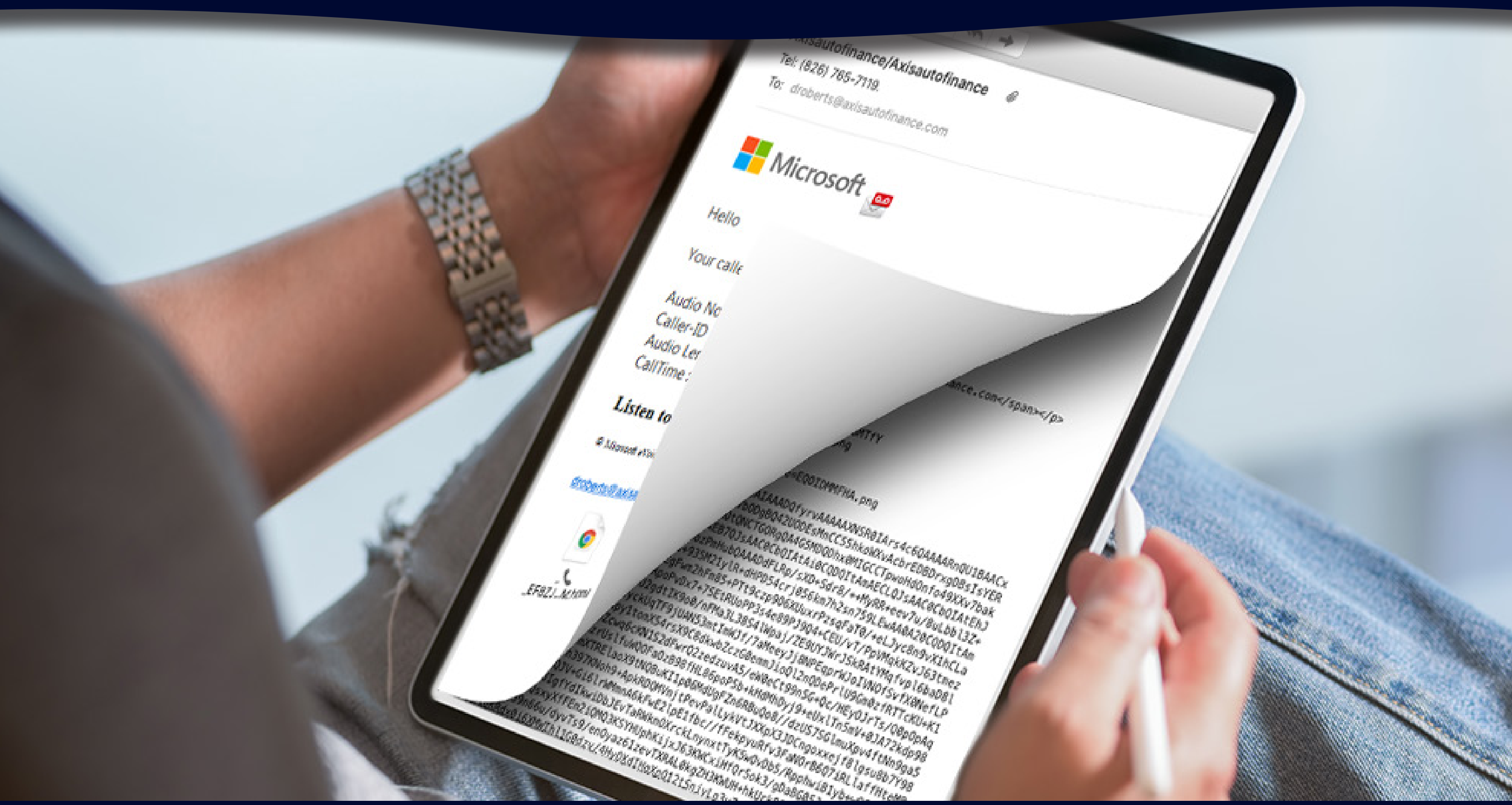
REPORT

Understanding Phishing: Computer Vision

Computer vision is a key part of INKY's ability to determine whether an email is safe or not. But few people understand how a computer can "see" the way a human does. It's not enough to reproduce a picture of the object in question. Computer vision involves deriving meaning from that image. This guide explains how INKY uses visual analysis to figure out important features of each incoming email and matches it with other information to ward off phishing attacks.

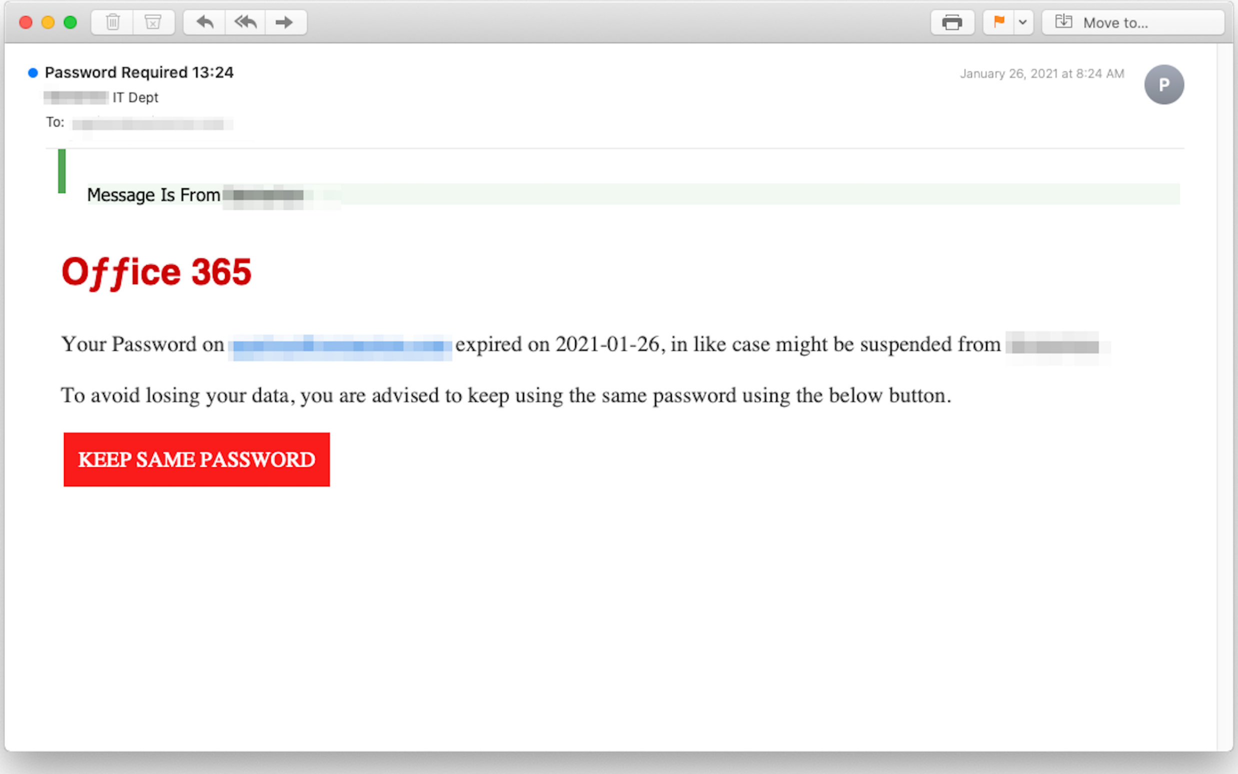


INKY Uses Computer Vision to Thwart Phishing Attacks

INKY, an in-line software module that monitors traffic from a secure email gateway (SEG) before it gets to the recipient, is super good at stopping phishing attacks. An important feature of INKY, one that sets it apart the competition, is that it analyzes an email (in less than 2 seconds!) both as a machine would and as a human would. In order to do the human part of the analysis, INKY needs to “see” the email. This “seeing” is done with computer vision.

Text is Actually an Image

A simple example would be a homographic (homo = like; graph = character) attack, one in which the perpetrator substitutes characters for a given domain with others that look the same or similar. With the broad acceptance of Unicode — the encoding system that gives every character in every language, living and ancient, a unique number — such attacks are easy. For example, the Japanese character set has in it a complete Latin alphabet, called, in Japanese, Romaji (like Roman). These letters look just like Latin letters, but the numbers underneath (the codepoints) that call them are entirely different. You can think of these fonts as numbered sets of pictures of letters. A glyph is just lit-up pixels in a pattern that makes a picture of a letter.



Password Required 13:24

[redacted] IT Dept

To: [redacted]

January 26, 2021 at 8:24 AM

Message Is From [redacted]

Office 365

Your Password on [redacted] expired on 2021-01-26, in like case might be suspended from [redacted]

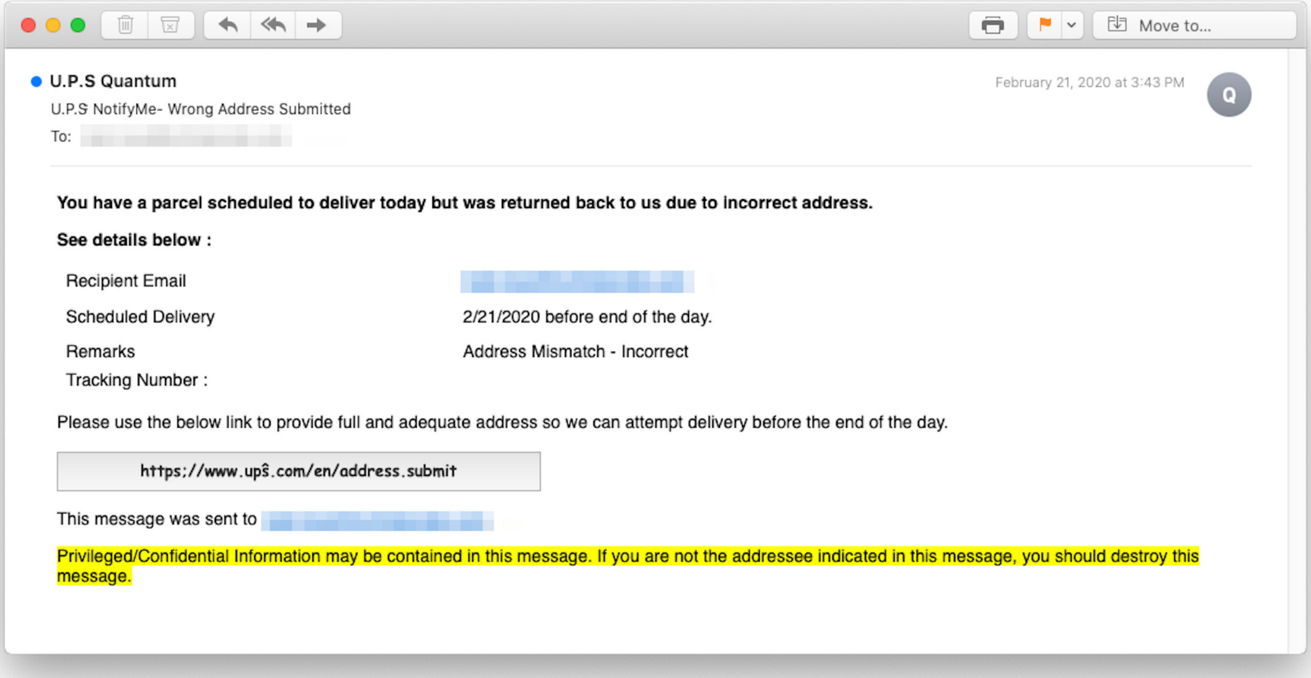
To avoid losing your data, you are advised to keep using the same password using the below button.

KEEP SAME PASSWORD

The Office 365 logo in this example is actually just big red text characters in a strange font.

An email from what looks like the domain Yahoo.com can come from somewhere completely else. The human looking at the sender information in the email header would see “Yahoo.com,” and generally think everything was fine. A machine analyzing the codepoints (numbers) underneath would check to see whether the domain was real and was sending from a legitimate range of IP addresses, but would not stop to ask, “Is this the right domain?”

INKY uses visual analysis (the way a human would) to determine that the email purports to come from Yahoo and then checks (the way a machine would) to see whether the underlying codepoints represent a domain under Yahoo’s control. If the two don’t match, INKY throws a flag.



U.P.S Quantum
U.P.S NotifyMe- Wrong Address Submitted
To: [redacted]

February 21, 2020 at 3:43 PM

You have a parcel scheduled to deliver today but was returned back to us due to incorrect address.

See details below :

Recipient Email	[redacted]
Scheduled Delivery	2/21/2020 before end of the day.
Remarks	Address Mismatch - Incorrect
Tracking Number :	

Please use the below link to provide full and adequate address so we can attempt delivery before the end of the day.

https://www.upŝ.com/en/address.submit

This message was sent to [redacted]

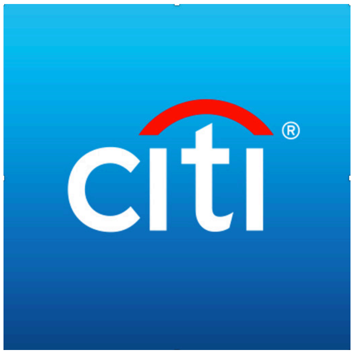
Privileged/Confidential Information may be contained in this message. If you are not the addressee indicated in this message, you should destroy this message.

In this example, the link does not lead to UPS (www.ups.com), the shipping company, but to www.upŝ.com, a domain controlled by phishers.

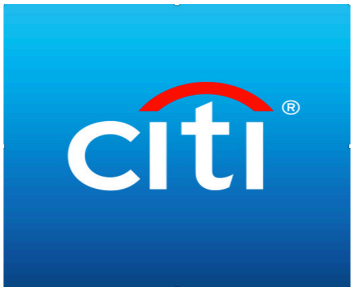
Images Can Be Deceptive

Visual analysis can get more complicated, of course. Imagine that part of fooling people into believing that an email is legit involves the usage of actual imagery. The bad guys know that text strings can often be checked against known threats and are aware that imagery is harder for search functions to “see” and understand. Thus, a Citibank logo in a message might slip past a machine and hoodwink a human. And the bad guys know that primitive image-matching software might detect just an ordinary logo. So, they introduce distortions designed to deceive such programs.

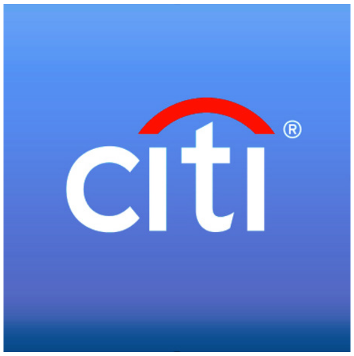
For example, this is the actual Citibank logo:



This image shows a version that is distorted dimensionally:



And here’s an example of the logo shown with altered colors:



Unsophisticated visual analysis programs might catch the first one, but INKY’s computer vision software is flexible enough to snag the other two as well. Even with elongated imagery or substituted colors, INKY still concludes: “This image is trying to be a Citibank logo,” and tells the machine-analysis module to check Citibank's associated domains to see whether the email came from one of the bank’s legitimate email servers. When the IP address turns up a steel fabrication factory in Venezuela, INKY throws a flag.

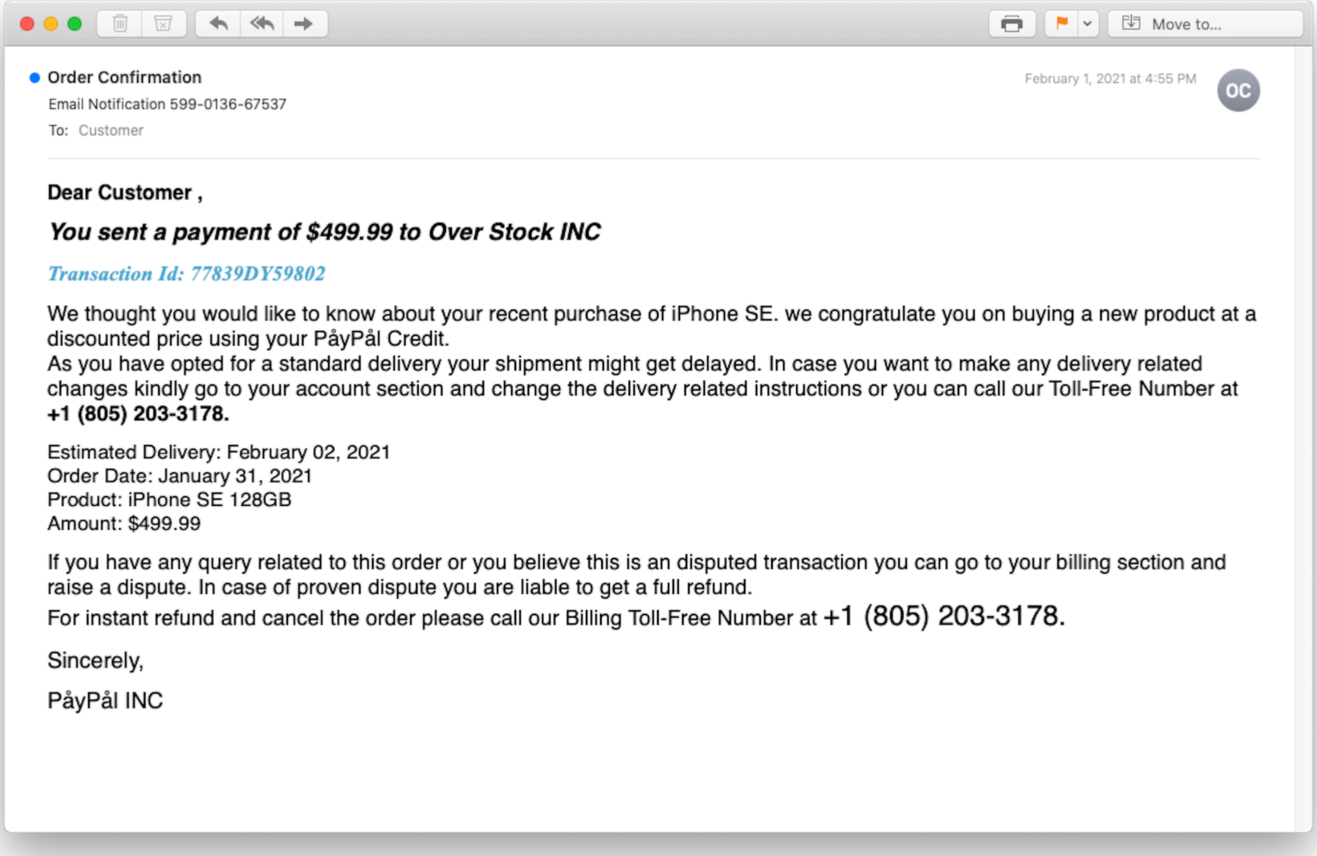
Modules Working Together Like a Brain

Computer vision is a key technique that Inky uses to build an overarching model for what the email is trying to be. Like a human, INKY is trying to “see” the email as the intended recipient would. Is this email telling me scary things about some aspect of my business or personal life? Is it saying it’s from a bank? Is it asking for something from me? Login information? Other personal details? Does it have reassuring logos from reputable brands or government bodies?

To winnow out fraudulent messages, INKY runs each email through a variety of models to assess images, colors, text, and other brand-indicative features. These models can be seen as basically a giant decision tree or an elaborate voting mechanism. Each model contributes what it thinks (depending on its mandate, e.g., Is this one supposed to be from a known brand? Yes, no, maybe) and gives a confidence level (e.g., I’m 60% sure this one is asserting that it comes from Citibank). When the overall assessment reaches some confidence threshold, the aggregator module says, I think this one is asserting that it comes from Citibank.

Effectively the models vote, much like the human brain does. In lay terms, we have our animal instincts: does that pitch seem too friendly; and our higher reasoning: I’ve never gotten something from these people before; perhaps this isn’t safe. And your fear module and your logic module give their best shot to the overall decision maker in you: perhaps I shouldn’t open that attachment.

Over time, with machine learning in a particular context (e.g., at a customer location), the model gets more accurate at drawing the correct conclusion.



Order Confirmation
Email Notification 599-0136-67537
To: Customer

February 1, 2021 at 4:55 PM

Dear Customer ,

You sent a payment of $499.99 to Over Stock INC

Transaction Id: 77839DY59802

We thought you would like to know about your recent purchase of iPhone SE. we congratulate you on buying a new product at a discounted price using your PåyPål Credit.
As you have opted for a standard delivery your shipment might get delayed. In case you want to make any delivery related changes kindly go to your account section and change the delivery related instructions or you can call our Toll-Free Number at **+1 (805) 203-3178.**

Estimated Delivery: February 02, 2021
Order Date: January 31, 2021
Product: iPhone SE 128GB
Amount: $499.99

If you have any query related to this order or you believe this is an disputed transaction you can go to your billing section and raise a dispute. In case of proven dispute you are liable to get a full refund.
For instant refund and cancel the order please call our Billing Toll-Free Number at +1 (805) 203-3178.

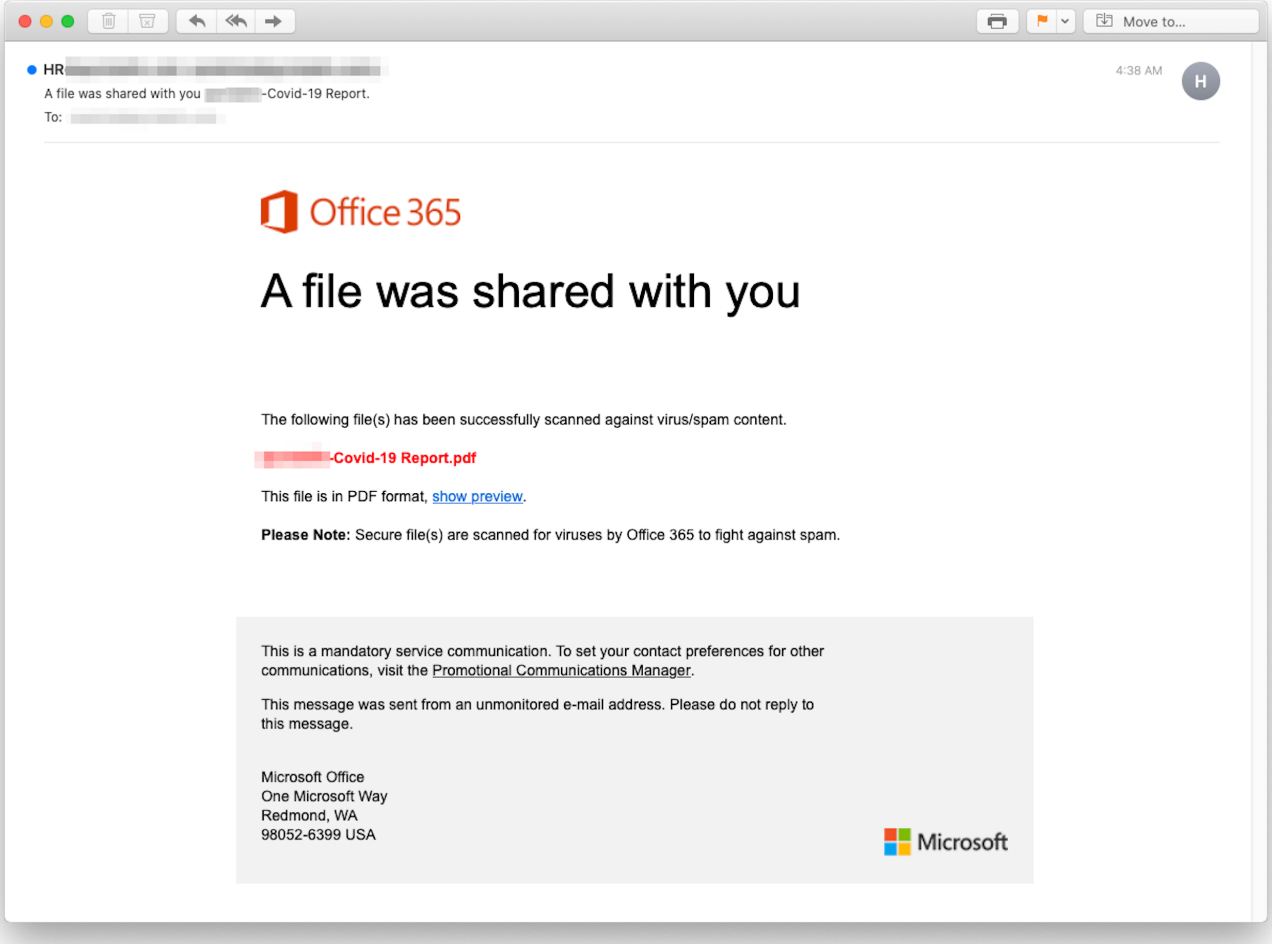
Sincerely,

PåyPål INC

A person seeing this phishing attempt might think it was sent from PayPal because we’re great at resolving near matches. A pattern-matching algorithm analyzing this email would not detect it as PayPal because Unicode text substitution has turned it into “PåyPål.” It may legitimately come from www.påypål.com, but that domain is outside the control of the actual PayPal.

DKIM, SPF, and DMARC are Not Enough

A human recipient has no way to verify that a plausible-sounding domain is legitimate as opposed to one that some attacker registered — for example, citibank-loan-customer.com. Now, that's a pretty plausible domain name, and it may pass the normal security checks that a SEG runs. There really is such a domain and its admin may even have turned on DomainKeys Identified Mail (DKIM) authentication (which acts like an encryption seal on all outbound mail streams) and set up Sender Policy



HR[redacted]

A file was shared with you [redacted]-Covid-19 Report.

To: [redacted]

Office 365

A file was shared with you

The following file(s) has been successfully scanned against virus/spam content.

[redacted]-Covid-19 Report.pdf

This file is in PDF format, show preview.

Please Note: Secure file(s) are scanned for viruses by Office 365 to fight against spam.

This is a mandatory service communication. To set your contact preferences for other communications, visit the Promotional Communications Manager.

This message was sent from an unmonitored e-mail address. Please do not reply to this message.

Microsoft Office
One Microsoft Way
Redmond, WA
98052-6399 USA

Microsoft

The visual information in this email seems to indicate that it comes from Microsoft.

Framework (SPF) records (which contain the exact server IP addresses and ranges that can legitimately send email from the domain). Such good citizens!

Thus, even if the target company has enabled Domain-based Message Authentication, Reporting & Conformance (DMARC, a method that can check either the SPF record or the DKIM signature on the incoming mail for a base level of legitimacy), it may let a bad one through because the domain is real, is identified correctly by a public key, and is sending from an address within the authorized range. If either SPF or DKIM passes, the email is usually delivered to the recipient.

Even security experts get confused about the value of SPF and DKIM, which prove cryptographically that an email came from a particular server in a domain and that said server is allowed to send from that particular address. However, these measures don't reveal who controls that domain.

In order for INKY's insight to work in the real world, the determination with respect to an email's legitimacy has to be done by correlating multiple forms of corroborating information. A Facebook logo in a message does not necessarily imply that the mail is trying to impersonate Facebook. It might just contain some text that says "Like us on Facebook." INKY "knows" that and doesn't "overreact" (weigh that information too heavily).

Although we, as human recipients, gloss over such details, our brains are doing an analysis quite similar to INKY's computer vision. We look at context, layout, features, and other subtle visual cues to make our own determination of legitimacy. What we can't do is then compare that to what the hidden parts of the email (i.e., the "header" information) tell INKY: where the email actually originated.

Generally, INKY wants to examine both imagery and text, assessing how images are placed, their size, colors, and other attributes, and how text blocks and images are positioned or associated with each other. These capabilities — built up over the years by a team of experts in artificial intelligence, machine learning, and computer vision and drawing on academic work in these areas — make INKY the most formidable anti-phishing technology on the market today.

Why INKY?

INKY provides the most comprehensive malware and email phishing protection available. To see INKY's anti-phishing solution in action, request a demo. Let us show you what a difference it can make.

INKY Phish Fence uses a proprietary blend of Machine Learning and Artificial Intelligence that blocks even the most sophisticated phishing attacks that get past other systems.

INKY Phish Fence uses proprietary technology and algorithms to "see" each email as the recipient would. Unlike a person, however, it can detect an email forgery and/or malicious or suspicious content. Once detected, it can redirect the email to a quarantine area or deliver it with disabled links and warnings.

Alerts show within the email itself, which allows it to be viewed on desktop or mobile. This is a significant difference from other systems, which display warnings in headers and may not render properly in mobile applications.

INKY Phish Fence sits on top of any email system, including Microsoft Office 365 and Google Suite.

INKY Phish Fence scans every sent and delivered email automatically and flags malicious emails.

A comprehensive dashboard allows admins to see both the bigger pictures and to drill down to specific attacks, individuals, and individual messages. A robust search allows for detailed reporting at the granular level.

It can be set up and ready to go in just a few hours.

We're passionate about email.

Ready to talk about an issue you're facing with email security at your organization?

www.inky.com