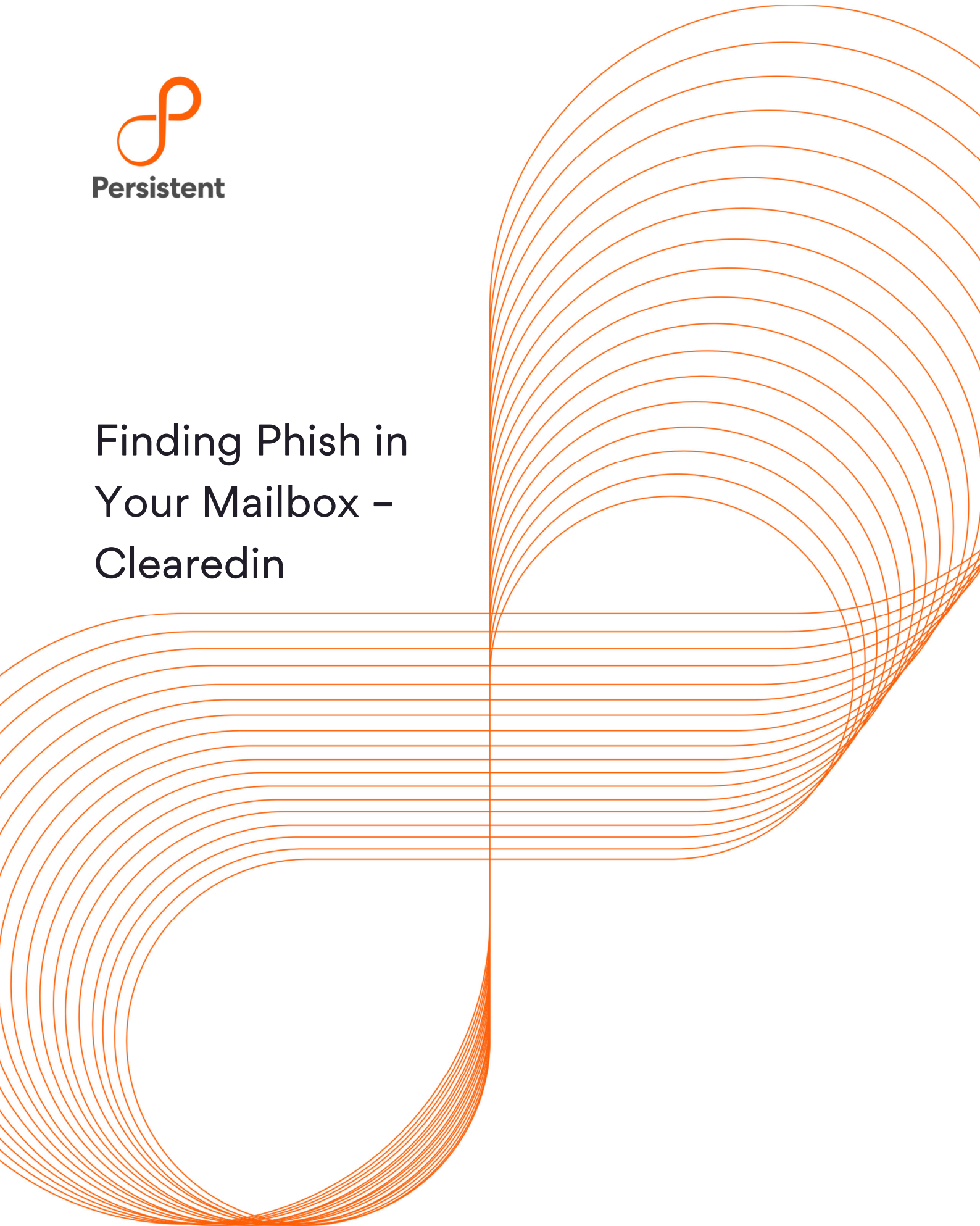




Persistent

Finding Phish in Your Mailbox – Clearedin



About the Client

Clearedin is a start-up with the business goal of building the secure business communications graph to unlock the value of millions of emails and messages sent daily. By processing the communications metadata, the client is helping businesses to be more secure and grow faster.

Problem Statement

- The goal was to build profiles of individuals and companies listing all channels of communication, starting with business emails.
- This will result in building a relationship graph among individuals, services (banks, insurance, healthcare, ecommerce, social network, etc.) and companies.
- By processing the communications metadata, Clearedin wants to help businesses to be more secure and grow faster.
- Large amount of semi-structured data with diverse access pattern was a challenge along with deduplication logic for emails/profiles and compliance issues around email data privacy with required authorization solution.
- Phish detection in real time should happen instantly as soon as the email lands in the mailbox.

What We Proposed

- Persistent Systems partnered with Clearedin to develop an end-to-end data processing pipeline to extract email metadata for historic and Realtime email using serverless architecture with Amazon stack.
- The solution locks Phishing/Suspicious messages in user mailbox whenever detected. User's feedback in terms of categorization of messages/contacts for good/bad/unsure categories is captured and incorporated with Phish detection mechanism.
- The solution also covers communication via 'Slack' and has extensions to support document sharing solutions like box.com, dropbox.com, OneDrive and Google Drive.
- Clearedin Dashboard that shows various actionable reports for various personas. Persistent team also developed user agents like Chrome extension, Gmail Add-ons, Office 365 Add-ins, COM Add to get the user's feedback on email categorization as well as providing vital information like crowd wisdom.

How AWS Services were used as Part of the Solution

- **Amazon API Gateway** – Provides REST API interface for the UI components as well as REST API endpoints for the email service providers to publish notifications for changes to the mailboxes.
- **Amazon S3** – Provides highly available cheap block storage used to hold historical email header data used for batch processing.
- **Amazon EMR** – SPARK cluster used to process historical email header data and build relationship graph.
- **Amazon Lambda** – Serverless framework by AWS used to process email headers in Realtime. Also used to serve the REST APIs for the UI components.
- **Amazon SQS** – Simple Queuing Service to help faster onboarding of users.

Outcome/Results and Benefits

- The solution provides email security by detecting phishing emails, sender's impersonation, spoofed domains etc. to secure business communication.
- Microservices + Serverless architecture provides great flexibility and easy scale up/down depending on the workload.
- The solution is completely hosted in AWS. This helped in reducing time for development cycle.

About Persistent Systems

Persistent Systems (BSE & NSE: PERSISTENT) builds software that drives our customers' business; enterprises and software product companies with software at the core of their digital transformation.

www.persistent.com

India

Persistent Systems Limited
Bhageerath, 402,
Senapati Bapat Road
Pune 411016.
Tel: +91 (20) 6703 0000
Fax: +91 (20) 6703 0009

USA

Persistent Systems, Inc.
2055 Laurelwood Road, Suite 210
Santa Clara, CA 95054
Tel: +1 (408) 216 7010
Fax: +1 (408) 451 9177
Email: info@persistent.com