Insights Report

AI-Enhanced Phishing Evolution

Adaptive Multi-Stage Threats Targeting Modern Enterprises

mailguard

Executive Summary

A new generation of phishing campaigns is exploiting large-language-model (LLM) tooling and automated brand-cloning services to create convincingly authentic emails, webpages and entire multi-step user journeys in minutes. These attacks:

- Impersonate high-trust brands (Microsoft 365, Meta, Spotify, Disney+, cPanel, Adobe and more) with pixel-perfect accuracy.
- Dynamically personalise logos, colour palettes and copy to the recipient's own organisation by calling open-API enrichment services such as Clearbit on-the-fly.
- Chain multiple "verification" steps, credential capture, payment capture, 2FA interception and final redirection to a legitimate site, to avoid user suspicion.

MailGuard is consistently detecting and neutralising these LLM-generated phishing kits hours, often days, before Microsoft Defender, Proofpoint and other well-known security stacks release new signatures. For boards and security leaders, the question is no longer if AI-driven phishing will breach your perimeter, but when, and whether current controls are adaptive enough to keep pace.

Email is the starting point for 91% of cyberattacks.

Source: Microsoft.com

References:
https://www.microsoft.com/en-au/security/business/security-101/what-is-business-email-compromise-bec

Key Threat Insight

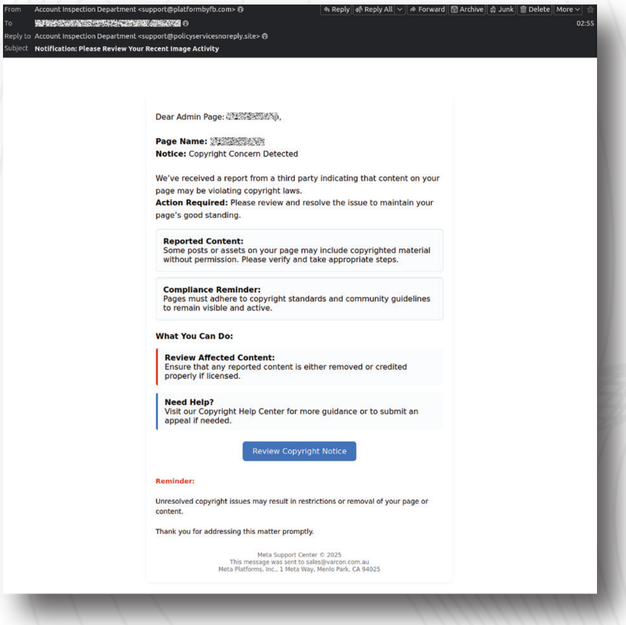
Attack Method

Stage	Tactic	Technical Details & Example
Initial Lure	Brand-faithful email spoof	Compromised domains deliver malicious email, e.g. “Microsoft Teams voicemail” alerts with benign-looking WAV links.
Phishing Site Generation	AI templating + live brand assets	Real-time retrieval of logo to theme a fake portal that perfectly matches the user’s company branding.
Adaptive Flow Control	Conditional pages & repeat prompts	JavaScript checks user agent / dev-tools, forces double-entry of credentials (“password incorrect, try again”) then auto-redirects to the real service (Adobe, Facebook, cPanel).
Multi-factor Bypass	“Enter 2FA / approve bank app”	Disney+ workspace scam instructs the victim to confirm a charge inside their banking app, piggybacking on push-based MFA to approve fraudulent payments.
Evasion	Polymorphic URLs & code obfuscation	Cloudflare-fronted hosts rotate sub-domains, embed payloads in SVG/MHTML containers, and self-destruct after predefined hit-counts to dodge retro-scans.

“Generative-AI toolchains now assemble full phishing playbooks including HTML, CSS, localisation, even vanity TLS certificates in under 90 seconds. Static detection is obsolete; we stop these threats with proprietary advanced AI before day-zero becomes hour-zero.”

— Anwar Ibrahim, CTO, MailGuard

Multi-Stage Phishing Attack Example - Facebook



From Account Inspection Department <support@platformbyfb.com>
To [redacted]
02:55
Reply to Account Inspection Department <support@policyservicesnoreply.site>
Subject **Notification: Please Review Your Recent Image Activity**

Dear Admin Page: [redacted],

Page Name: [redacted]
Notice: Copyright Concern Detected

We've received a report from a third party indicating that content on your page may be violating copyright laws.
Action Required: Please review and resolve the issue to maintain your page's good standing.

Reported Content:
Some posts or assets on your page may include copyrighted material without permission. Please verify and take appropriate steps.

Compliance Reminder:
Pages must adhere to copyright standards and community guidelines to remain visible and active.

What You Can Do:

Review Affected Content:
Ensure that any reported content is either removed or credited properly if licensed.

Need Help?
Visit our Copyright Help Center for more guidance or to submit an appeal if needed.

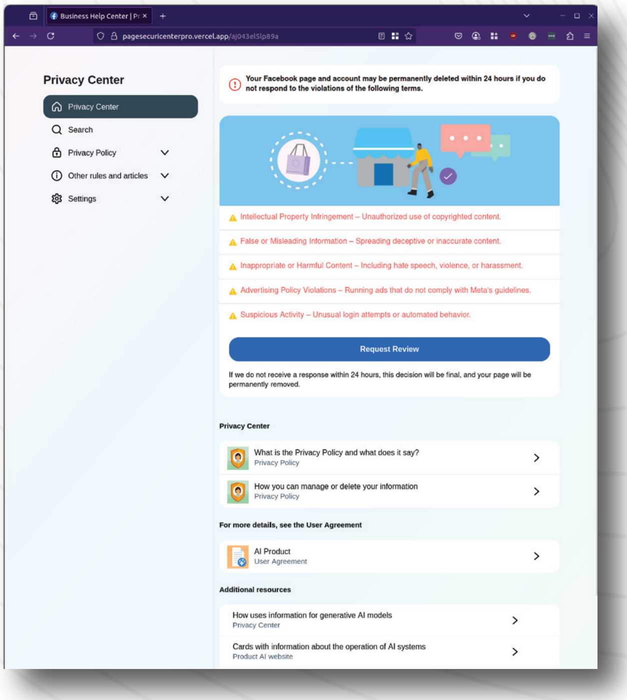
Review Copyright Notice

Reminder:

Unresolved copyright issues may result in restrictions or removal of your page or content.

Thank you for addressing this matter promptly.

Meta Support Center © 2025
This message was sent to sales@varcon.com.au
Meta Platforms, Inc., 1 Meta Way, Menlo Park, CA 94025



pagesecuricenterpro.vercel.app/aj043el5lp89a

Your Facebook page and account may be permanently deleted within 24 hours if you do not respond to the violations of the following terms.

- ⚠ Intellectual Property Infringement – Unauthorized use of copyrighted content.
- ⚠ False or Misleading Information – Spreading deceptive or inaccurate content.
- ⚠ Inappropriate or Harmful Content – Including hate speech, violence, or harassment.
- ⚠ Advertising Policy Violations – Running ads that do not comply with Meta's guidelines.
- ⚠ Suspicious Activity – Unusual login attempts or automated behavior.

Request Review

If we do not receive a response within 24 hours, this decision will be final, and your page will be permanently removed.

Privacy Center

What is the Privacy Policy and what does it say?
Privacy Policy

How you can manage or delete your information
Privacy Policy

For more details, see the User Agreement

AI Product
User Agreement

Additional resources

How uses information for generative AI models
Privacy Center

Cards with information about the operation of AI systems
Product AI website

Masquerading as Meta's support team, is designed to deceive Facebook Page administrators into handing over sensitive login credentials and two-factor authentication (2FA) codes.

Stage 1: Email advising Facebook Admins of a Copyright Infringement violation.

Stage 2: Phishing page mimicking Meta Privacy Center where targets are asked to click to 'Request Review'

Stage 3: Personal Details Capture

Stage 4: Password Capture

Stage 5: Password Loop, Password Failed Re-Enter (Tactic to confirm details are accurate)

Stage 6: MFA Capture

Stage 7: MFA Loop, MFA Failed Re-Enter

Stage 8: Request Sent confirmation

Stage 9: Return to legitimate Facebook page to avoid detection.

Multi-Stage Phishing Attack Example - Microsoft Teams Voicemail

Deceptive phishing campaign impersonates Microsoft Teams, luring users with a fake voicemail alert.

Stage 1: Email advising you have a voicemail.

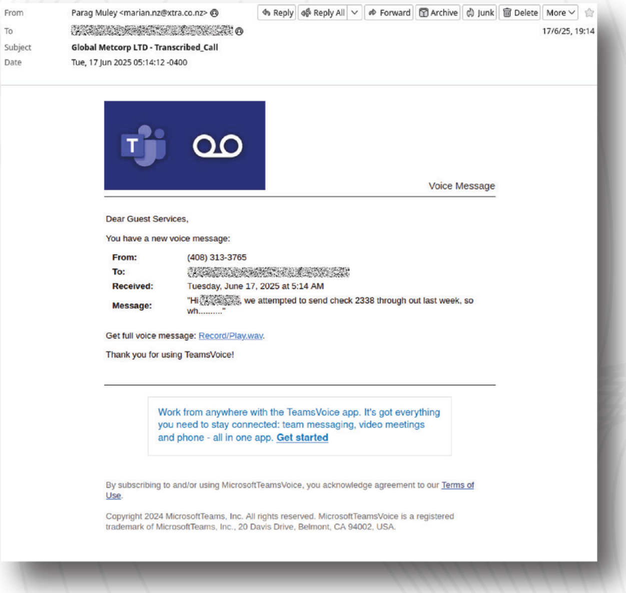
Stage 2: Voicemail player.

Stage 3: Interaction with the player launches a spoofed Microsoft login screen.

Stage 4: Password Capture

Stage 5: Password Loop, Password Failed Re-Enter (Tactic to confirm details are accurate)

Stage 6: Return to legitimate eBay or Amazon page to avoid detection.



From Parag Muley <marian.nz@xtra.co.nz>
To [redacted]
17/6/25, 19:14
Subject **Global Metcorp LTD - Transcribed_Call**
Date Tue, 17 Jun 2025 05:14:12 -0400

Voice Message

Dear Guest Services,

You have a new voice message:

From: (408) 313-3765
To: [redacted]
Received: Tuesday, June 17, 2025 at 5:14 AM
Message: "Hi [redacted], we attempted to send check 2338 through out last week, so wh.........."

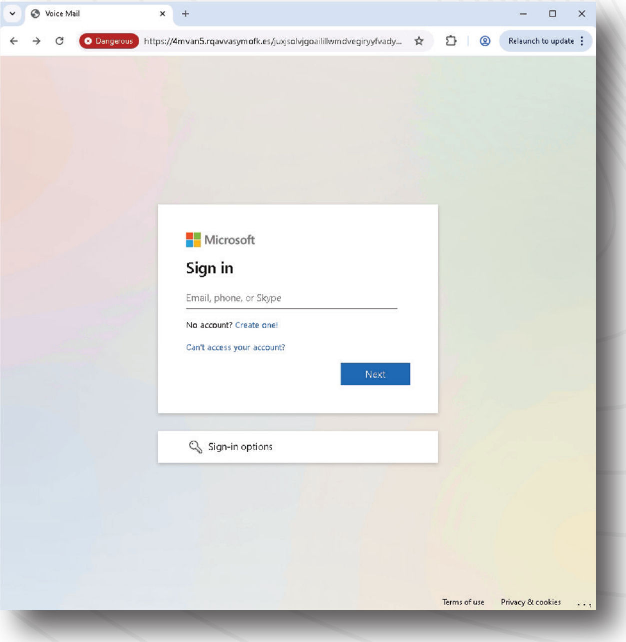
Get full voice message: Record/Play.wav.

Thank you for using TeamsVoice!

Work from anywhere with the TeamsVoice app. It's got everything you need to stay connected: team messaging, video meetings and phone - all in one app. **Get started**

By subscribing to and/or using MicrosoftTeamsVoice, you acknowledge agreement to our Terms of Use.

Copyright 2024 MicrosoftTeams, Inc. All rights reserved. MicrosoftTeamsVoice is a registered trademark of MicrosoftTeams, Inc., 20 Davis Drive, Belmont, CA 94002, USA.



https://4mvan5.rqawasymofk.es/juxjsolvjgoailillwmdvegiryyfvady...

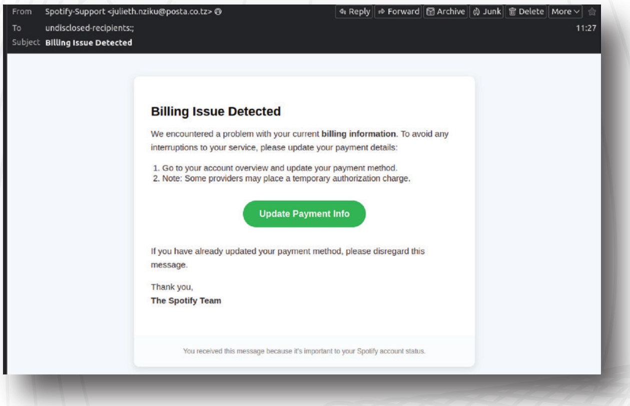
Sign in

Email, phone, or Skype

No account? Create one!

Can't access your account?

Multi-Stage Phishing Attack Example - Spotify



From Spotify-Support <julieth.nziku@posta.co.tz>

To undisclosed-recipients:;

Subject **Billing Issue Detected**

11:27

Billing Issue Detected

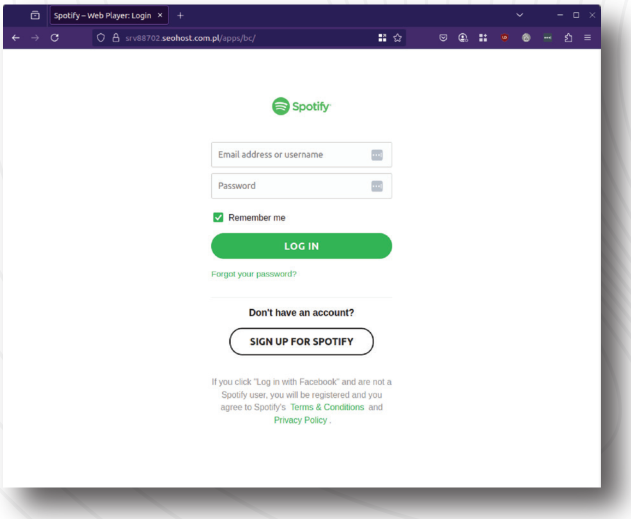
We encountered a problem with your current **billing information**. To avoid any interruptions to your service, please update your payment details:

1. Go to your account overview and update your payment method.
2. Note: Some providers may place a temporary authorization charge.

If you have already updated your payment method, please disregard this message.

Thank you,
The Spotify Team

You received this message because it's important to your Spotify account status.



Spotify – Web Player: Login

srv88702.seohost.com.pl/apps/bc/

Don't have an account?

If you click "Log in with Facebook" and are not a Spotify user, you will be registered and you agree to Spotify's Terms & Conditions and Privacy Policy .

This variant impersonates Spotify Support in an attempt to steal user login credentials and financial data.

Stage 1: Email advising that a billing issue has been detected.

Stage 2: Spotify login page captures username and password.

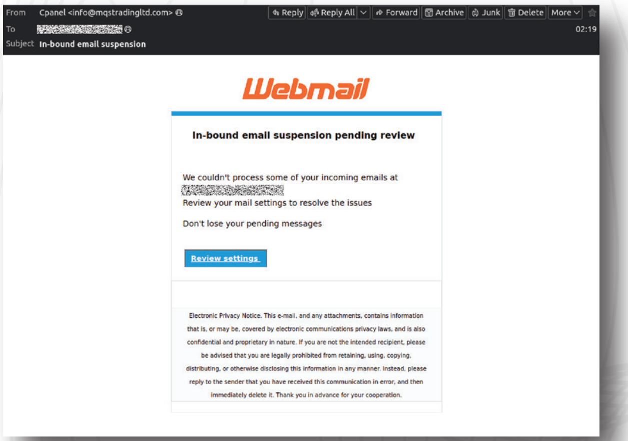
Stage 3: Update payment details captures credit card information.

Stage 4: SMS verification capture.

Stage 5: Success message confirming update of account details.

Stage 6: Return to legitimate Spotify page to avoid detection.

Multi-Stage Phishing Attack Example - cPanel Webmail Notification



From Cpanel <info@mqstradingltd.com>
To [redacted]
Subject In-bound email suspension
02:19

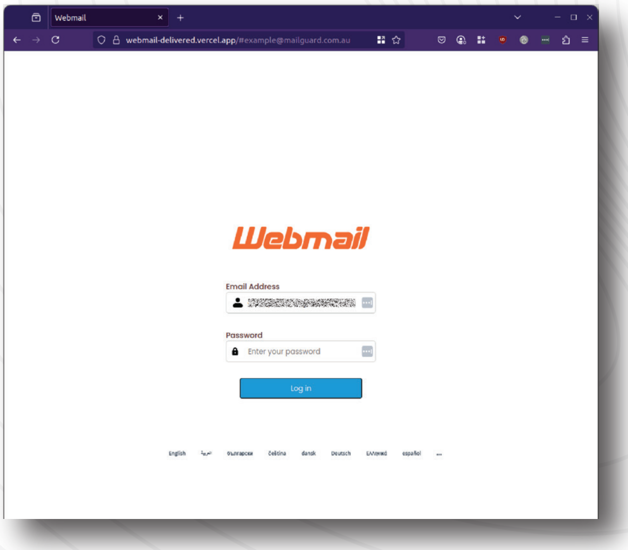
In-bound email suspension pending review

We couldn't process some of your incoming emails at [redacted]
Review your mail settings to resolve the issues

Don't lose your pending messages

Review settings

Electronic Privacy Notice. This e-mail, and any attachments, contains information that is, or may be, covered by electronic communications privacy laws, and is also confidential and proprietary in nature. If you are not the intended recipient, please be advised that you are legally prohibited from retaining, using, copying, distributing, or otherwise disclosing this information in any manner. Instead, please reply to the sender that you have received this communication in error, and then immediately delete it. Thank you in advance for your cooperation.



webmail-delivered.vercel.app/#example@mailguard.com.au

Email Address [redacted]

Password Enter your password

Under the guise of a cPanel webmail notification, this email uses social engineering to create a sense of urgency and manipulate users.

Stage 1: Email advising inbound email suspension pending review.

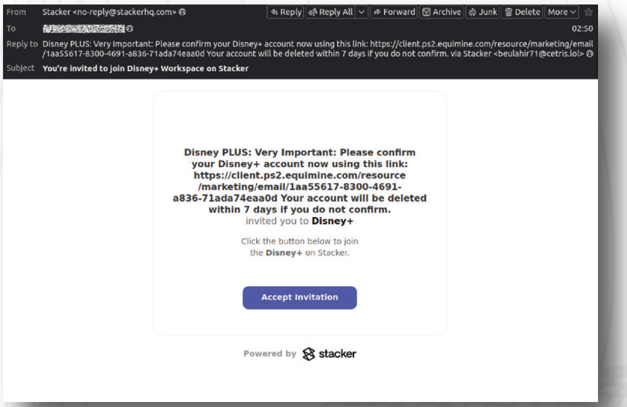
Stage 2: Review Settings button takes users to a webmail login page to capture username and password.

Stage 3: Login Loop, Login Failed Re-Enter (Tactic to confirm details are accurate).

Stage 4: Delivery Update success page.

Stage 5: Return to legitimate cPanel page to avoid detection.

Multi-Stage Phishing Attack Example - Stacker Disney+ Workspace



From Stacker <no-reply@stackerhq.com>

To [redacted]

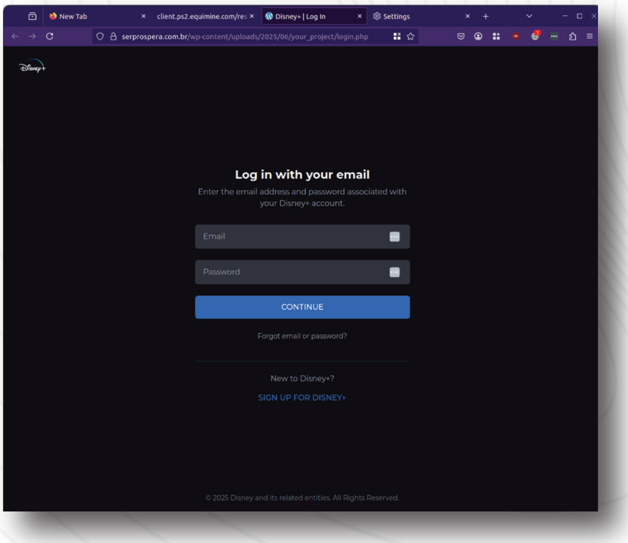
02:50

Reply to Disney PLUS: Very Important: Please confirm your Disney+ account now using this link: https://client.ps2.equimine.com/resource/marketing/email/1aa55617-8300-4691-a836-71ada74eaa0d Your account will be deleted within 7 days if you do not confirm. via Stacker <beulahir71@cetris.lol>

Subject You're invited to join Disney+ Workspace on Stacker

Disney PLUS: Very Important: Please confirm your Disney+ account now using this link: https://client.ps2.equimine.com/resource/marketing/email/1aa55617-8300-4691-a836-71ada74eaa0d Your account will be deleted within 7 days if you do not confirm.
invited you to **Disney+**

Click the button below to join the **Disney+** on Stacker.



serprospera.com.br/wp-content/uploads/2025/06/your_project/login.php

Log in with your email

Enter the email address and password associated with your Disney+ account.

© 2025 Disney and its related entities. All Rights Reserved.

A Disney+ workspace invitation from "Stacker", the message appears to be a simple collaboration invite but is in fact a multi-stage phishing attack.

Stage 1: Email invite from Stacker to join Disney+ Workspace.

Stage 2: Account Confirmation Page.

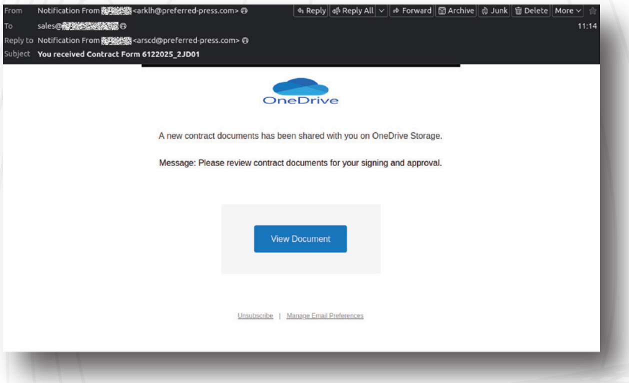
Stage 3: Redirect to Login to capture username and password.

Stage 4: Payment Details page captures credit card information.

Stage 5: Requests transaction approval via the victims mobile banking app (presumably to bypass MFA).

Stage 6: Return to legitimate Disney+ page to avoid detection.

Multi-Stage Phishing Attack Example - Microsoft OneDrive

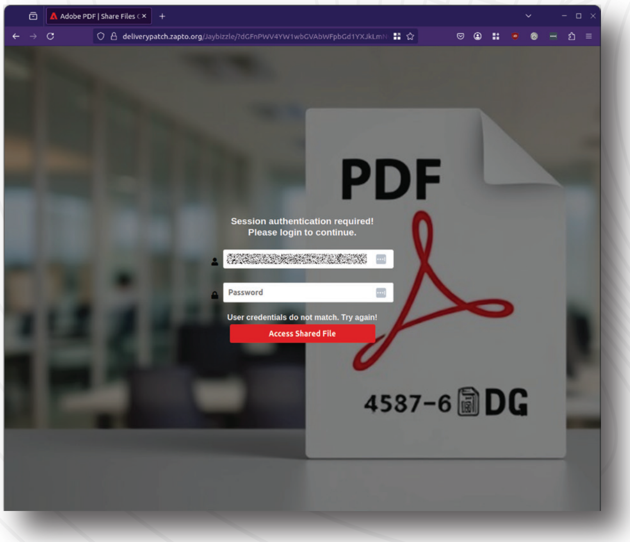


From Notification From [redacted] <arklh@preferred-press.com>
To sales@[redacted]
Reply to Notification From [redacted] <arscd@preferred-press.com>
Subject You received Contract Form 6122025_2JD01

11:14

A new contract documents has been shared with you on OneDrive Storage.

Message: Please review contract documents for your signing and approval.



deliverypatch.zapto.org/Jaybizzle/?dGFnPWV4YW1wbGVAbWFpbGd1YXJkLmN

Session authentication required!
Please login to continue.

[redacted]

Password

User credentials do not match. Try again!

Emails spoofing OneDrive lure users into revealing their login credentials to access an Adobe PDF file that has been shared with them.

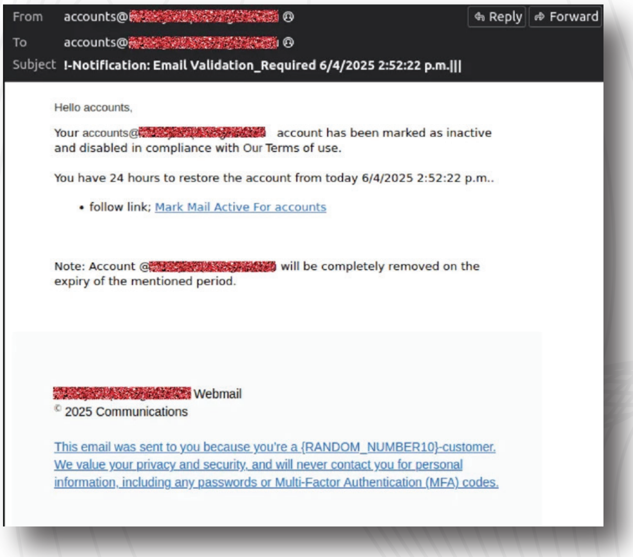
Stage 1: Email advising that a new contract document has been shared with you via OneDrive.

Stage 2: Clicking View Document directs users to an Adobe PDF-themed login screen capturing username and password.

Stage 3: Login Loop, Re-Enter required to access the files (confirming details are accurate).

Stage 4: Return to legitimate Adobe PDF Reader page to avoid detection.

Multi-Stage Phishing Attack Example - Email Validation



From accounts@[redacted]

To accounts@[redacted]

Subject !-Notification: Email Validation_Required 6/4/2025 2:52:22 p.m.|||

Hello accounts,

Your accounts@[redacted] account has been marked as inactive and disabled in compliance with Our Terms of use.

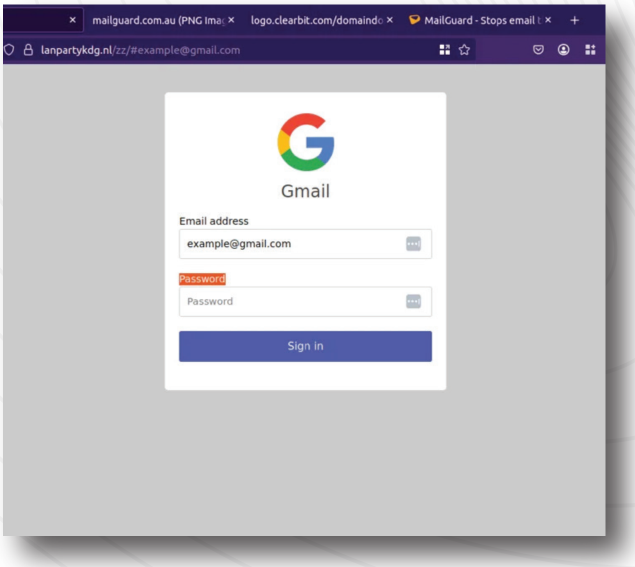
You have 24 hours to restore the account from today 6/4/2025 2:52:22 p.m..

- follow link; Mark Mail Active For accounts

Note: Account @[redacted] will be completely removed on the expiry of the mentioned period.

[redacted] Webmail
© 2025 Communications

This email was sent to you because you're a {RANDOM_NUMBER10}-customer. We value your privacy and security, and will never contact you for personal information, including any passwords or Multi-Factor Authentication (MFA) codes.


mailguard.com.au (PNG Imag
logo.clearbit.com/domaindo
MailGuard - Stops email t
lanpartykdg.nl/zz/#example@gmail.com
Gmail
Sign in

Email address
example@gmail.com

Password
Password

Impersonating a legitimate system notification, using deceptive tactics to steal user credentials, this attack appears simple at first glance, but its execution is cunning. It presents as an email validation notification, relying on dynamic redirects and branding lookups to improve the scam's credibility and avoid detection.

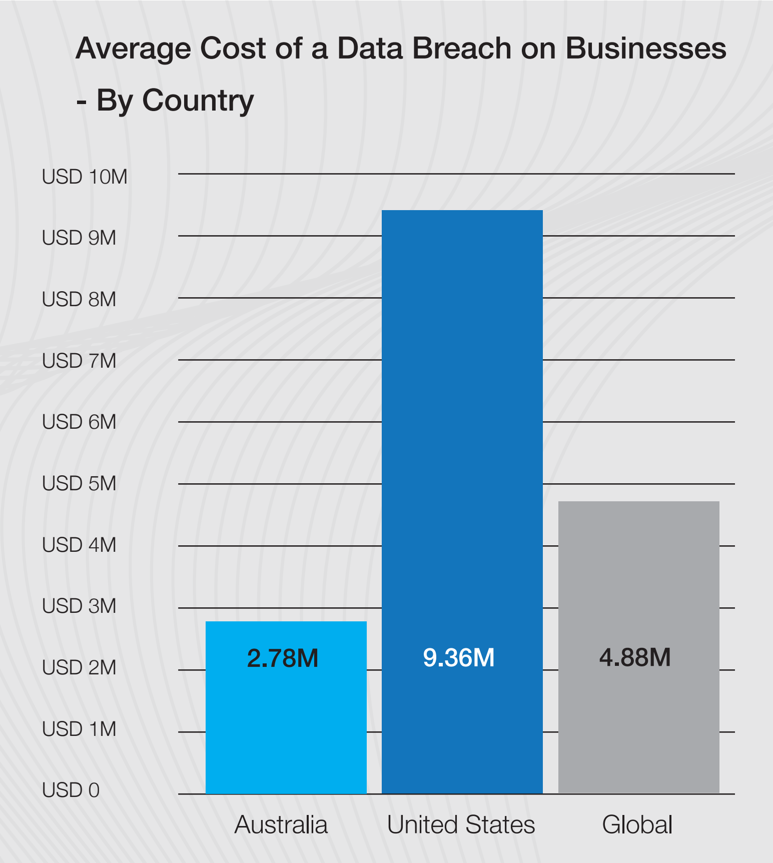
Stage 1: Email advising your account has been marked as inactive and disabled.

Stage 2: Clicking the mark email as active link redirects users to a Gmail Login phishing page to capture usernames and passwords.

Stage 3: Reports a login failure and redirects to a login with the real domain for the target victims company, causing the user to believe the initial login was merely a technical glitch.

Why It's So Effective

- **Hyper-realism at scale**: LLMs ingest brand guidelines and spit out idiomatic copy, correct colour gradients and locally relevant legal jargon.
- **Dynamic personalisation:** Logo-enrichment APIs plus recipient-specific email variables (name, domain) erode the "generic spam" tell-tale.
- **Filter evasion:** No malware attachments; links point to newly registered but clean domains, often protected by Cloudflare or Azure Front Door.
- **Psychological pressure:** Urgency ("billing issue", "copyright strike", "email suspension"), combined with familiar UI, accelerates reflex clicks.



IBM Cost of a Data Breach Report 2024

References:
https://www.ibm.com/reports/data-breach

Consequences of Inaction

Account Takeover & Lateral Movement: Compromised Microsoft 365 tenants provide single-sign-on keys to SharePoint, Teams and CRM data.

Payment Diversion & Fraud: Stolen card details plus 2FA codes enable instant transactions or payroll redirection before anomalies surface.

Data Breach & Compliance Penalties: Sensitive email archives exfiltrated; potential violation of ISO 27001, GDPR, and Australian Privacy Act.

Brand & Shareholder Impact: Reputational fallout drives customer churn and erodes market trust; the mean cost per email-initiated breach now exceeds US $4.45 million (IBM 2024 DBIR).

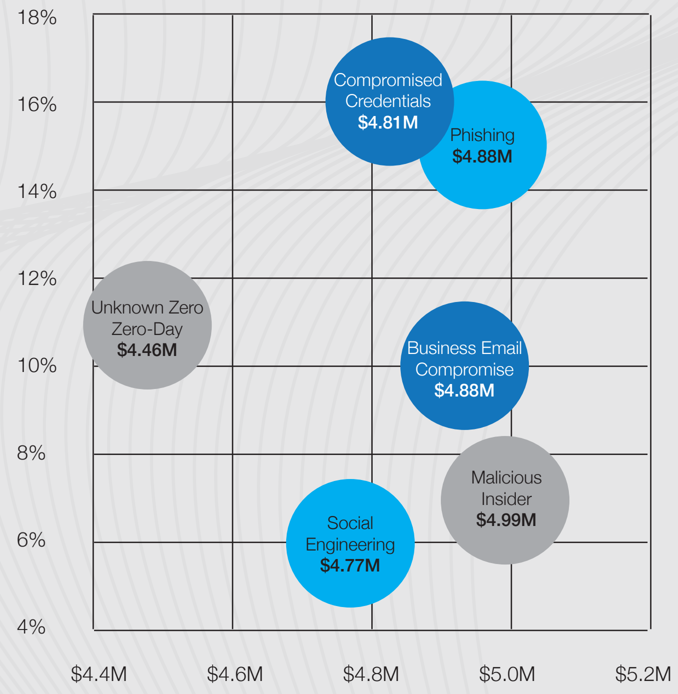
> *"LLM-assisted phishing kits now stitch together reverse-proxy frameworks such as EvilProxy or Modlishka with on-demand brand-cloning and Cloudflare Workers delivery. In one command, attackers launch a site that transparently relays the entire Microsoft 365 or Google Workspace OAuth flow and captures the returned session cookies. MailGuard's sophisticated protection is built to detect subtle variations in content and delivery patterns to neutralise such threats on first encounter."*
>
> **— Prathik Chandrashekar, Head of Engineering, MailGuard**

Why MailGuard Is Critical

- **Rapid protection:** Advanced threats blocked in-flight, moved out of harm's way.
- **Predictive lead:** Proprietary AI models detect novel attack blueprints (QR-code phish, Clearbit logo abuse) months before industry signatures converge.
- **Continuous AI/LLM counter-training:** Real-time ingestion of adversarial templates ensures detection models evolve as quickly as attackers innovate.
- **Seamless integration** and deployment inline with Microsoft 365 and Google, blocking threats before they reach users.
- This is **precision cyber defence at speed**, engineered to intercept threats before human error comes into play.

All campaigns detailed in this report were **100% stopped by MailGuard.**

Cost and Frequency of Initial Attack Vectors that are Email-Related



IBM Cost of a Data Breach Report 2024

292 Days

to identify and contain breaches involving stolen credentials.

References:
https://www.ibm.com/reports/data-breach

A Strategic Call To Action

Cybercriminals have automated creativity; protection cannot rely on human vigilance alone. Executive leadership must:

1. **Audit detection latency:** Measure how many minutes (not days) your stack takes to learn and block a brand-new phishing kit.
2. **Adopt advanced AI defence:** Insert MailGuard as an additional control point before the inbox. No rip-and-replace required.
3. **Champion a "trust-nothing email" culture:** Embed continuous training and zero-trust principles, backed by technical enforcement.

Modern resilience demands pre-emptive speed.

Let's **schedule a time** to review your organisation's security posture, and assess how **MailGuard can complement your defences**, strengthening control, boosting resilience and tackling new strategic challenges**, to stay ahead of evolving threats**.

Built in Australia.
Trusted Globally.

MailGuard is a global leader in email threat detection. A pioneer in cloud email security since 2001, MailGuard invented the concept of pre-filtering email threats before inbox delivery, laying the foundation for the Secure Email Gateway (SEG) category.

Today, MailGuard protects organisations globally with AI-powered threat detection, seamlessly deployed inline with Microsoft's ecosystem and Google, among other email providers.

At the heart of our platform is **MyGuard** — our proprietary AI threat engine developed with over **A$35 million in R&D**. MyGuard combines:

- Gen-AI powered LLMs
- Bayesian and fingerprint-based classifiers
- Real-time behavioural heuristics

…to stop advanced threats on first encounter before they reach staff inboxes — including those that bypass Microsoft and other cloud email security vendors.

MailGuard is **ISO/IEC 27001:2022 certified**, trusted by **over 5,500 organisations**, including governments, law firms, banks, hospitals, and ASX-listed companies. Recognised for our unmatched speed in detecting zero-day email threats, we have consistently stopped sophisticated exploits, like **QR code phishing, Dropbox-based malware**, and **Azure AD Guest Invite fraud**, months ahead of Microsoft, and other leading platforms.

In an era of rising cyber regulation and board-level accountability, MailGuard enhances your Microsoft 365 or Google security stack with minmal disruption, easy activation, and elite-speed protection, fulfilling your fiduciary and operational responsibilities.

Trusted by Global Leaders.
Since 2001.

- A **leader in advanced 'zero zero-day' email threats** missed by Microsoft 365 and other 3rd party vendors.
- Achieve peace of mind with MailGuard, a solution **trusted by global leaders** that ensures your email is secure.
- Benefit from **A$35M+ in R&D, including proprietary AI & ML-powered threat detection**, to boost your cybersecurity confidence.
- AI-powered email threat detection and inline architecture intercepts and blocks threats hours faster, on first encounter.

"It's the type of innovation that we want to see."

— Satya Nadella, CEO & Chairman, Microsoft

"MailGuard has developed world-leading cloud and email security IP. This is IP that is unique to Australia; it's among the leading cloud and email security solutions anywhere in the world."

— Hon. Malcolm Turnbull, Former Australian Prime Minister

"You are being led by what I see as one of the world's best, at preventing and protecting your secure infrastructure, securing your people, and securing your business"

— Steve Miller, COO, Microsoft Asia

mailguard

www.mailguard.com.au

Let’s Connect

Make time today to talk to our local team of experts about fortifying your inboxes.
expert@mailguard.com.au