Visual Signals: Critical Enhancement To Slam The Door On Scammers

Visual-AI In Anti-Phishing Whitepaper

Abstract

This document concerns itself with any form of phishing that relies on visual elements in the communication to gain the trust of the reader. This does not entirely rule out spear-phishing attacks, but the use of Visual-AI does best lend itself to attacks that leverage brands and those techniques where the content of emails, websites and documents are converted to graphics to fool phishing detection systems (PDS).

In this document, we use the term 'phishing' not just to describe techniques that try to gather personal and confidential information, but also to describe attempts to use email, websites and other techniques where the end result is the installation of malware that can gain access to sensitive systems/data, and ransomware that can cripple systems for the purpose of extortion.

The concept we propose is one that operates in near real-time and views content as a human would see it but at 'machine-speed'. Able to handle massive volumes quickly and accurately. Delivering a valuable scoring system to a PDS, which can then be used in their own filtering and scoring system.

Terms Used In This Document

APWG	Anti-Phishing Working Group
BEC	Business Email Compromise
BYOD	Bring Your Own Device
PDS	Phishing Detection Systems
TTC	Time To Confirmation
VAP	Very Attacked Person/People

Problem Statement

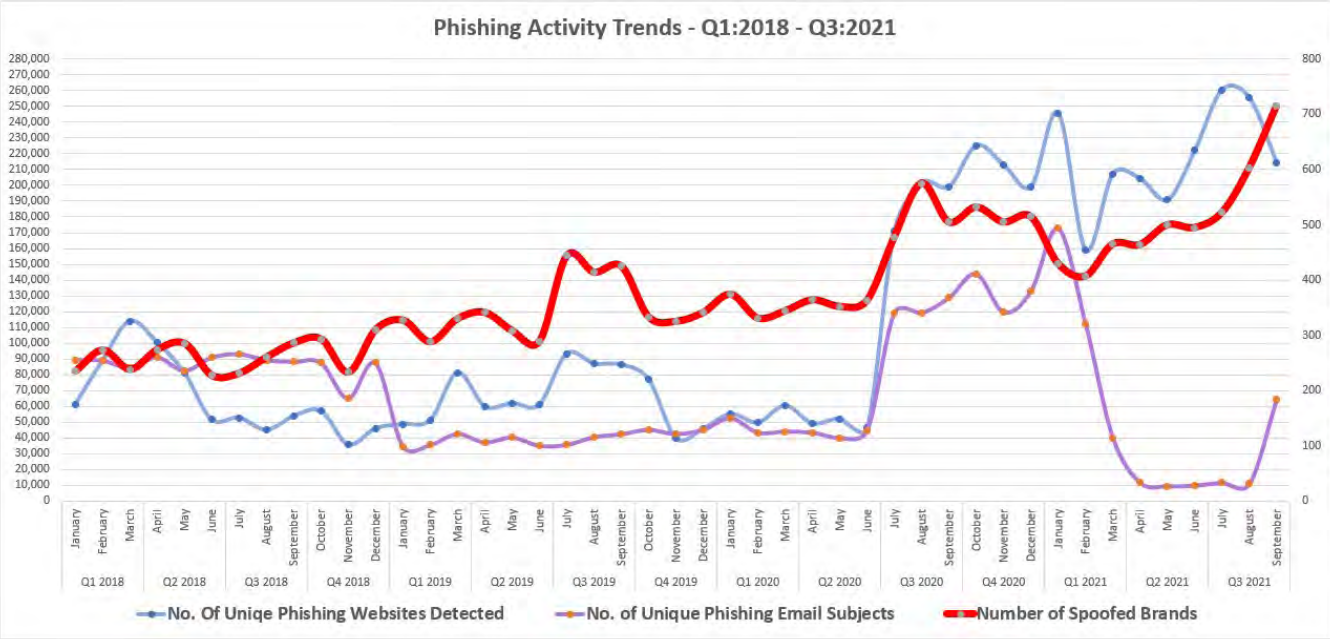
Phishing, in all its forms, has become of critical concern to so many companies and individuals, and no wonder given that cybercrime is predicted to cost the world $10.5 trillion annually by 2025, up from $3 trillion a decade ago and $6 trillion in 2021. This has fuelled massive growth in the cybersecurity sector. In fact, global spending on cybersecurity products and services will grow to $1.75 trillion cumulatively for the five-year period from 2021 to 2025. All according to Cybersecurity Ventures.

One of the fastest growing trends is phishing, which has seen a massive rise since the beginning of the COVID era. APWG saw 260,642 phishing attacks in July 2021, which was the highest monthly in APWG's reporting history. And the overall number of phishing attacks has doubled from early 2020.

If we look at data collected by APWG from January 2018 to September 2021, we see a very interesting trend:

	Q1 2018			Q2 2018			Q3 2018			Q4 2018		
	January	February	March	April	May	June	July	August	September	October	November	December
No. Of Unique Phishing Websites Detected	60,887	88,754	113,897	100,382	81,257	51,401	52,613	44,855	53,546	56,815	35,719	45,794
No. of Unique Phishing Email Subjects	89,250	89,010	84,444	91,054	82,547	90,882	93,078	89,323	88,156	87,619	64,905	87,386
Number of Spoofed Brands	235	273	238	274	285	227	231	260	286	293	233	310
	Q1 2019			Q2 2019			Q3 2019			Q4 2019		
	January	February	March	April	May	June	July	August	September	October	November	December
No. Of Unique Phishing Websites Detected	48,663	50,983	81,122	59,756	61,820	60,889	93,194	86,908	86,276	76,804	39,580	45,771
No. of Unique Phishing Email Subjects	34,630	35,364	42,399	37,054	40,177	34,932	35,530	40,457	42,273	45,057	42,424	45,072
Number of Spoofed Brands	327	288	330	341	308	289	444	414	425	333	325	341
	Q1 2020			Q2 2020			Q3 2020			Q4 2020		
	January	February	March	April	May	June	July	August	September	October	November	December
No. Of Unique Phishing Websites Detected	54,926	49,560	60,286	48,951	52,007	46,036	171,040	201,591	199,133	225,304	212,878	199,120
No. of Unique Phishing Email Subjects	52,407	43,270	44,008	43,282	39,908	44,497	119,181	119,180	128,926	143,950	119,700	133,038
Number of Spoofed Brands	374	331	344	364	352	363	478	575	505	532	505	515
	Q1 2021			Q2 2021			Q3 2021					
	January	February	March	April	May	June	July	August	September			
No. Of Unique Phishing Websites Detected	245,771	158,898	207,208	204,050	190,762	222,127	260,642	255,385	214,345			
No. of Unique Phishing Email Subjects	172,793	112,369	39,918	11,400	9,239	9,669	11,384	10,716	64,233			
Number of Spoofed Brands	430	407	465	464	500	495	522	603	715			

The chart below highlights this more visibly, showing that the number of spoofed brands has seen consistent growth across this period, reaching its peak in September, 2021. This trend is expected to continue and highlights the efficacy of spoofing as a technique used by bad actors. Similarly, although the use of phishing websites has seen some fluctuations across this period, it too reached a peak in Q3 2021.



Phishing technologies and methods underlying these trends continue to improve. According to Ironscales, it has allowed attackers to spoof the world's top 200 brands to create 50,000 fake login pages. Nearly 5% (2,500) of the 50,000 fake login pages were polymorphic, with one brand spinning out more than 300 permutations.

Another worrying trend is the growth in targeting mobile endpoints, with Lookout highlighting that mobile phishing threats in the energy sector (a key target for bad actors, along with other infrastructure, utility and healthcare organisations) surged 161% in 2021.

PDS are tasked with identifying and blocking phishing attacks but are faced with numerous challenges:

It's An Arms Race

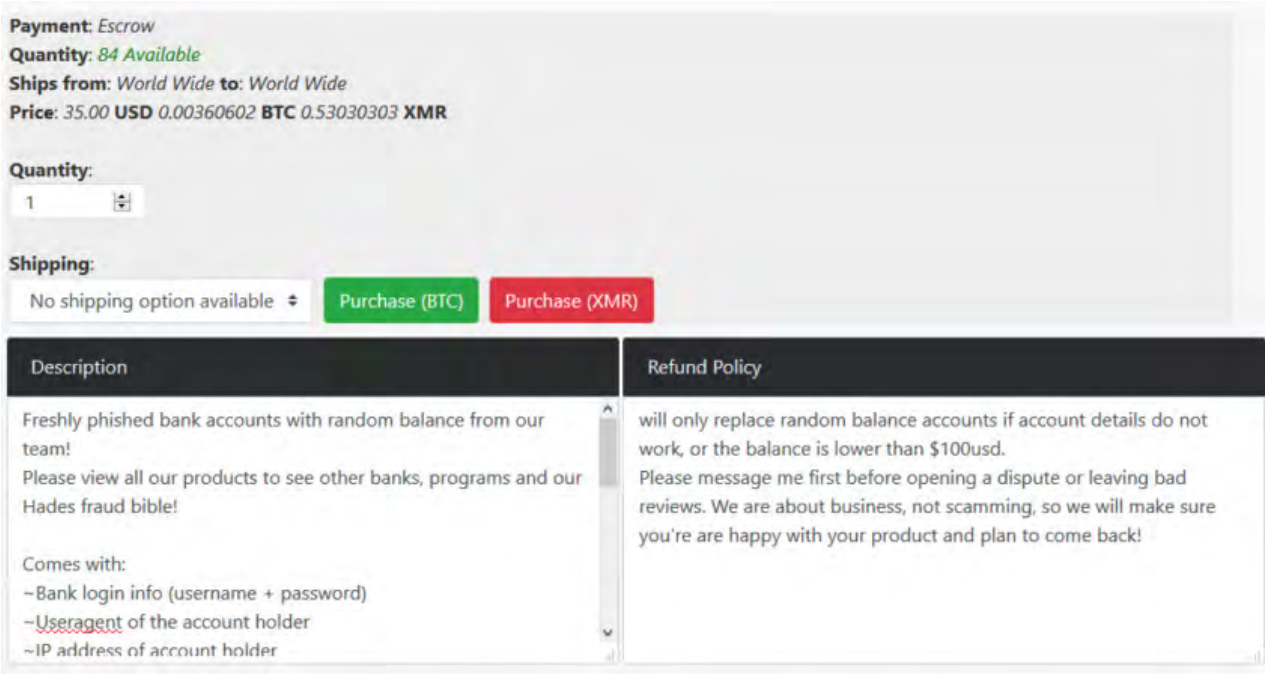
Anti-phishing companies are in an arms race with bad actors. From their ever more ingenious obfuscation techniques to avoid detection, to their targeting of more channels (voice, social, malware, web ads and text messages), and their more recent shift to targeting communications endpoints, especially with the shift to home-working where defenses are weakest. The list of checks and balances for PDS (Phishing Defence Systems), grows ever longer and this means the TTC (Time To Confirmation) also grows longer.

Readily Available, Weaponised Technologies

As ingenious as many of the phishing techniques are, being a successful bad actor needs little more than a good working knowledge of IT systems and access to the dark web. Some of the darker corners of the dark web are reported to be truly horrifying, but the more common areas of the dark web are a haven for nefarious commerce. Hacking and phishing 'kits' can be purchased from as little as $2, and whole cybercrime solutions can be bought at what most would consider 'very reasonable prices'. The suppliers of these solutions spare no effort in ensuring their clients are happy, providing spoofed pages that imitate real companies and even full guides on how to launch an email phishing scam.

Meanwhile, access to whole dark-server based services, such as spam email servers, make it very easy to get up and running. These are advanced servers that make it hard for junk filters to identify that it's a phishing email. In addition, the "From" address in the emails may look legitimate and use a valid domain like @gmail.com.

Remember also that the dark web isn't just a place to buy. Bad actors can also make money selling the data they gather.



Take this buying page shown above for example (as researched by Privacy Affairs), where you can purchase a batch of 'freshly phished bank accounts with random balance' of at least $100. They even offer a satisfaction guarantee, with the truly ironic statement, 'We are about business, not scamming...'

This means that bad actors can move very quickly to adapt to new detection and blocking methodologies.

The Odds Are Stacked Against You

In this arms race, PDS providers can understandably feel like they can't win. After all, their customers need them to detect and stop every phishing attempt, ensuring that not one single breach is allowed because just one breach can create a million euro/dollar issue. The bad actors, however, have a much lower bar, where just one breach is a major success.

Brute Force Method

A sure-fire method of stopping phishing attacks where the bad actor is using links to a malicious site or file, is to simply block any communication from unknown sources that contain a file attachment or a link to an external site. Don't worry about checking and testing the content or links, just bounce the email and let genuine senders reach out to try and fix the issue with their intended recipient.

As effective as this is, it stops a lot of genuine communications and causes issues with systems that send everything from verification links to invoices for services. The recipient doesn't receive the email and the sender (being automated) never sees the bounce! Clearly, a better solution is required.

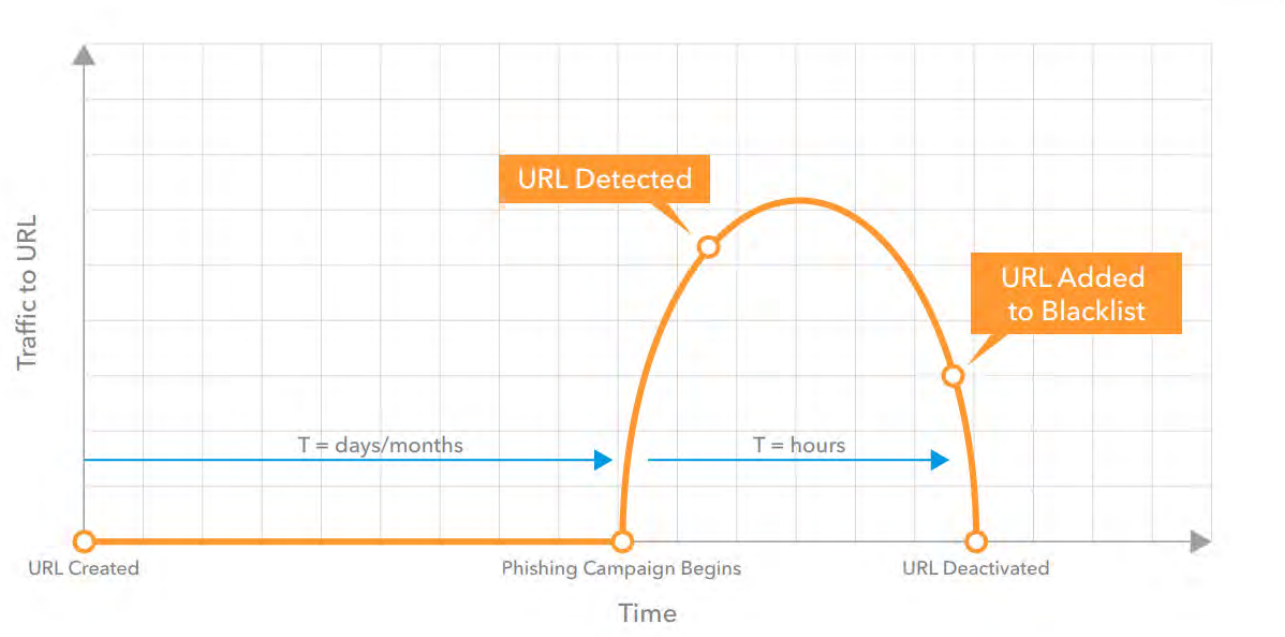
Challenges

The PDS industry faces numerous challenges in delivering a system that can be relied on by their clients:

Time & Speed

Any PDS must operate as close to real-time as possible, with absolute minimal delay, so as to be transparent to the user. Additionally, bad actors and scammers are consummate opportunists and they know that it takes time for relevant data about domains, senders, URLs and other sources/flags to be gathered, scored, and added to relevant blacklists.

So they switch and change often, with sites existing, in some cases, for only hours (as highlighted in a whitepaper by Akamai).



Above we see a quite typical phishing campaign profile. By the time the URL is detected, analysed and added to the blacklist, the campaign is finished and the site is moved to a new URL.

Any system, therefore, needs to be able to deliver results within this framework and not rely on blacklisted factors as the singular element of its scoring system.

Complex Systems Take Time To Patch

Cyber-security systems have traditionally relied on patches for things like malware signatures and blacklists, but IT professionals have a natural reluctance to patch, and especially to patch too often. Ignoring the planning and logistics of implementing patches (often requiring taking servers offline), IT staff hate 'rocking the boat'. Patches can cause major outages and introduce other unforeseen issues, which then cause stress and anxiety to resolve. This introduces delays in closing gaps in the defence cordon, giving bad actors the time they need to get a win.

Systems need to be able to flag possible threats without the need for constant patching. It should offer its detection functionality outside of the patching based methodology and offer critical value-add to patching-based defence systems.

Massive Volumes

As shown at the beginning of this document, bad actors can generate 50,000 fake login pages for the world's top 200 brands with relative ease. Also, highlighted by the APWG's (Anti-Phishing Working Group) Phishing Activity Trends Report for Q3 2021 Phishing attacks are at their highest level in recorded history. The issue, therefore, that PDS providers have is quickly and efficiently sifting through all communications and links to validate genuine communications and sites while blocking bad ones.

Any system that can help filter out the noise and focus on the highest-risk threats will help to deliver timely detection and blocking in a transparent way.

Multiple Attack Vectors, Channels And Devices/End-Points

Phishing used to be limited only to email on PC. Today every device and every channel can now be a conduit for phishing. From mobile malware delivered through SMS (SMiShing), the gathering of data through voice-calls (Vishing) and even through malicious apps or ads served in apps, websites and social media.

Social engineering is one of the latest techniques used. Employing gamification and the power of sharing, they operate through social media sites and chat apps to snare victims, who then share a fake link (which is trusted by other users because it came from a trusted source). Each person then inadvertently installs malware or gives away perhaps minor details that provide the next link the scammers need for their BEC (Business Email Compromise) campaign.

The era of BYOD (Bring Your Own Device) and home-working also brings challenges, where devices or personal email accounts may not be protected, yet are operating inside the business perimeter.

The ideal solution must therefore sit behind and between all the devices and channels to check sites and files when accessed inside the perimeter, and do all this in real-time.

Bad Actors Employ Legitimate Technologies

As cloud services become more popular and new techniques are employed by users and made standard by social sites (like Twitter's shortening of all URLs to t.co versions), bad actors are exploiting the trust that users have in these services, making it more likely that they'll click on links to files hosted on Google Docs, Microsoft OneDrive or URLs that have been shortened.

This is another way that bad actors introduce noise in their attacks, so detection systems have more checks to do.

Systems must therefore look at the content beyond the link to identify its legitimacy.

Staff Training

Of course, companies spend a great deal of time, effort and money providing training to staff in an effort to forewarn them about the typical attack vectors and educating them on what red flags to look out for. However, research shows that even highly computer-literate individuals were unable to successfully identify fake websites, even when expecting to see them!

The best-produced fake site was able to fool more than 90% of participants, while the average was just under 40% of participants thinking that a fake site was genuine. This means that of every 10 employees in a company, 4 of them would be likely to take an action on a phishing website, compromising themselves and the company!

The research highlights how 23% of participants looked only at the content of the website and did not use any other visual cues and it showed that the same percentage of participants did not have a good understanding of the padlock icon, with some thinking that it was more trustworthy when shown within the content area than in the browser header.

Further, the research showed that "indicators that are designed to signal trustworthiness were not understood (or even noticed) by many participants." This highlights that using techniques to flag dangerous or high-risk content, may simply go unnoticed or even ignored.

So, ultimately, we see that trying to rely on the user as the defence barrier, especially when they are inside the castle gates, is not the best strategy. The goal must be to ensure that they never see a phishing email or website in the first place and provide training as a nice-to-do action.

Any solution that can reduce an organisation's reliance on using their internal staff as amateur cyber-security blockers will be a very attractive proposal.

The Visual-AI Solution

Visual-AI's strength is its ability to see the world as humans see it, but at machine speed. Never tiring, never making mistakes and to a much higher accuracy than humans can achieve. Visual-AI, therefore, enables a new paradigm in combating phishing attacks where the attackers rely on:

1. Building trust through the use of familiar and authoritative visual cues
2. Evading detection by using graphical elements to replace machine-readable elements
3. Evading detection by adding noise to, or eliminating/obfuscating, code

The use of Visual-AI is effective for both email-based attacks and also for phishing websites and documents that may contain phishing content.

Examples of Key Visual Elements Exploited By Bad Actors

Company Logo

Used in email, documents and websites, the logo of a bank, service or system can inspire confidence. Identifying the highest risk/most phished brands in attacks quickly allows the PDS to take priority action on that message or site.

It is important to note that brands can have multiple versions of their logo and bad actors may often use older versions in their attacks:

According to Vade Security, the top 25 impersonated brands are as follows:

Most Impersonated Brands - Top 10

#	Brand	Unique Phishing URLs	QoQ Growth
1	PayPal	11,392	-31.2%
2	↑2 Facebook	9,795	-18.7%
3	↓1 Microsoft	8,565	-38.2%
4	↓1 Netflix	6,758	-50.2%
5	↑63 WhatsApp	5,020	13,467.6%
6	↓1 Bank of America	4,375	-21.5%
7	↑1 CIBC	2,414	11.2%
8	↑4 Desjardins	2,243	54.4%
9	↓3 Apple	2,126	-57.9%
10	↓1 Amazon	2,110	0.6%

Added: WhatsApp, Instagram, Square, Comcast, M&T Bank
Dropped: Yahoo!, Wells Fargo, SunTrust Bank, AT&T, Societe Generale

Vade Secure
Predictive Email Defense

Most Impersonated Brands - 11 to 25

#	Brand	Unique Phishing URLs	QoQ Growth
11	↓4 Chase	2,012	-14.6%
12	↑3 BNP Paribas	1,512	23.1%
13	↑16 Instagram	1,401	187.1%
14	↑19 Square	1,315	246.1%
15	↑1 Dropbox	1,233	0.7%
16	↑1 ATB Financial	1,229	0.7%
17	↓7 DHL	1,161	-31.1%
18	↑8 Comcast	1,012	47.1%
19	Orange	992	6.4%
20	↑5 Adobe	872	11.8%
21	↑1 Impots	867	4.2%
22	↑25 M&T Bank	849	469.8%
23	↓10 Docusign	837	-40.3%
24	↓4 Google	795	-12.9%
25	↓7 Credit Agricole	710	-30.0%

Added: WhatsApp, Instagram, Square, Comcast, M&T Bank
Dropped: Yahoo!, Wells Fargo, SunTrust Bank, AT&T, Societe Generale

Vade Secure
Predictive Email Defense

Marks

Similar to logos, Scammers may use recognised and authoritative marks to increase trust. Safety marks, security cert marks, and padlock icons specifically in website content, etc. can be combined to ratchet up trust factors in recipients. The presence of these marks, in conjunction with other elements, can be a key indicator of a phishing attempt.

Favicon

The mini logo-based icon used in the tab of a website is another often-used technique to try and confuse victims. It can be another effective ranking factor.

QR Codes

A most recent trend is the introduction of QR codes, specifically in mobile communications. Systems can struggle to both identify them and interpret their data, which when decrypted will expose a link that takes the user to a malicious site.

QR Code embedded into email to mobile user

Context

The role of context in ranking possible attacks is also very important. A logo found on a page or email is ambiguous without taking into account context.

High Threat

Low Threat

High Threat

Low Threat

These examples above highlight how context plays a vital role, above simply detecting the presence of a brand, in ranking an email or page as high risk.

Examples of Evasion Techniques Involving Graphical Elements

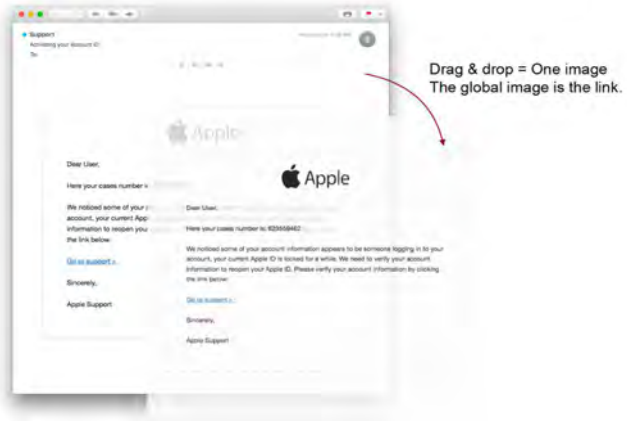
Visual analysis of embedded graphical content can also help to detect bad actors who are trying to evade detection. Typical examples are:

Text Converted To Graphics

They will convert key ‘trigger’ words in the content to a graphic. For instance, words like ‘Username’, ‘Password’, ‘Login’, and ‘Credit Card Number’ will be converted from readable text to a JPG or PNG, but in such a way to be indistinguishable from the normal text to the user.

Sections Converted To Images

Rather than converting just a word at a time, they may convert an entire form to a graphic, overlaying the input fields above the graphic. In some cases, as outlined in the image to the right by Vade Secure, they will convert the entire email or site into a single graphic.



URLs Converted To Graphics

Similar to key ‘trigger’ words, a genuine URL is converted to a graphic, however, a link is then attached to the graphic that points to the fake site. In this way a user may see www.paypal.com, but the link behind the image will point to www.paypa1.com.

6 Common Detection-Evasion Techniques

In developing a resource like this, one can't help but be impressed by the ingenuity of these bad actors in developing new and innovative ways to hide from, and fool, PDS platforms. There are a multitude of techniques they use, but for the purposes of this whitepaper we have focused on the most common techniques that can be overcome by Visual-AI:

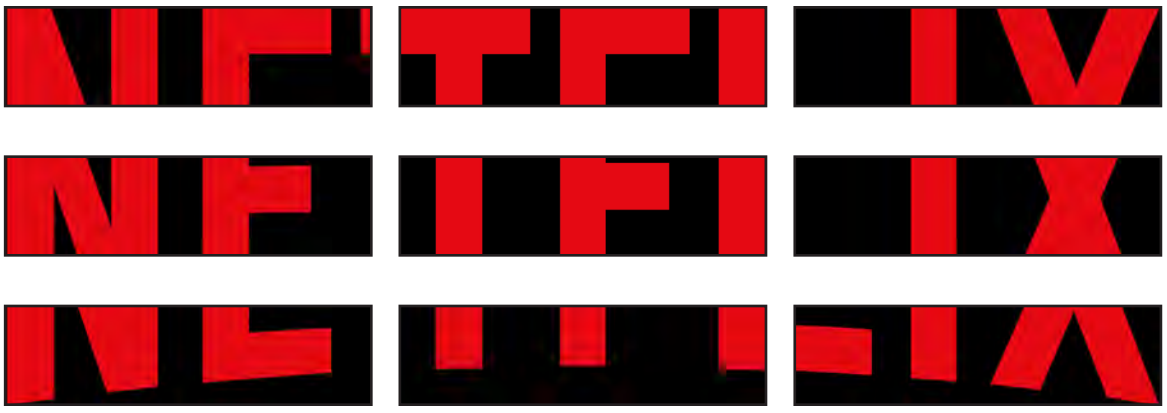
1. Keep The Lifespan Short

As outlined in a previous section, many systems rely on blacklists to identify threats. Bad actors will therefore run a site for a very short period before they change servers, IPs, domains etc., thereby avoiding being flagged. 2020 has also seen the rise of single-use URLs that therefore only last seconds!

2. Add Noise

A great way to evade is to confuse. They achieve this by adding significant 'noise' to the code of an email or site. They will change key attributes of graphics, like filenames and HTML attributes and metadata. They can even break up graphics into multiple small parts, as shown below

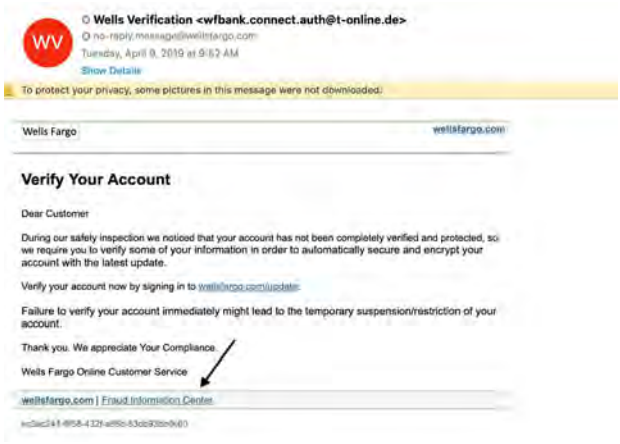
.



Using HTML or javascript these separate parts can be displayed seamlessly to a visitor in their web browser.

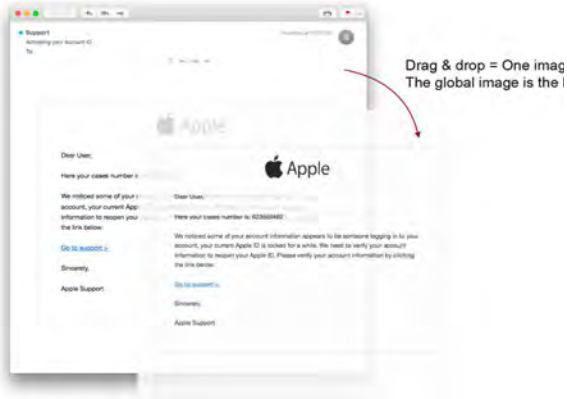
3. Legitimise The Illegitimate

As well as being a fantastic method for tricking victims, bad actors will often use legitimate links in emails and websites, such as help, legal and even anti-fraud pages. They will also use a legitimate reply-to address, all of which helps to confuse the PDS.



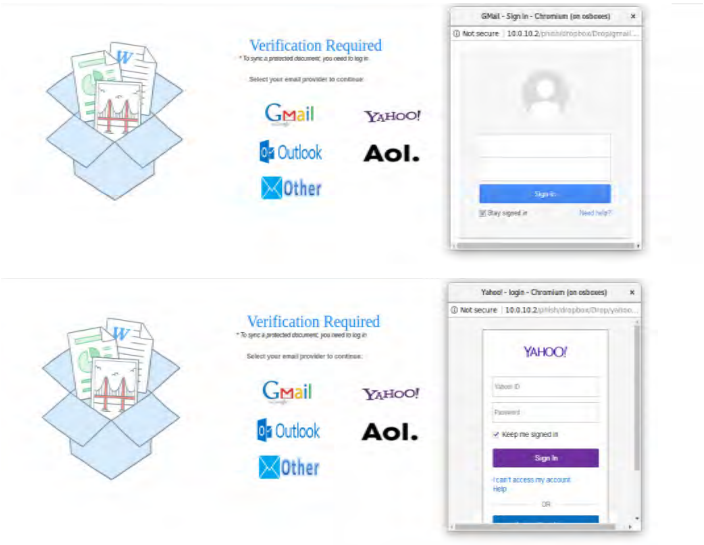
4. Nothing To See Here

As shown in the Apple example previously, bad actors will take the most sensitive section/s of content that can trigger a detection, or even the entire email/website and convert them into graphics. This greatly reduces or eliminates machine-readable content/code, giving the PDS nothing to work on.



5. Dynamic Content

They will use (often obfuscated) Javascript to dynamically generate the contents of a page. This stops HTML parsers from being able to extract elements, such as form fields, for analysis because the page is only displayed when rendered by a web browser. This issue has been made worse by the introduction of WebAssembly, which provides many advantages for developers, but has been massively adopted by bad actors who are delighted with the fact that they can compile code, which makes it even easier to obfuscate.



6. Open The Windows

Bad actors can drive victims to genuine sites, but trigger an additional window, or a popup, to open above the legitimate login form in order to substitute their bad form, as shown in the examples shown here.

How Visual-AI Detects The ‘Undetectable’

Visual-AI sees all content with human eyes. It is something that the bad actors cannot avoid because their goal is to show humans an email, site or document that looks as natural and genuine as possible. By fully rendering the content (perhaps in a sandbox) and converting the output to an image, the PDS will have captured what a human will see. This can then be passed to VISUA’s Visual-AI engine for analysis.

Logo and Marks detection can be employed to identify brands and other important markings. Additionally, Text Detection can be used to convert all text into machine-readable content and find ‘trigger’ words that indicate risk.

All this data can be compiled, scored and returned to the PDS system for addition to its own gathered data for an aggregate risk score. Content from the highest risk brands, or with the highest risk trigger words can be prioritised for further analysis.

1 Render & Capture

Render the web page/email and save it as a flat image for processing.

2 Process The Image

Identify high-risk brands (for priority processing) and any anomalous attributes within the page/email.

3 Visual Risk Scoring

Calculate a risk score and pass it, with the identified anomalies, back to the master phishing detection system for final actions.

Conclusion

This document has highlighted the challenges faced by cyber-security companies and PDS providers in stopping phishing attacks. It shows the extreme and ingenious lengths bad actors go to in order to avoid detection and achieve their goals. And further outlined the role that Visual-AI has to play in enhancing detection and blocking of phishing attacks.

Humans To The Rescue?

We know that trained humans have the capabilities to detect phishing attacks to a very high degree of accuracy. The problem is the volumes they can process are very low. An army of humans would therefore be required to keep up with the volume of checks required. The work would also be tedious and tiring. As such it would be both economically and physically unsustainable. Humans are also not great at connecting the dots in massive volumes of data.

Ultimately, working at human speed is like **Knowing A Gang Of Intruders Are Inside The Castle Gates And Closing Them Too Late.**

Artificial Intelligence To The Rescue?

AI addresses the issue of speed because it operates at 'machine speed'; able to process millions of pieces of data in a fraction of the time it would take a human to do the same work. AI can also be trained to quickly and efficiently compare multiple strands of data and data-points to identify correlations that would otherwise be missed. The weakness of AI is that bad actors can relatively easily evade the employed data gathering techniques, which non-visual AI systems rely on, to produce their results.

This means that employing AI is like **Knowing A Gang Of Intruders Are Opening The Castle Gates And Closing Them Just In Time, Most Of The Time (But Sometimes One Or Two Get In).**

Visual-AI + Standard AI Is The Answer!

By supplementing the current AI systems with advanced and targeted Visual-AI, you achieve a perfect combination that delivers 'human' visual analysis (but at machine speed) with AI data analysis. This combination is far more difficult for bad actors to evade and can be effective at near real-time (one second or less), allowing all content and traffic requests to be processed in real-time.

If this is combined with intelligent protection, i.e. priority analysis for VAP (Very Attacked People) in an organisation, a very high level of protection can be achieved without disruption to workflow.

This is the equivalent of **Knowing Intruders Are Approaching Your Castle Gates And Locking Them Out Before They Even Get Close!**

About VISUA

VISUA develops best-in-class, enterprise Visual-AI that powers the world's leading brand protection, authentication and monitoring platforms. VISUA delivers solutions ranging from logo/mark detection and counterfeit product detection to holographic authentication and ad detection. Its Visual-AI technology is proven to deliver the highest precision with instant learning, at unlimited scale, and is adaptable for any use case. VISUA believes in People-First AI, they see a world where Visual-AI will lift humanity out of the mundane, empowering a society that focuses more on creativity and collaboration and less on binary tasks, and empowering services and solutions that humans alone simply can't deliver.

VISUA

Harness the power, speed and accuracy of Visual-AI in your Phishing Detection Platform with the most precise and scalable solution on the market

Contact VISUA to learn more about how
our tech combined with our flexible,
hands-on approach can benefit your business.

sales@visua.com

Or find out more at VISUA.COM