PhishHook: A tool to detect and prevent phishing attacks

Michael Stepp

steppm@cs.arizona.edu

University of Arizona

Introduction

- Common phishing attacks
- Defense strategies
- PhishHook, which implements one of these strategies
- Evaluation of PhishHook

Why do they exist?

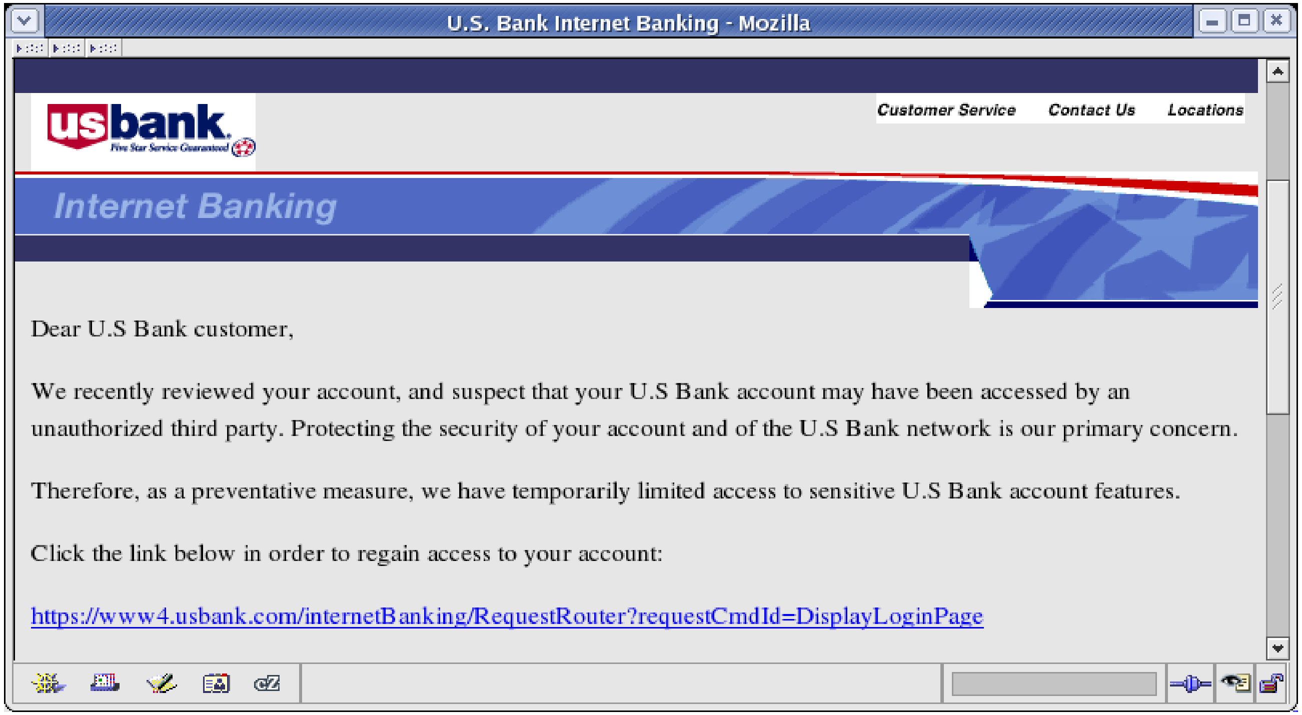
Phishing is an effective way to get a user to reveal his/her personal information:

- Name, address, telephone number
- User ID and password for some secure system
- Social security number
- Credit card number
- Mother’s maiden name
- Other indirect means of accessing user’s information

Why do they work?

Phishing attacks rely on:

- Concealing information
- Presenting misinformation
- Taking advantage of user's trust/gullibility


U.S. Bank Internet Banking - Mozilla
usbank
Five Star Service Guaranteed
Customer Service
Contact Us
Locations
Internet Banking

Dear U.S Bank customer,

We recently reviewed your account, and suspect that your U.S Bank account may have been accessed by an unauthorized third party. Protecting the security of your account and of the U.S Bank network is our primary concern.

Therefore, as a preventative measure, we have temporarily limited access to sensitive U.S Bank account features.

Click the link below in order to regain access to your account:

https://www4.usbank.com/internetBanking/RequestRouter?requestCmdId=DisplayLoginPage

Methods of deceit

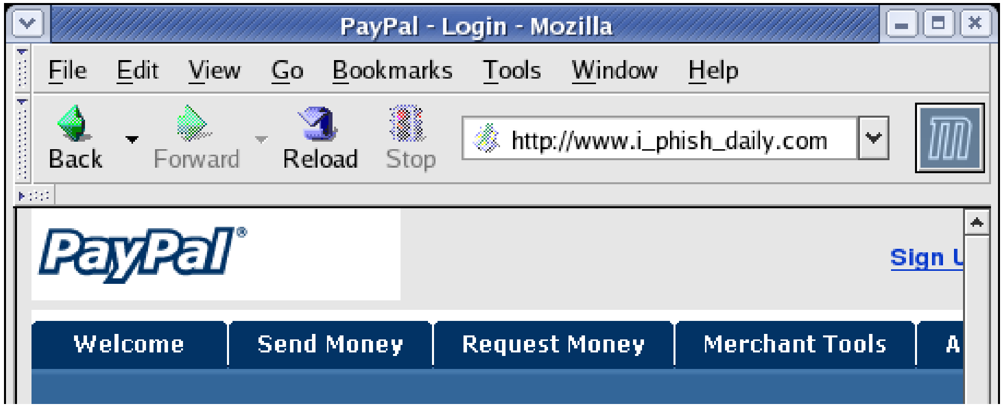
- Using an IP address instead of a domain name

  `http://68.142.197.80/` ≡ `http://www.yahoo.com/`

- Using a domain name that is very similar to a real one

  `http://www.paypa1.com/`

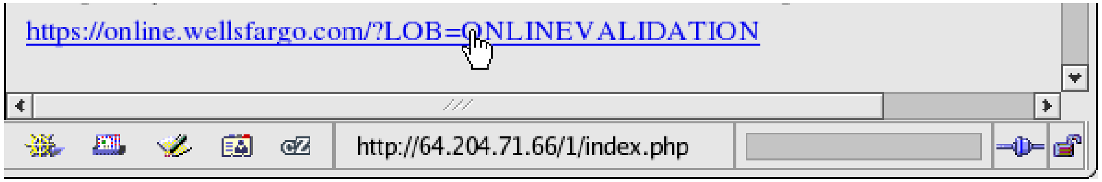
- Copying the appearance of another website



http://www.i_phish_daily.com

Methods of deceit (cont'd)

- Misleading hyperlink text



https://online.wellsfargo.com/?LOB=ONLINEVALIDATION

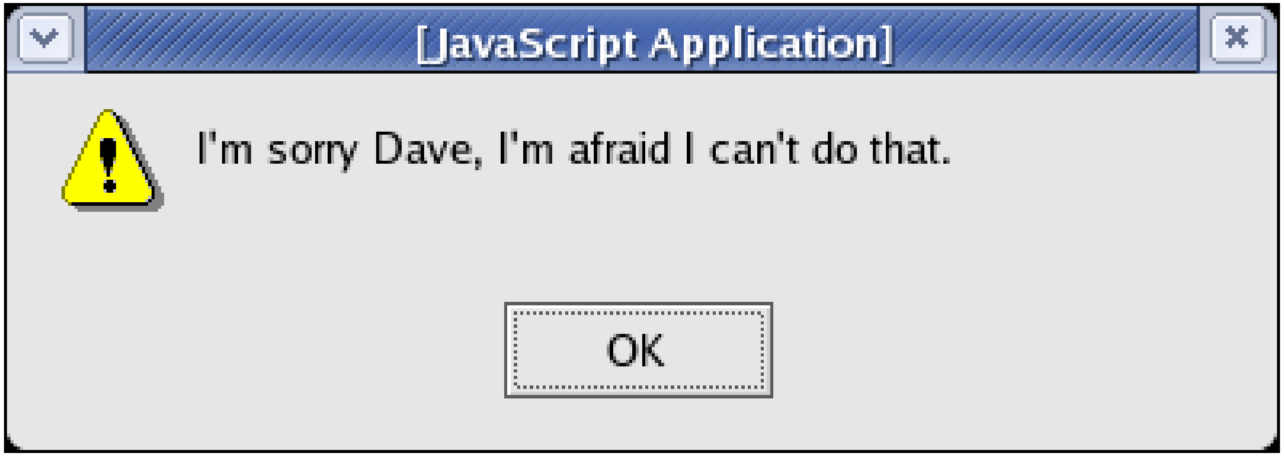
http://64.204.71.66/1/index.php

- Hiding the status bar text
- Using images in lieu of HTML
- Making everything a link

Possible Solutions

Idea #1:

Prevent posting sensitive information on a suspicious website



I'm sorry Dave, I'm afraid I can't do that.

Idea #1

Pros:

- Prevents all possible phishing attacks
- Lets the user know when a site is malicious

Cons:

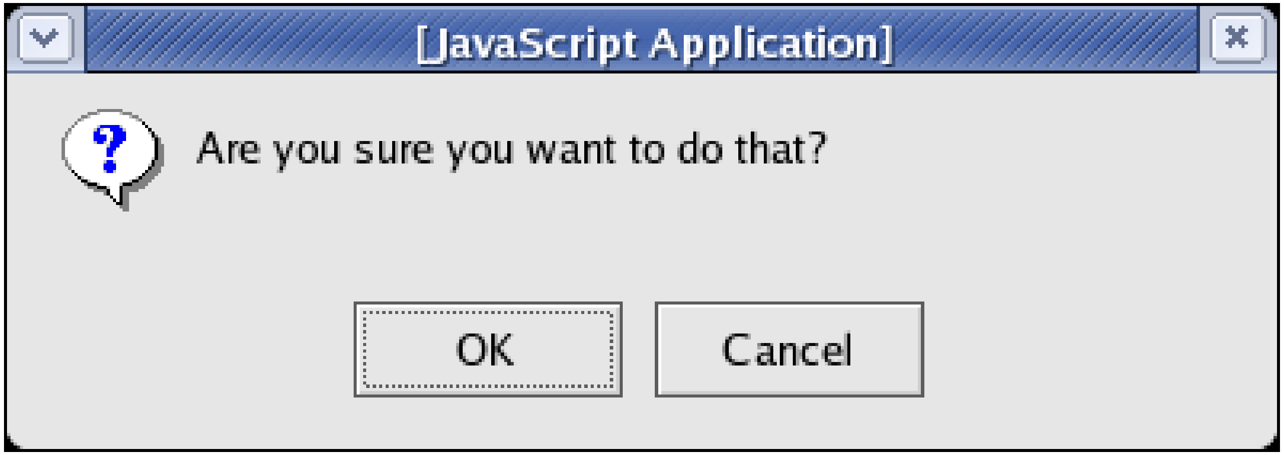
- Relies on the phish detector being 100% accurate
- False positives prevent user from accessing legitimate sites
- False negatives that are still phishy are not reported

Conclusion: BAD IDEA!

Idea #2

Idea #2:

Display warning prompts for all unsafe actions



[JavaScript Application]

Are you sure you want to do that?

Idea #2 (cont'd)

Pros:

- False positives not restricted
- Notifies user of specific dangers on a website

Cons:

- Most actions on a website are unsafe in some way
- The number of prompts would make browsing cumbersome

Conclusion: BAD IDEA!

Conclusion:

- Too aggressive!
- Better solution: passive approach
  - Alert user about dangers
  - Do NOT restrict user’s actions
  - Do NOT force user to acknowledge warnings

PhishHook

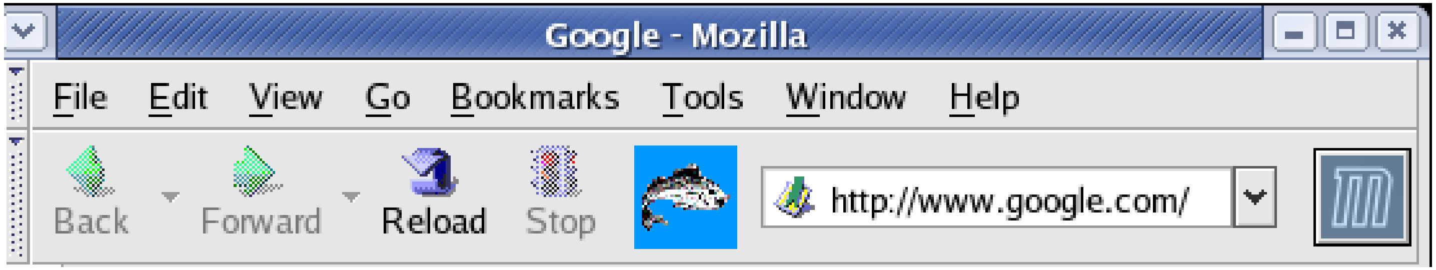
PhishHook: extension to Mozilla web browser

Why Mozilla?

- Setting of most phishing attacks, good place to intercept them
- Provides library of useful functions
- Uses DOM (Document Object Model), represents HTML in a simple tree structure

PhishHook User Interface

- Just one button: the phish button



- Toggles between clean and dirty webpage
- A “clean” page will be converted to “normal form”
- Visualizes possible phishy behavior
- Educates the user about phisiness

Transformations

Text Transformations:

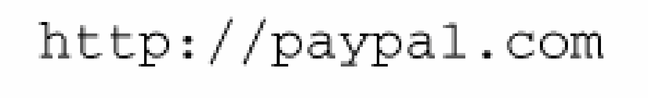
- All text is set to a default font and size
- All background colors ⇒ white
- Text colored by content

normal text		normal text
hyperlink text	⇒	hyperlink text
phishy text		phishy text

Transformations (cont’d)

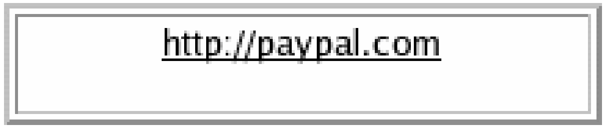
Image Transformations:

- All images processed by OCR library
- Images that contain text will be replaced by the text itself.



http://paypal.com

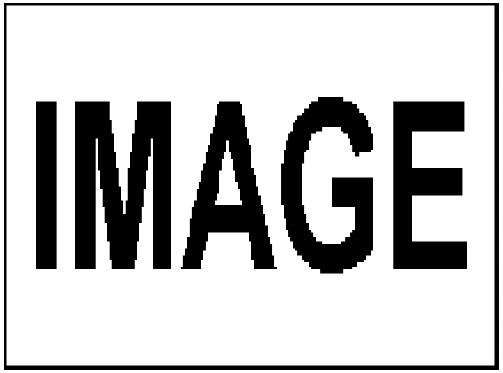
⇒



http://paypal.com

- Others replaced by default image, colored **purple** if inside a hyperlink and **black** otherwise.

⇒

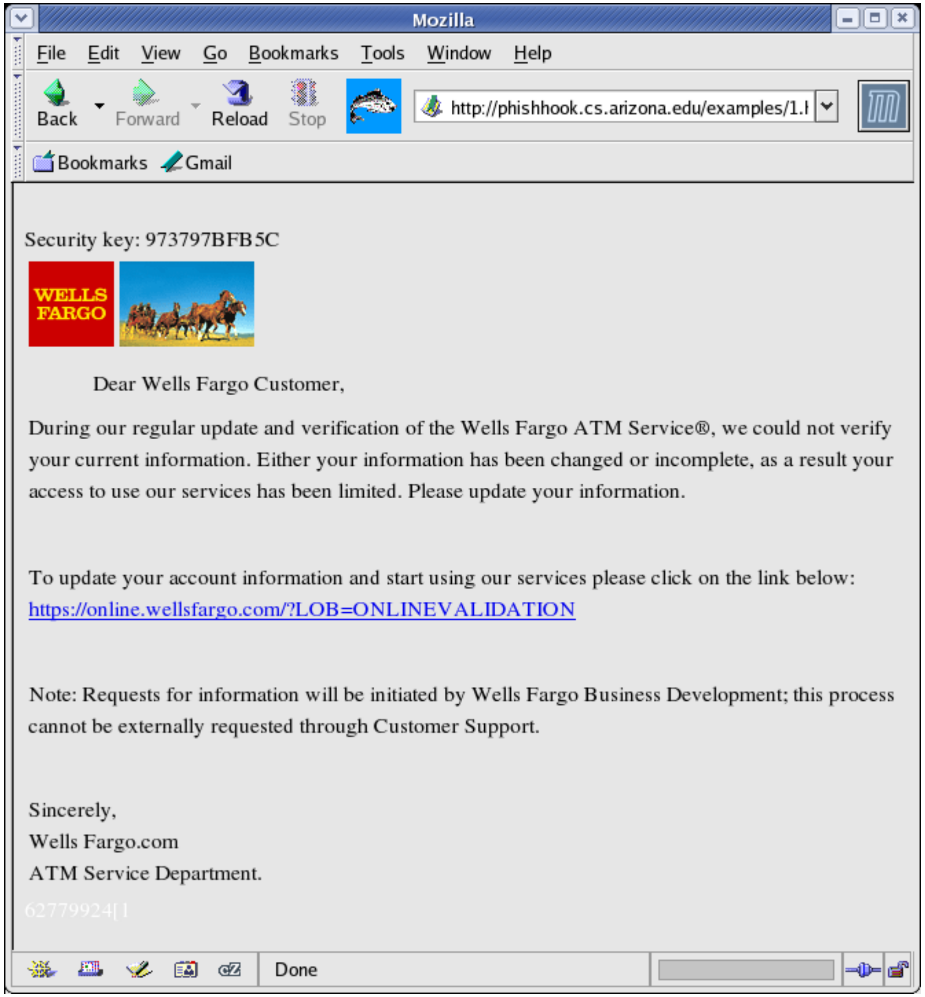


Transformations (cont'd)

Hyperlink Transformations:

- Hyperlink targets compared against their contents:
  - if they do not match, replace text with warning
- If hyperlink target is offsite, highlight it
- If hyperlink target = IP address, highlight it

Example


Mozilla
File Edit View Go Bookmarks Tools Window Help
Back
Forward
Reload
Stop
http://phishhook.cs.arizona.edu/examples/1.l
Bookmarks
Gmail
WELLS FARGO
Done

Security key: 973797BFB5C

Dear Wells Fargo Customer,

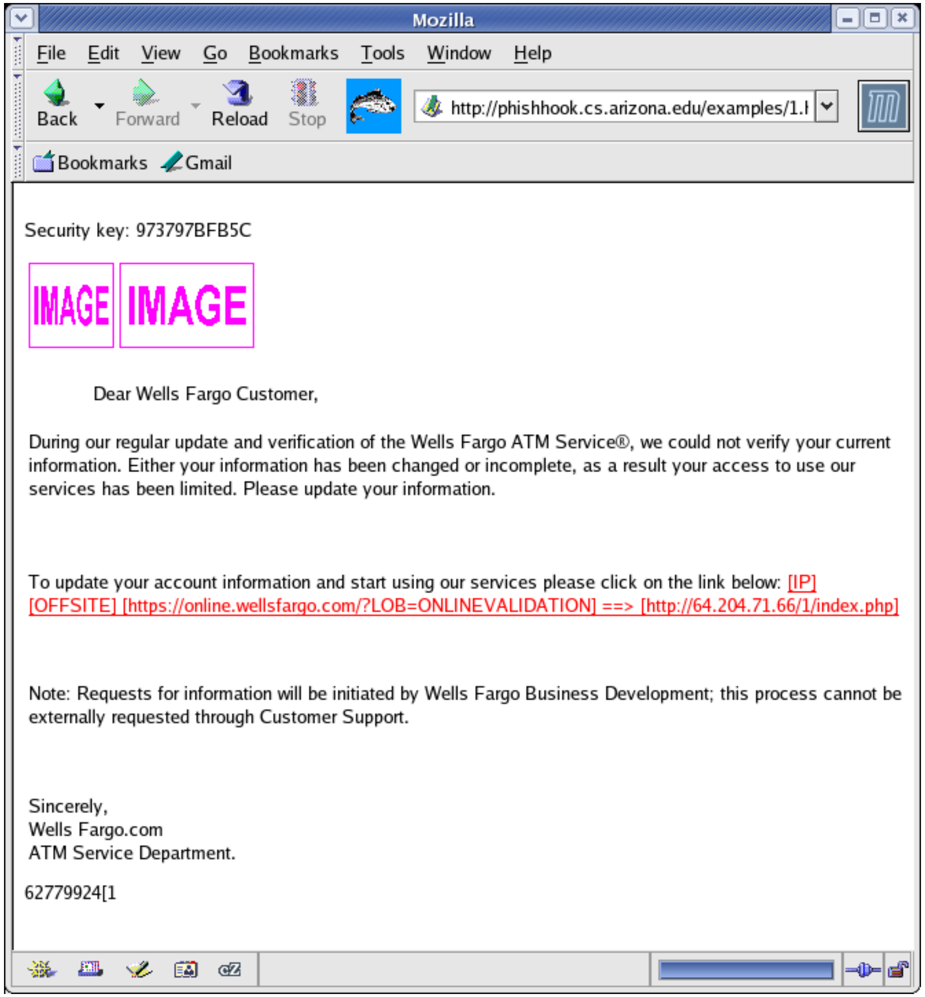
During our regular update and verification of the Wells Fargo ATM Service®, we could not verify your current information. Either your information has been changed or incomplete, as a result your access to use our services has been limited. Please update your information.

To update your account information and start using our services please click on the link below:
https://online.wellsfargo.com/?LOB=ONLINEVALIDATION

Note: Requests for information will be initiated by Wells Fargo Business Development; this process cannot be externally requested through Customer Support.

Sincerely,
Wells Fargo.com
ATM Service Department.

62779924[1


Mozilla
File Edit View Go Bookmarks Tools Window Help
Back
Forward
Reload
Stop
http://phishhook.cs.arizona.edu/examples/1.l
Bookmarks
Gmail
IMAGE
IMAGE

Security key: 973797BFB5C

Dear Wells Fargo Customer,

During our regular update and verification of the Wells Fargo ATM Service®, we could not verify your current information. Either your information has been changed or incomplete, as a result your access to use our services has been limited. Please update your information.

To update your account information and start using our services please click on the link below: [IP] [OFFSITE] [https://online.wellsfargo.com/?LOB=ONLINEVALIDATION] ==> [http://64.204.71.66/1/index.php]

Note: Requests for information will be initiated by Wells Fargo Business Development; this process cannot be externally requested through Customer Support.

Sincerely,
Wells Fargo.com
ATM Service Department.

62779924[1

Effects of PhishHook

We can now examine the effectiveness of PhishHook on the methods of deceit:

- Using an IP address instead of a domain name
  ⇒ Hyperlink transformations
- Copying the appearance of another website
  ⇒ All transformations
- Misleading hyperlink text
  ⇒ Hyperlink transformations

Effects (cont’d)

- Hiding the status bar text
  $\Rightarrow$ Hyperlink transformations
- Using images in lieu of HTML
  $\Rightarrow$ Image transformations
- Making everything a link
  $\Rightarrow$ Color coding: purple $\equiv$ hyperlink

Drawbacks

- Problems with OCR:
  - No good open-source package
  - Most deal with limited cases: i.e. 1-bit color, fixed-width font
  - Anti-aliased fonts
  - Text of different sizes
  - Text on different baselines
  - Special characters: i.e. `http://www.site.com/`
- Result: text-on-image stripped out in most cases

Evaluation

- PhishHook addresses common methods of deceit

- Exposes them in passive way:
  - Only acts when requested by the user
  - Does not restrict actions of the user
  - User free to ignore all warnings if irrelevant
  - User not forced to acknowledge warnings

- Incorporated into established web browser

Future Work

- Address technique of using URLs similar to legitimate ones:
  - Have database of commonly spoofed URLs
  - Compare given URL against database URLs
  - Small edit distance $\Rightarrow$ probable spoofed site
- Add objective “phishiness” rating: tells likelihood that the webpage is malicious
- Similar extension to Thunderbird mail client, to detect phishy emails (in progress)

Related Work

- SpoofGuard
  - Extension to Internet Explorer
  - Evaluates current webpage, indicates risk level with warning light
  - Relies on 5 measurements, done in 2 rounds
  - Overall risk = weighted sum of measurements
  - Caches data from commonly spoofed sites
  - Compares images and URLs to cached versions

Related Work (cont'd)

- PhishGuard
  - Background process, monitors your internet activity
  - Maintains database of known phishy websites
  - When user visits phishy website, warning popup appears
  - User can report new phishy websites, information disseminates to all users

References

- Yuka Teraguchi, Dan Boneh, Neil Chou, Robert Ledesme, and John C. Mitchell. Client-side defense against web-based identify theft. `http://crypto.stanford.edu/SpoofGuard/`
- PhishGuard. `http://www.phishguard.com`
- MailFrontier Phishing IQ Test. `http://survey.mailfrontier.com/survey/quiztest.cgi?themailfrontierphishingiqtest`
- Mozilla/Gecko/XPCOM. `http://www.mozilla.org/`, `http://xulplanet.com/references/xpcomref/`
- Monkey image courtesy of `http://www.cnn.com`.