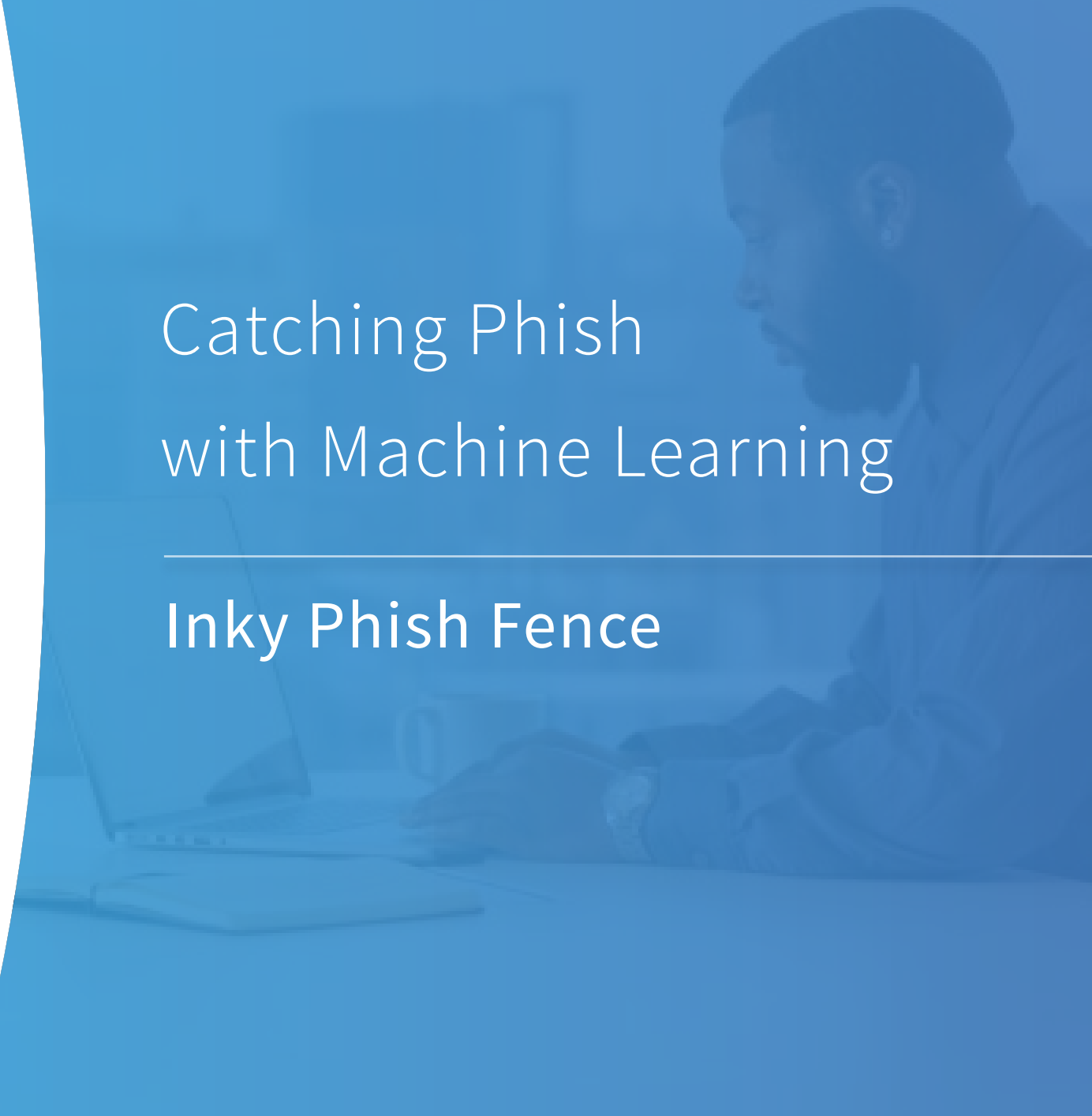
inky®

Inky.com

Catching Phish with Machine Learning

Inky Phish Fence



Problem Definition

Email-based phishing is the #1 cybersecurity attack vector, generating over **$1B / year of wire fraud**.

Our goal is to identify these emails “zero-day” — without relying on user reports of phishing inks or bad attachments.

“Phishing” implies Impersonation

Brand impersonation emails look like legitimate transactional emails from DocuSign, Microsoft, etc. They’re usually aimed at credential harvesting.

Spear phishing or “Business Email Compromise (BEC)” emails look like personal emails from important people: wire requests, invoices, requests for W-2s, attachments “for review”.

Spear Phishing

Spear Phishing Detection Example

Expense Report Problem

Friday February 23rd

DR

David Ridder
david.ridder@gmail.com

Inky thinks this message may be fraudulent

Report Phish Feedback Powered by Inky

Hey John,

Can you take care of this for me? Can you just login here and approve it?
www.inky.com/expense-approvals/123432

Thanks,
Dave

Spear Phishing Detection
In One Sentence

To automatically identify spear phishing emails, we maintain models of legitimate mail from senders and look for new, outlier emails using a range of standard anomaly detection algorithms.

Spear Phishing Detection

Solution Architecture

Our system maintains a company-wide social graph and detects anomalies to catch spear phishing.

1. Maintain social graph of senders and recipients by monitoring all email traffic.
2. When each new mail comes in, compute a profile capturing its writing style, geographical route, and other properties.
3. Compare the profile with profiles we've seen before for that sender.
4. If the new email has a different enough profile, add a red or yellow warning banner.
5. New profiles are learned over time as more examples arrive.

Spear Phishing Detection
Details

We represent new and historical emails as feature vectors.

Example features:

- Presence/absence of specific headers
- Received headers and implied geolocation of referenced hosts/IPs
- Recipients (To:/Cc:)
- MIME structural details and naming conventions
- Originating mail client type
- Header token frequencies

Spear Phishing Detection
Challenges

- Infinite possible email headers implies variable feature vocabulary.
- Feature hashing is necessary to handle unseen features.
- Many senders = many models.
- Intrinsic complexity of email leads to sizable model per sender.
- Senders send mail from multiple places and devices.
- Body replay attacks limit the value of body text analysis.
- Email language is often abbreviated, therefore hard to model with NLP.
- Some anomolies matter more than others in practice.
- Mail usage patterns and mail infrastructure naturally drift over time.

Spear Phishing Detection
Example

Next we'll look at three emails purportedly from John Doe about a 40th birthday party. To the user, they look identical. However, the first two are legitimate, but the third is spoofed.

Header analysis and historic profiling reveal that the 3rd message is actually very different.

The first 2 come from different servers, but they're both Gmail's and located in the US (209.85.220.41, 209.85.220.48).

The 3rd comes from Brazil (150.165.253.150). It also made several other hops along the way.

Spear Phishing Detection

Legitimate Email #1

```
Return-Path: <john@gmail.com>
Received: from mail-sor-f41.google.com (mail-sor-f41.google.com. [209.85.220.41])
        by mx.google.com with SMTPS id 63sor1271989qth.102.2018.04.11.09.44.25
        (Google Transport Security);
        Wed, 11 Apr 2018 09:44:25 -0700 (PDT)
Received-SPF: pass (google.com: domain of john@gmail.com designates 209.85.220.41 as permitted sender) client-ip=209.85.220.41;
Authentication-Results: mx.google.com;
       dkim=pass header.i=@gmail.com header.s=20161025 header.b=SKd8nAlO;
       spf=pass (google.com: domain of john@gmail.com designates 209.85.220.41 as permitted sender) smtp.mailfrom=john@gmail.com;
       dmarc=pass (p=NONE sp=QUARANTINE dis=NONE) header.from=gmail.com
DKIM-Signature: v=1; a=rsa-sha256; c=relaxed/relaxed; d=gmail.com; s=20161025;
        h=mime-version:references:in-reply-to:from:date:message-id:subject:to:cc;
        bh=...; b=...
X-Google-DKIM-Signature: v=1; a=rsa-sha256; c=relaxed/relaxed; d=1e100.net; s=20161025;
        h=x-gm-message-state:mime-version:references:in-reply-to:from:date:message-id:subject:to:cc;
        bh=...; b=...
X-Gm-Message-State: ALQs6tBv...
X-Google-Smtp-Source: AIpwx490T...
X-Received: by 10.200.53.164 with SMTP id k33mr8405274qtb.37.1523465064900; Wed, 11 Apr 2018 09:44:24 -0700 (PDT)
MIME-Version: 1.0
References: <CAL+9f6CR7xwS4-Wo2wYyqk+xniQgkoPwoRHyTLW+=82gx9sRdQ@mail.gmail.com>
In-Reply-To: <CAL+9f6CR7xwS4-Wo2wYyqk+xniQgkoPwoRHyTLW+=82gx9sRdQ@mail.gmail.com>
From: John Doe <john@gmail.com>
Date: Wed, 11 Apr 2018 16:44:14 +0000
Message-ID: <CAGskw+-JvZin0mh-P+sm7WCFeLyxBpfU8KK3wgyT7MSgONsiLw@mail.gmail.com>
Subject: Re: 40th birthday
To: Jane Doe <jane@gmail.com>
Content-Type: multipart/alternative; boundary="001a113f275a056ed10569955ad2"
```

Spear Phishing Detection

Legitimate Email #2

```
Return-Path: <john@gmail.com>
Received: from mail-sor-f48.google.com (mail-sor-f48.google.com. [209.85.220.48])
        by mx.google.com with SMTPS id 63sor1271989qth.102.2018.04.10.12.24.25
        (Google Transport Security);
        Tue, 10 Apr 2018 12:24:25 -0700 (PDT)
Received-SPF: pass (google.com: domain of john@gmail.com designates 209.85.220.48 as permitted sender) client-ip=209.85.220.48;
Authentication-Results: mx.google.com;
       dkim=pass header.i=@gmail.com header.s=20161025 header.b=SKd8nAlO;
       spf=pass (google.com: domain of john@gmail.com designates 209.85.220.48 as permitted sender) smtp.mailfrom=john@gmail.com
       dmarc=pass (p=NONE sp=QUARANTINE dis=NONE) header.from=gmail.com
DKIM-Signature: v=1; a=rsa-sha256; c=relaxed/relaxed; d=gmail.com; s=20161025;
        h=mime-version:references:in-reply-to:from:date:message-id:subject:to:cc;
        bh=...; b=...
X-Google-DKIM-Signature: v=1; a=rsa-sha256; c=relaxed/relaxed; d=1e100.net; s=20161025;
        h=x-gm-message-state:mime-version:references:in-reply-to:from:date:message-id:subject:to:cc;
        bh=...; b=...
X-Gm-Message-State: ALQs6tBv...
X-Google-Smtp-Source: AIpwx490T...
X-Received: by 10.200.53.163 with SMTP id k33mr8405274qtb.37.1523465064800; Tue, 10 Apr 2018 12:24:24 -0700 (PDT)
MIME-Version: 1.0
From: John Doe <john@gmail.com>
Date: Tue, 10 Apr 2018 19:24:14 +0000
Message-ID: <CAGskw+-JvZin2mh-M+sm7WCFeLyxBpfU8KK3wgyT7MSgONseLw@mail.gmail.com>
Subject: 40th birthday
To: Jane Doe <jane@gmail.com>
Content-Type: multipart/alternative; boundary="001a113f275a056e546345645ae4"
```

Spear Phishing Detection

Spoofed Email

```
Return-Path: <fabiolabrazaquino@cchla.ufpb.br>
Received: from mx1.ufpb.br (mx1.ufpb.br. [150.165.253.150])
        by mx.google.com with ESMTPS id m38si2763821qta.396.2018.04.03.20.48.08
        (version=TLS1_2 cipher=ECDHE-RSA-AES128-GCM-SHA256 bits=128/128);
        Tue, 03 Apr 2018 20:48:09 -0700 (PDT)
Received-SPF: pass (google.com: domain of fabiolabrazaquino@cchla.ufpb.br designates 150.165.253.150 as permitted sender
Authentication-Results: mx.google.com;
       dkim=pass header.i=@cchla.ufpb.br header.s=mailcchla header.b=YhPUXuIL;
       spf=pass (google.com: domain of fabiolabrazaquino@cchla.ufpb.br designates 150.165.253.150 as permitted sender) s
Received: from email.ufpb.br (email.ufpb.br [150.165.253.99]) by mx1.ufpb.br (Postfix) with ESMTP id 04425B78; Wed,
  4 Apr 2018 00:47:51 -0300 (-03)
Received: from localhost (localhost [127.0.0.1]) by email.ufpb.br (Postfix) with ESMTP id 4C0D340631; Wed,
  4 Apr 2018 00:47:51 -0300 (BRT)
Received: from email.ufpb.br ([127.0.0.1]) by localhost (email.ufpb.br [127.0.0.1]) (amavisd-new, port 10032) with ESMTP
  4 Apr 2018 00:47:49 -0300 (BRT)
Received: from localhost (localhost [127.0.0.1]) by email.ufpb.br (Postfix) with ESMTP id 1508A40674; Wed,
  4 Apr 2018 00:47:49 -0300 (BRT)
DKIM-Filter: OpenDKIM Filter v2.10.3 email.ufpb.br 1508A40674
DKIM-Signature: v=1; a=rsa-sha256; c=relaxed/relaxed; d=cchla.ufpb.br; s=mailcchla; t=1522813669;
  bh=...; h=MIME-Version:To:From:Date:Message-Id; b=...
X-Virus-Scanned: amavisd-new at email.ufpb.br
Received: from email.ufpb.br ([127.0.0.1]) by localhost (email.ufpb.br [127.0.0.1]) (amavisd-new, port 10026) with ESMTP
  4 Apr 2018 00:47:48 -0300 (BRT)
Received: from [172.20.10.6] (unknown [197.210.25.123]) by email.ufpb.br (Postfix) with ESMTPSA id 496F040655; Wed,
  4 Apr 2018 00:47:13 -0300 (BRT)
MIME-Version: 1.0
X-Mailer: Thunderbird
Content-Transfer-Encoding: quoted-printable
Content-Description: Mail message body
Subject: Re: 40th birthday
To: Jane Doe <jane@gmail.com>
From: John Doe <john@gmail.com>
Date: Wed, 04 Apr 2018 11:46:48 +0800
Reply-To: mikh.fridman@gmail.com
Message-Id: <20180404034714.496F040655@email.ufpb.br>
Content-Type: text/plain; charset="iso-8859-1"
```

Spear Phishing Detection

All the messages are DKIM-signed and receive a passing result. However, the 3rd is signed by cchla.ufpb.br not gmail.com.

Other differences include headers added and removed. Legitimate Gmail messages have X-Gm-Message-State, X-Google-Smtp-Source.
Spoofed has X-Mailer, X-Virus-Scanned, DKIM-Filter.

MIME structure differences: legitimate messages are multipart/alternative while the spoofed is just a single text/plain.

Brand Forgery

Brand Forgery Example

Alert: Your American Express

Friday February 23rd

AE

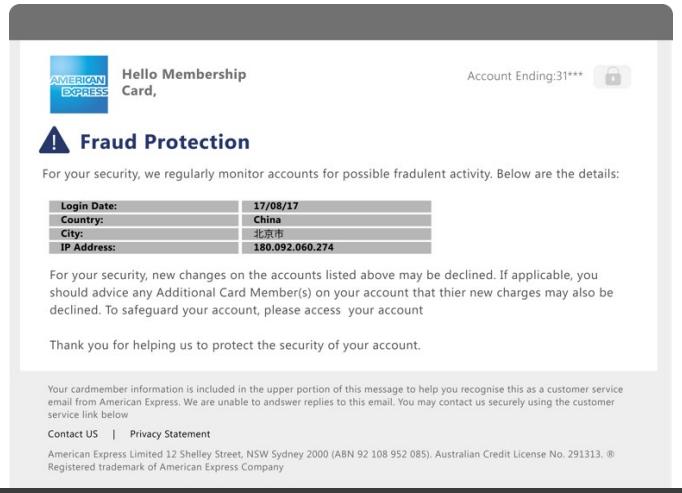
Àmerican Express Support
àmericanexpresssupport@aexp-ip.com

WARNING!

Inky Phish Fence Thinks this message looks suspicious. (From: id196@aexp-ip.com)

Brand Impersonation: The message appears to be impersonating American Express, but was not sent from an authorized domain controlled by American Express.

Report Phish Feedback Powered by Inky



Hello Membership Card,

Account Ending:31***

Fraud Protection

For your security, we regularly monitor accounts for possible fradulent activity. Below are the details:

Login Date:	17/08/17
Country:	China
City:	北京市
IP Address:	180.092.060.274

For your security, new changes on the accounts listed above may be declined. If applicable, you should advice any Additional Card Member(s) on your account that thier new charges may also be declined. To safeguard your account, please access your account

Thank you for helping us to protect the security of your account.

Your cardmember information is included in the upper portion of this message to help you recognise this as a customer service email from American Express. We are unable to andswer replies to this email. You may contact us securely using the customer service link below

Contact US | Privacy Statement

American Express Limited 12 Shelley Street, NSW Sydney 2000 (ABN 92 108 952 085). Australian Credit License No. 291313. ® Registered trademark of American Express Company

Brand Forgery Detection
In One Sentence

To automaticaly identify a brand forgery email, first determine what brand a human would say the email appears to be from, then prove that the email was not sent from a mail server operating on behalf of that brand.

Brand Forgery Detection

Solution Architecture

Our system identifies brand forgery attempts in two steps.

Black Box 1: Brand Identifier

1. (Preprocessing)
   Train machine learning models on brand text and HTML email properties.
2. (Preprocessing)
   Train computer vision models on brand imagery (like Facebook does for faces).
3. This black box takes a raw email and outputs what brand a human would say the mail is from.

Black Box 2: Fraud Detector

1. (Preprocessing)
   Construct database of valid sending domains for major brands.
2. Verify email is sent from a valid sender for the brand output by black box 1 using cryptographic standards.
3. If email is not from a valid sender, add a red or yellow warning banner.

Brand Forgery Detection

Details

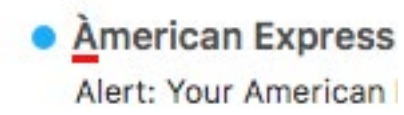
Attackers cleverly design brand forgery emails to fool both the humans and the machines.

Techniques that fool humans:

- Resending the exact HTML from a branded transactional mail with slight modifications.
- Using brand-indicative imagery and text
- Registering new typo domains or similar (e.g., arnericanexpress.com)

Techniques that fool machines:

- Cloaking brand-indicative text with Unicode, etc.
- Stuffing text and keywords into the HTML to fool Bayesian classifiers
- Sending mail from Office 365, G Suite, or other high-reputation accounts.
- Randomizing URLs or malware attachments.



Brand Forgery Detection

Details

To foil the attackers, we must model the way humans view each message and implement countermeasures against the techniques meant to fool machines.

Techniques to thwart the attackers:

- Recognizing brand-indicative imagery, colors, styling with ML
- Approximate matching of brand-indicative text and domain names
- Recognizing machine-generated URLs and domain names
- Requiring alignment of sending domain with sending brand

Brand Forgery Detection

Challenges

Deep learning works well for the basic task, but there are challenges.

- Prior work on computer vision has targeted photographic imagery
- Image matching/hashing algorithms tend to focus on exact matching for dedupuing

Brand Forgery Detection

Challenges

Brand imagery evolves over time.

1954

1954-1957

1957-1969

1969-1994

1994-1999

1999-present

Source: logos.wikia.com

Brand Forgery Detection

Challenges

Human memory of brand imagery is poor.

Less Accurate

More Accurate

Brand Forgery Detection

Challenges

A brand logo in an email doesn't necessarily imply impersonation.

BioHealth Innovation

Forward Subscribe

April 9, 2018

Third Annual Crab Trap: Finalists Announced Tomorrow

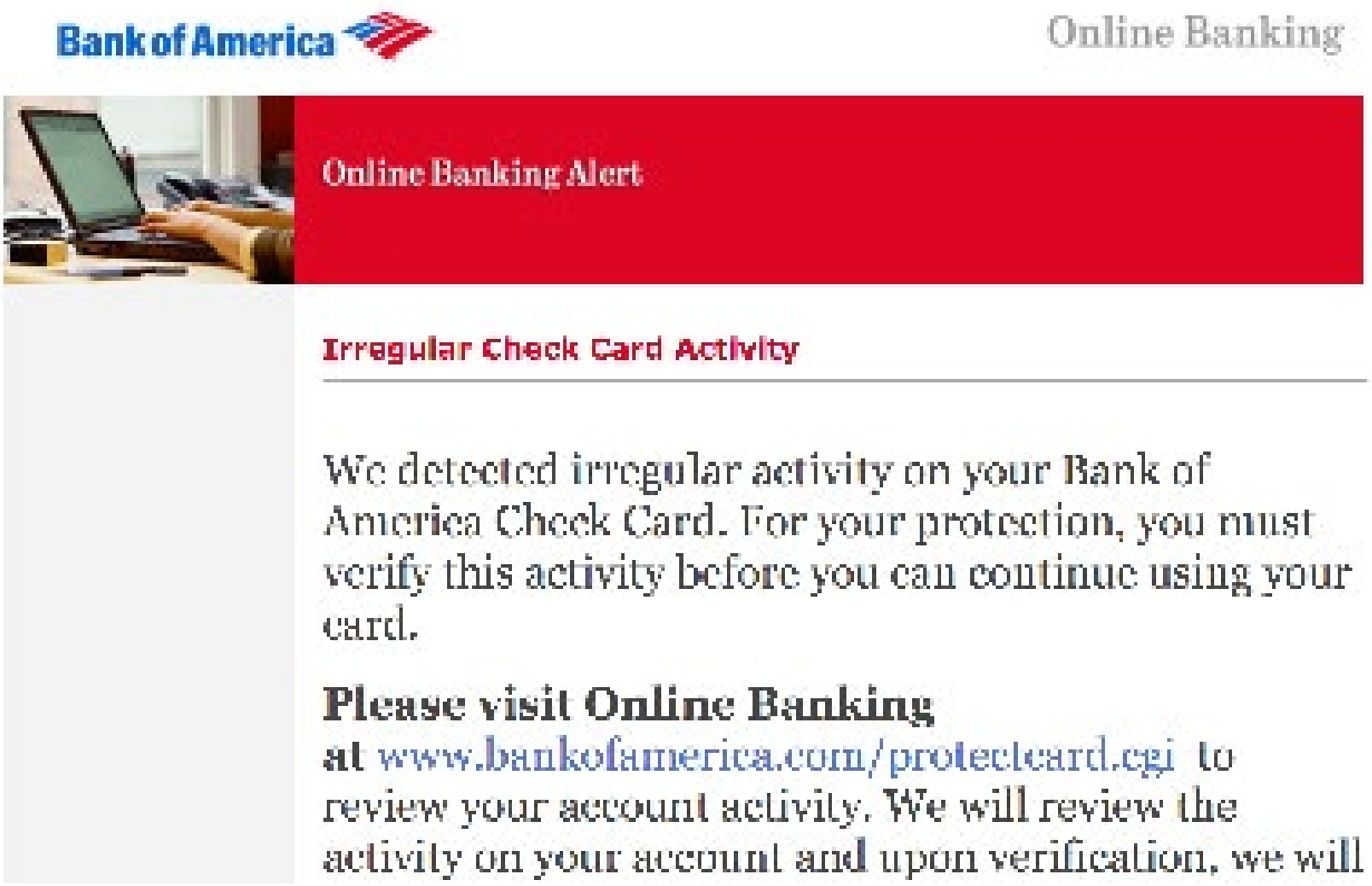
You thought the Shark Tank was tough...
wait until the BioHealth Capital Region Crab Trap!
Five finalists will have a chance to win the grand prize by presenting infront of a panel of prominent industry funding experts and executives including:

- **Rich Bendis** (Moderator), President & CEO, BioHealth Innovation, Inc.
- **Christian Barrow**, Executive Director, Life Sciences, J.P. Morgan
- **Shaun Grady**, VP, Strategic Partnering & Business Development, AstraZeneca
- **Sara Nayeem**, Partner, New Enterprise Associates (NEA)
- **Dr. Paul Silber**, Founding Principal, Blu Venture Investors
- **Robert Silverman**, Head Externalized Drug Discovery

Brand Forgery Detection

Challenges

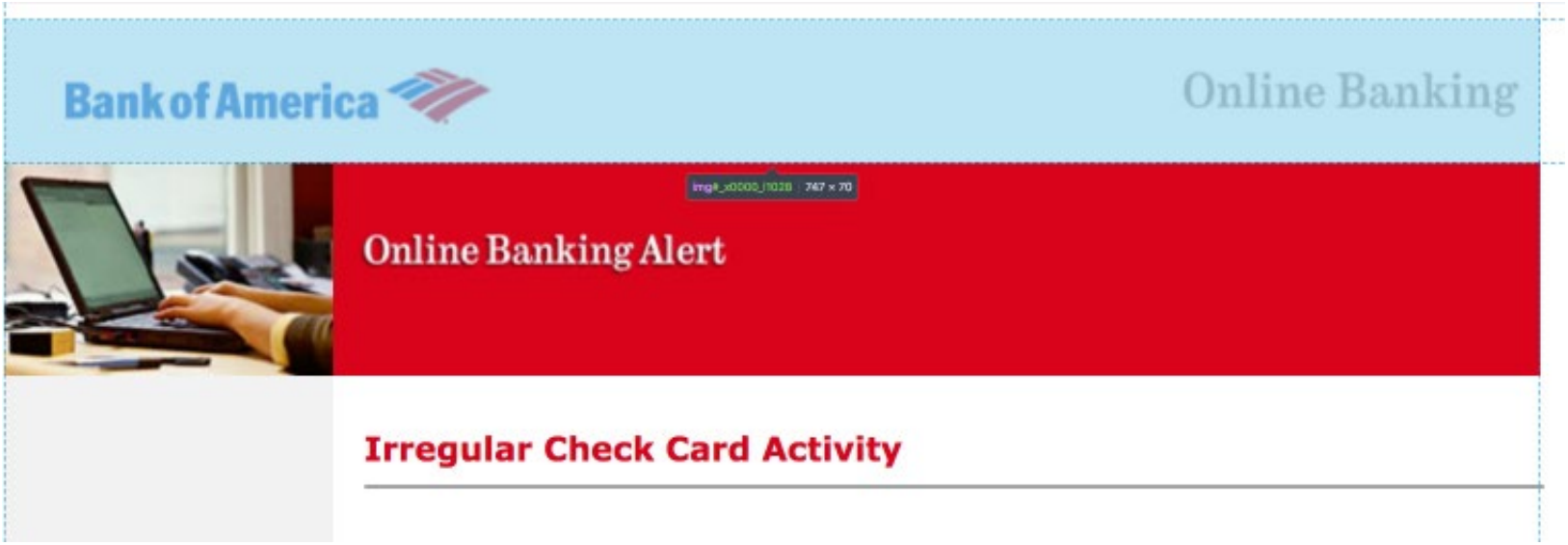
Images in HTML emails are rarely nicely isolated.



Brand Forgery Detection

Challenges

The highlighted blue area is a single image.



Q&A

Inky
Dave Baggett
dave@inky.com