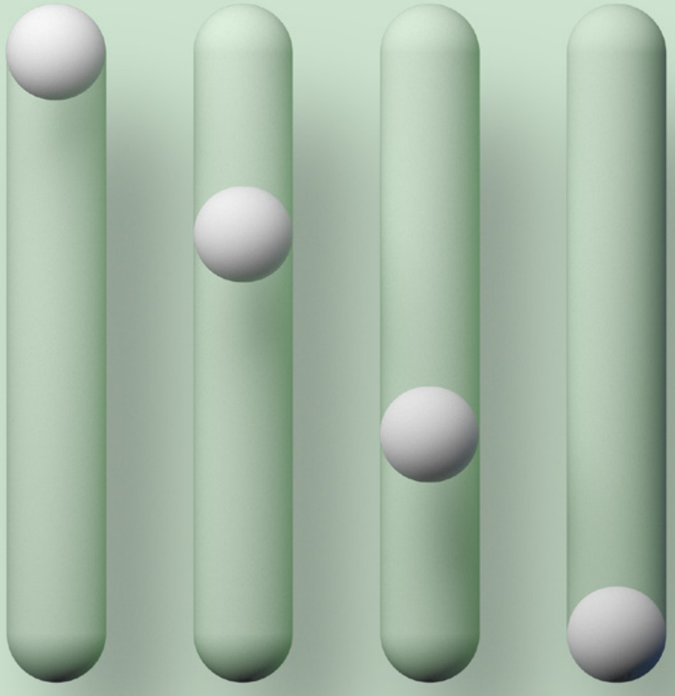
Leveraging AI and Automation to Stop Typosquatting Attacks

Top Level Domain Expansion Expands Attack Surface for Cyber Criminals



BOLSTER

Executive Summary

Safeguarding corporate assets is an ongoing challenge. Typosquatting has emerged as a major problem, one that top management may not recognize but can significantly damage a brand. Criminals use it to install malware; steal sensitive, personal, and financial information; and hold computers hostage.

As ICANN expanded the number of Internet naming domains, problems have risen. The increase was needed because businesses rely on the global network to market, deliver, and support their products and services. With the number of domains expanding so did the number of potential entryways for hackers seeking to tarnish a brand. Today, there are more than 1,400 generic Top Level Domains, (each offers many potential entryways), and the number continues to increase. This creates a huge attack surface, which now results in about 50 million users falling victim to typosquattting ruses every year.

Companies have struggled to address the problem because they lacked a viable defensive mechanism. The old strategies of buying domain variations is quite costly, and trying to remove fraudulent activity chews up time and money; is manpower intensive, and usually ineffective.

The only possible way to address the problem is through artificial intelligence and automation. However, most fraud detection systems only solve a piece of the problem, offer rudimentary learning models, and are largely manual processes with no automation. In response, Bolster developed the industry’s most accurate algorithm with a false positive rate of 1 in 100,000. Its system takes down over 99% of fraudulent sites within 24 hours and does so without requiring manual intervention. Problem solved.

New Top Level Domains Widen the Attack Surface

Hackers rely on domain name missteps to tarnish brands

For every action, there is an equal and opposite reaction, noted Sir Isaac Newton. As the Internet grew in popularity, naming convention shortcomings arose as demand for domain names outstripped the supply. In 2017, ICANN began releasing what grew to more than 1,400 new generic Top-Level Domains (gTLDs), and that number continues to increase. The influx provided organizations with much needed Web addresses but also became home to hackers who carved out lucrative businesses in typosquatting.

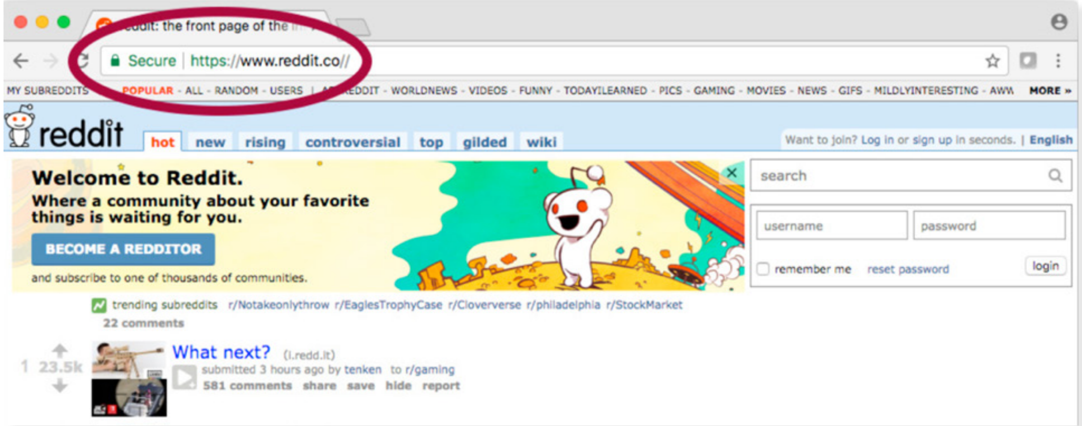
The domain name shortage limited corporations' ability to market their services via the Internet. Rather than continue to play catch up, ICANN constantly expands the domain name list—adding two in Q4, 2019. In addition, additional names arise from extending the existing 312 country code Top Level Domains (ccTLDs) and numerous sponsored TLDs (sTLD's). These extensions address business needs but widen the potential attack surface by creating more potential variations of a company's domain to be used by the bad guys.

Typosquatting (also known as domain squatting and URL hijacking) is registering, trafficking in, and using an Internet domain name in bad faith. Here's how it works. Human beings make mistakes, and hackers leverage them for their own malfeasance. Sometimes, individuals misspell a Web address. The bad guys know this mix-up occurs regularly and set up fraudulent sites that take advantage of the miscues. Their site names are very close to legitimate ones.

Typo	Example
A different top-level domain	.cm vs .com
Addition to end of domain	domains.com
Hyphenation	do-main.com
Insertion of va owel or consonant	domaiin.com
Swapping vowel or consonant	domian.com
Subdomains broken by period	domain.domain.com

In one case, the criminals went to great lengths to create a fake Reddit site that looked nearly identical to the authentic site. They used the .co suffix for Columbia to mimic the popular site, Reddit (reddit.co instead of reddit.com) and even included an illicitly-acquired Reddit SSL certificate. These steps gave site visitors the impression that they were visiting a legitimate Reddit site since the browser would be showing a "secure" site with the green lock icon.

Customer Facing Fraud Site with Matching SSL Certificate Screen Shot



Secure | https://www.reddit.co//

Welcome to Reddit.

Where a community about your favorite things is waiting for you.

and subscribe to one of thousands of communities.

Want to join? Log in or sign up in seconds. | English

What next? (i.redd.it)

submitted 3 hours ago by tenken to r/gaming

581 comments share save hide report

eddit: the front page of the
MY SUBREDDITS
POPULAR - ALL - RANDOM - USERS | ASKREDDIT - WORLDNEWS - VIDEOS - FUNNY - TODAYILEARNED - PICS - GAMING - MOVIES - NEWS - GIFS - MILDLYINTERESTING - AWW
MORE »
reddit
hot
new
rising
controversial
top
gilded
wiki
BECOME A REDDITOR
search
username
password
remember me
reset password
login
trending subreddits r/Notakeonlythrow r/EaglesTrophyCase r/Cloververse r/philadelphia r/StockMarket
22 comments
1
23.5k

All of these sites are legally registered but rely on whisking unsuspecting users to bogus locations where they fall victim to a variety of ruses. Once there, their views could be used to generate fraudulent ad clicks, download malicious files, or install malware on the user's computer. These attacks are often achieved through a phishing campaign that uses a typo squatted domain to fool the user into thinking it is from a legitimate source.

The explosion of new gTLDs allows criminals to widen their attack points and have more chances to fool an unsuspecting user. In the second quarter of 2020, Bolster examined nearly ten million suspicious domains. As expected, the percent of attacks using the .com domain decreased, and we saw an increase in the use of other domains. Only 29% used legacy .com domain names, representing a 25% drop from the 56% seen in 2019 . gTLDs, such as .info, .ru, and .link, have emerged as popular domains for typosquatting attacks, and the frequency of new gTLDs being used is only increasing.

Leveraging AI and Automation to Stop Typosquatting Attacks

MX Records: Weaponizing a Typosquatting Domain for Business Email Compromise Attacks

In addition, new sophisticated TLD-based typosquatting email scams targeting partners and employees emerged. In this case, the bad guys send phishing emails from their bogus primary domains. These domains consist of the A (Address) records (determines which IP address belongs to a domain name) and MX (Mail eXchange) records (for sending email from). The MX record contains the host name of the computer(s) that handle the emails for a domain and a prioritization code, such as Mail.business.com.

MX record essentially weaponizes the domain for another type of attack called business email compromise (BEC). The criminal creates a malicious domain that mimics a legitimate email source, which allows them to bypass many of today's advanced email security products. The MX record gives the appearance of authenticity, so executives are more likely to open them and be lured into phishing targets where they click on a link, visiting a malicious site. This method is very common for BEC attacks. In some cases, mail reaches company executives during business hours and features urgent requests asking, for instance, to pay an outstanding invoice ASAP. Given that individuals are busy, they often take only a cursory look at such messages and fall victim to the attack.

Example of an A-record:

DOMAIN: business.com
HOST NAME: mail
IP-ADDRESS: 11.22.33.222

Example of an MX-record:

DOMAIN: business.com
MAIL EXCHANGER: mail.business.com
PRIORITY: 10

Emails are routed through to the IP address that is set in the A-record of the host.

The practice is quite lucrative. In a ransomware scheme reported by Brian Krebs, typosquatters used .cm instead of .com to setup their email and web servers. The scammers targeted iTunes and AOL users claiming to be fake security alerts, then netted 12 million hits at their sites in three months in early 2018.

Humans CCost to RegisterPossible

Why can't enterprises stop the practice? The primary challenge of fighting fraud on typosquatting domains is one of scale. As an example, a six-letter domain has 12,000 different variants on such TLDs. It is impossible for a brand to defensively register all of these domains, nor can these be monitored manually. And the problem is growing.

1,600
TLDS

Today, there are around 1,600 TLDs, which is five times more than what they were 10 years ago.

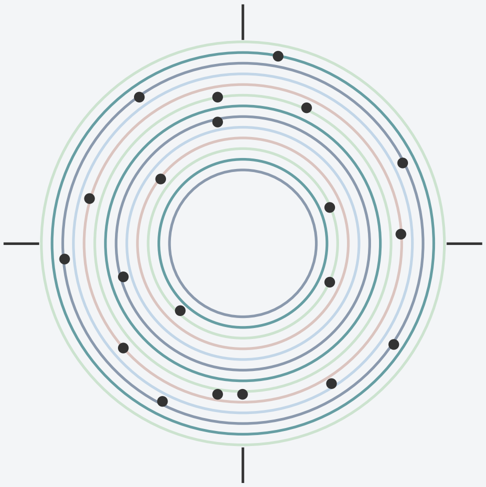
One strategy to fight typosquatting is to preempt criminals by purchasing all of the possible domains. The average cost of annual registration is between $10 and $20 per domain, according to the Whois lookup and domain ranking service. Using the low end ($10 per domain), purchasing every possible six-digit domain name is $120,00. For a 10-character name, the price rises to $250,000.

However, the reality is many names will not be available for $10. Typosquatters gobble up new domains as soon as they become available. The cost for their sites (when they are willing to sell them) averages about $1,000 per name, according to GoDaddy documentation. So acquiring 100 from a typosquatter represents a $100,000 annual investment at minimum, leaving thousands of other possible domains still available for cyber criminals. Preemptive registrations as a defense strategy is an economically unfeasible strategy.

Leveraging AI and Automation to Stop Typosquatting Attacks

Characters in URL	Possible Domains	Cost to Register
6______	12,000	$120,000
7_______	15,000	$150,000
8________	20,500	$205,000
9_________	22,000	$220,000
10__________	25,000	$250,000

In addition to the registration fees, there is also the administrative costs associated with tracking and keeping up with new TLDs. The cost estimates above could quickly spiral if the preemptive registrations in multiple TLDS. With these economics, a monitoring and takedown approach is the best strategy and much more cost effective that preemptive domain registrations.



Preemptive domain registration is an economically unfeasible defense strategy.

AI-based monitoring and automated takedowns costs only a fraction of what it would cost to preemptively register domains. It also provides a consistent, predictable cost, which simplifies planning and budgeting.

The Hands On Approach: Site Takedowns

Corporations can try to identify and take down bogus sites, but hurdles arise here also. Legacy fraud detection tools are not designed to address this issue because they sorely lack the accuracy needed to detect fraudulent or phishing sites. The inaccuracy prohibits these products from automatically reporting and taking down typosquatted sites. Since the volume of sites that need to be taken down can quickly get into the thousands, companies today face a daunting problem that is only expected to get worse.

The challenges evolved in a gradual manner, so the tools to address them grew in an autonomous, hodgepodge fashion. One solution identifies a fraudulent site, but much more work is needed to determine where it is housed and take it down. Consequently, enterprises find themselves with a handful of solutions that address one segment of the problem and need to cobble together and maintain all of the other pieces themselves.

Compounding the problem, current tools require a lot of manual processing. These products rely on search support, meaning that they do not automatically scan but rather provide search options that humans input. Scanning via manual searches, crawlers, and other legacy tool options eats up time, manpower, and money. With TLD numbers swelling, staff becomes overwhelmed and unable to monitor the scams effectively, proactively, and cost effectively.

Purchasing monitoring and takedown service may sound like an attractive option, but the reality is that most companies providing this service have manual processes. That is why most monitoring and takedown services require customer confirmation on the maliciousness of a site and approval to request a takedown. Because of the manual process, the most common pricing model is to charge per takedown.

Getting sites taken down adds more complications. Theoretically, a corporation can gather documentation and submit evidence that the site is fraudulent or abusive to the hosting company. Evidence usually entails screen shots or documentation that proves the site is being used for illegal activity. Just gathering this evidence is often laborious and difficult because these sites come and go so frequently.

After submitting the evidence and having no takedown occur, a company has to option to file a lawsuit against the transgressors, who may or may not be in a country where they do business. If the pages are housed abroad, that country's copyright laws may be inconsistent with the plaintiff's native country. At the end of this whole process, even if it is successful, criminals just create new pages on another server, and the corporation is back at square one, having to repeat the process all over again.

AI and Automation are the Solution

Technology advances created these problems, but they also offer a potential solution: automation at large scale. Corporations do not have the manpower to monitor malicious sites and thwart malicious sites, but artificial intelligence and machine learning solutions do. Yet not all AI systems are equal.

Bolster has developed artificial intelligence that delivers human intelligence at machine scale. It combines deep learning, computer vision, and natural language to understand the intent of a page rather than static criteria. This algorithm is highly accurate and is able to perform takedowns, removing the need for human intervention. The result is a highly accurate algorithm that operates at scale, putting the company on an equal footing against the criminals and removing their scale advantage.

Bolster Artifical Intelligence Technology

Deep Learning

Brings most accurate detection of brand hijacking based on deep analysis of image and text of the website

Natural Language Processing

Understand the intent of the website based on the natural language content.

Computer Vision

Fast image recognition algorithms that detect brand hijacking in just a few milliseconds

Threat Graph with 10 Billion Nodes

High quality proprietary threat intelligence collected over years adds important signals to the other AI models to amplify detection and accuracy

The Bolster algorithms are so sophisticated that they recognize minute differences between a legitimate Nike logo and a highly sophisticated illegitimate one. The system also recognizes how the site is handling sensitive information, like a username and password, and flag cases of potential fraud.

The Bolster platform also considers whether an MX record has been established, which is a clear indicator for whether a domain reserved by a typosquatter will be used for phishing campaigns. With an MX record, a typosquatter can become a trusted email source and use the fraudulent domain to phish and defraud a brand's employees, customers, and partners. The Bolster platform identifies these sites as being fraudulent and automatically submits takedown requests.

AI and machine learning, accuracy rates and scalability are vital features, especially when using automation to address the typosquatting problem. Criminals have had success because their technology is simple to deploy and works at scale. Until now, vendors have not been able to match them, and therefore corporations have been a few steps behind rather ahead of the bad guys. The reason? Artificial intelligence algorithm effectiveness varies dramatically. Bolster developed the industry's most accurate algorithm, one with a false positive rate of 1 in 100,000. Coupled with its leading-edge automation features, the solution has a takedown rate of over 99% within 24 hours – all without requiring manual intervention.

Zoom

Zoom has become an action, like Google, as well as a company. As its growth skyrocketed during the pandemic, online phishing and fraud increased exponentially.

The company wanted to get ahead of the threats and prevent the phishing and the defrauding of their user base. They worked with Bolster to deploy safeguards on a Sunday. Within the first 24 hours, the Bolster Detection Engine identified and took down 1,476 sites, a 99.3% takedown rate. The near-instant changes ensured that the vast majority of users were never exposed to Windows and Android malware, various phishing sites, or tech-support scams. In addition, Bolster added the malware binaries to VirusTotal and prevented bad actors from using similar malware elsewhere on the Internet. Underscoring the need for ongoing surveillance, 14,012 suspicious sites were identified and neutered in the first month of operation.

Zoom Brand Protection Results

1,476

SITE TAKEDOWNS
IN THE FIRST 24 HOURS

99.3%

TAKEDOWN RATE
IN THE FIRST 24 HOURS

14,012

SUSPICIOUS SITES
IDENTIFIED AND MONITORED
IN THE FIRST MONTH

Conclusion

Corporations are struggling to keep pace with the rapid rise seen in the number of TLDs. With ICANN's expansion of Internet domain names has come a rise in criminals using typosquatting to tarnish brands, large and small. Consequently, all companies need to proactively search for bogus web sites and take them down before they and their customers suffer.

However, until now, enterprises lacked both the financial as well as technical resources. The cost of purchasing the bogus sites was prohibitive. The sheer |volume of TLDs overwhelms them. Also, enterprises have not been able to move fast enough to identify and take down the fraudulent sites before they do damage. The Bolster platform addresses these vexing issues. It offers corporations an effective, automated, proactive way to combat this rapidly growing problem. Industry leaders rely on it to protect their brand. How will your company ward off the bad guys who clearly have your Website in their crosshairs?

For more information:

- Visit www.bolster.ai
- Sign up for a free trial
- Use our AI to scan URLs for free at checkphish.ai

BOLSTER

www.bolster.ai
4966 El Camino Real, Suite #101
Los Altos, CA, USA 94022
info@bolster.ai