TICKETING TECHNOLOGY FORUM

DISCUSSION PAPER #5

Protecting Ticket Sales from Bot Traffic

The fifth of our *#TTF16* Discussion Papers looks into protecting online sales from unpleasant bots - including a real life case study from Quakecon 2016.

BACKGROUND INFORMATION

While many ecommerce industries are impacted by automated bot traffic hitting their websites, bots present a broad and more complicated set of issues to ticketing websites.

WHAT IS A BOT?

A bot, also called a scraper, spider or crawler, is an automated program that carries out requests against your website. There can be a wide range of reasons why these programs exist, both positive and negative, or somewhere in between. Think of bots like a traffic light system – greens, ambers and reds.

Bots do not necessarily serve the same purposes or work in the same way, but they do have some things in common. Bots have all been written with a specific purpose in mind; whether that is good, bad or indifferent.

> ***"The volume of bot traffic has increased dramatically in recent years"***

They all add load to your web servers whilst carrying out their tasks; how much load depends on how many bots hit your servers at the same time and what their specific tasks are. In this respect the timing of when a bot or bots hit your website can be critical.

All bots also command operational decisions from every online business, such as whether each bot is to be authorised, blocked or that some other action is required.

The volume of bot traffic has increased dramatically in recent years, leading to third parties developing a wide variety of bots which

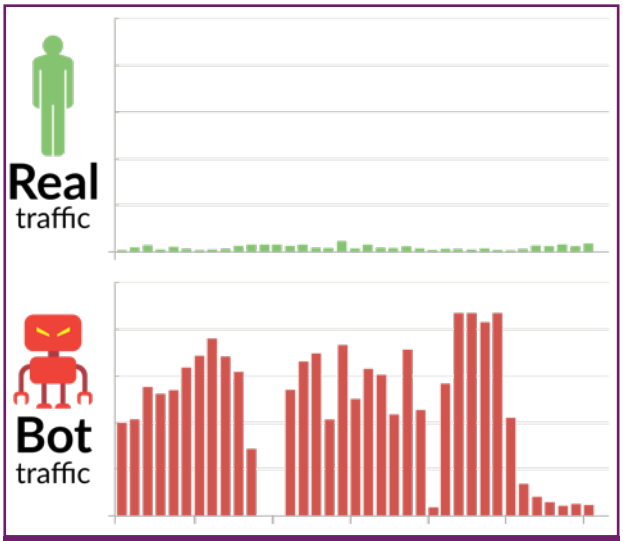


Fig 1: A real life example from a client who found that 96% of traffic to some parts of their site were bots"

people or businesses can purchase for their own usage. In short, no one now needs to be a developer to utilise bot technologies and this in itself has contributed to the vast bot traffic volumes now seen.

In light of this, awareness of what bots are hitting on your website and management strategies for each are important for the ongoing success of your website trading.

GOOD BOTS (GREEN)

It is important to note that not all bots are bad; search engine bots such as Google or Bing for example are generally welcomed. These bots rarely cause a problem, drive business to your website and can be controlled to an extent using tools provided by the search engines. Be aware though that they can generate a lot of load on your servers while they are scraping your site, so

"Unofficial crawlers are unlikely to be aiding your business, and unlike the search bots, the data will likely not be used to drive business directly your way"

for load sensitive sites they still need management.

For example, search bots will happily crawl your website at any time, generating load, and may be the straw that breaks the camel's back, either pushing the website offline or pushing the system into an unacceptable slowdown at busy times.

No ticketing site can afford for this to happen, with ticket allocations as well as loyal customers at stake. Therefore even the most benign of bots, the search bot, can be the cause of such problems and needs appropriate management.

"Bots have no concept of your peak time, they will happily start crawling your website, generating load, and potentially be the straw that broke the camel's back"

Other examples of good bots would be partners and services that you have entered into an arrangement with to gather data from your site for analytics, price shopping or ad retargeting purposes. These can inadvertently impact your site through introducing overhead on your servers during peak periods.

This can all cause operational issues if not managed carefully and, without good analysis for usage patterns, can actually cause unintended harm.

BAD BOTS (RED)

Not all bots are acting in your interest though. There are many bots that are automating activity against your website for unsolicited reasons, sometimes even illegally.

Negative impact of bots include:

- Data and content can be harvested for use on other sites or by competitors
- Site performance impacted when it matters most
- Staff time spent chasing bots instead of operating and enhancing your platform

There are third party services, open source tools or even custom scripts that can be employed to trigger a bot attack. These attacks can be launched intentionally from personal machines, from groups of machines hosted in physical data centres or in the cloud, or in extreme circumstances illegal botnets of up to tens of thousands of compromised PCs all over the world.

A growing proportion of these bots are intelligent, simulating real users and real browsers. The real browser bots can affect analytics tools, such as Google Analytics, and impact usage and conversion figures as they execute the JavaScript used by these tools.

These business metrics rely on the data being from real customers, not automated bots, so decisions made on the data will impact an online business as they no longer reflect real user behaviour.

Having visibility of all bots hitting your servers and their impact is critical in order to implement mitigation strategies and protection from malicious bots. Such retrospective analysis is important for ongoing bot management, however an upfront strategy and protection from known malicious bots is recommended to stave off an attack before it impacts your website.

UNSPECIFIED BOTS (AMBER)

As well as bots that are good and bad there are a selection of bots that are unknown, operating in the grey area between the two extremes. New bots are being released all the time and any hitting your website need to be identified, assessed for risk and managed accordingly.

These bots are unauthorised but not necessarily detrimental to you and could potentially be of benefit. An example of this would be a price comparison site that harvests data and then links users through to complete a purchase.

For such services a commercial decision needs to be made as to whether they return more value than the operational overhead of supporting the scraping activity.

Visibility of these bots and an understanding of their purpose is critical in order to take the appropriate action; authorise access to the website, block access or other appropriate action.

ARE BOTS ILLEGAL?

While bots that are used to bring websites offline are clearly illegal, the question of legality around other bots, such as those used for data harvesting, is an ongoing debate, with legislation continually being discussed but change not looking imminent.

To truly manage bots in an effective way, ticket selling website must stay a step ahead of the bad bots on the technical front, whilst managing any potentially unwanted side effects from the otherwise benign bot.

"Serving pages to bots adds overheads to servers. This is typically 20-30% of traffic but up to 96% has been seen"

The reality of the modern era is that bot traffic management needs to become an ongoing task of online operations.

THE COST OF BOTS

Serving pages to bots adds overheads to servers. This is typically 20-30% of traffic but up to 96% has been seen.

CASE STUDY: QUAKECON 2016 - BEATING THE TRAFFIC

QuakeCon is North America's largest annual "bring-your-own-computer" convention and tournament, based around the popular id Software game franchise Quake.

Described as the "Woodstock of gaming", it has been running for 20 years and is attended by up to 10,000 people each year from around the world.

The 2016 event will take place on August 4-7 in Dallas, Texas. While general admission is free, there are a limited number of reserved seating tickets available for $40 each, which are sold through the QuakeCon website.

Anticipating the high demand once tickets became available to buy, the team at QuakeCon sought a solution to ensure visitors could get onto the site and buy tickets in a fair, orderly manner, without bringing the site down and whilst avoiding expensive infrastructure burst costs.

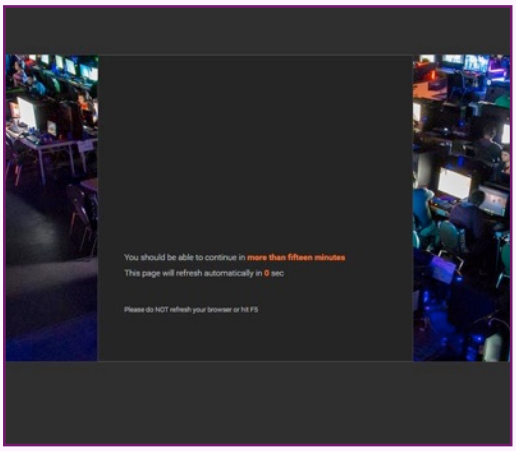
With TrafficDefender's online traffic management and queue system in place, QuakeCon were able to direct visitors into a waiting room ahead of the on-sale going live.

When the on-page countdown hit zero, the ticket sale went live and almost 2,000 visitors were placed into a queue to buy tickets.

Not only did this give visitors a fair system to buy tickets in an orderly fashion, it also allowed QuakeCon to control how much traffic hit their infrastructure and avoid performance issues or outages.

CASE STUDY CONTINUED:

QuakeCon were impressed by the technology behind TrafficDefender and how their tech-savvy visitors could not find a way to hack around the queue, and malicious bots and scrapers could be blocked to ensure only genuine ticket buyers were utilising their available resources, maintaining a fair experience for all.



Thanks to TrafficDefender, many more visitors were able to queue than could have been comfortably served simultaneously otherwise by the site, with a very small abandonment rate from the queue.

In real term this means that a large proportion of your operational budget may be being spent on servicing bots, not your customers.

BEATING THE BOTS

Just as there are methods to identify bots and determine their intent in visiting your site, there are also ways of controlling their access. But unfortunately it's more complicated than only blocking their IP addresses. Anyone wishing to scrape information from a website can do so with easily accessible tools like Shader.io, which automatically rotates IP addresses to circumvent IP blacklists.

IP addresses can change ownership without warning, which risks locking out real users, so you need to utilise additional means to manage bots whilst ensuring your customers and 'good' bots are not penalised.

Bot behaviour is getting more sophisticated all the time, so the methods of identifying bot behaviour need to move just as quickly.

Blanket blocking of public cloud IP Ranges (Amazon Web Services, Microsoft Azure, etc.) may seem like a solution but can have unintended consequences, such as inadvertently blocking your own traffic, legitimate traffic or partners who are providing a service you are paying for.

TRAFFIC MANAGEMENT

One solution that has been developed to solve these problems is by using a web traffic management system. A sophisticated system will manage the flow of traffic into websites, utilising strategies to distinguish between bots and real users to optimise server capacity, maintain uptime and deliver a smooth experience to customers during on-sales.

Advanced systems can also place excess traffic into a queue during peak times, protecting the website from becoming overwhelmed. Once capacity is freed up, waiting visitors are passed through to the website in a fair and orderly fashion.

However you address the problem of bots, be they green, amber or red, it's essential to have an awareness of the impact they are having on your business. ■

More on traffic management? Join us at #TTF16

ABOUT THE AUTHORS

Jeremy Gidlow is Managing Director, and Steve Vorley is Product Owner for TrafficDefender at Intechnica. Intechnica have a decade of experience designing, building and running business critical systems, enabling increased development throughput, and advising on how to transform software delivery. TrafficDefender is Intechnica's web traffic management and queuing system for retail and ticketing websites. TrafficDefender ensures websites can scale to meet any peak both by queuing excess visitors to keep websites from overloading, and by intelligently controlling the mix of real visitors and bot traffic. ■