. Despite the name, ESSL supports TLS and other modern encryption schemas.

Since the inception of ESSL, the desire to use secure, encrypted systems for all traffic has become the standard, rather than the exception. Because of this change, nearly 20% of all traffic served by Akamai uses the ESSL network – approximately 8 Tbps at this time. When we use hits to measure traffic instead of bits, ESSL is one of the main Akamai networks and serves the majority of our HTTPS-based traffic.

> **"**
>
> **Since the inception of ESSL, the desire to use secure, encrypted systems for all traffic has become the standard, rather than the exception.**

In our analysis of API and HTML traffic, we examined traffic on the ESSL network over a 10-day period in October 2018. Similar queries over shorter time spans have been performed since, and show nearly identical trends.

The queries were specifically looking for API and HTML traffic and do not include images or other traffic that is not a form of API traffic (JSON + XML) or HTML. We made a conscious effort to differentiate between hits (requests/responses) and traffic (the bits and bytes sent by volume).

Credits

State of the Internet / Security Team

Tony Lauro, Senior Manager, Security Strategy – Tools of Mass (Retail) Destruction
Martin Flack, Principal Architect – Rise of API Traffic
Moritz Steiner, Principal Architect – Rise of API Traffic
Kyle Schomp, Performance Engineer Senior II – Is IPv6 Being Underreported?
Rami Al-Dalky, Intern – Is IPv6 Being Underreported?

Editorial Staff

Martin McKeay, Editorial Director
Amanda Fakhreddine, Sr. Technical Writer, Managing Editor
Steve Ragan, Sr. Technical Writer, Editor

Creative

Benedikt Van Holt, Art Direction
Brendan John O'Hara, Graphic Design
Georgina Morales Hampe, Kylee McRae, and Murali Venukumar, Project Management

Akamai ***Intelligent Security Starts at the Edge***

Akamai secures and delivers digital experiences for the world's largest companies. Akamai's intelligent edge platform surrounds everything, from the enterprise to the cloud, so customers and their businesses can be fast, smart, and secure. Top brands globally rely on Akamai to help them realize competitive advantage through agile solutions that extend the power of their multi-cloud architectures. Akamai keeps decisions, apps, and experiences closer to users than anyone – and attacks and threats far away. Akamai's portfolio of edge security, web and mobile performance, enterprise access, and video delivery solutions is supported by unmatched customer service, analytics, and 24/7/365 monitoring. To learn why the world's top brands trust Akamai, visit www.akamai.com, blogs.akamai.com, or @Akamai on Twitter. You can find our global contact information at www.akamai.com/locations. Published 02/19.