

PhishMonger: A Free and Open Source Public Archive of Real-World Phishing Websites

David G. Dobolyi & Ahmed Abbasi

Center for Business Analytics

McIntire School of Commerce

University of Virginia

PhishTank


PhishTank is operated by [OpenDNS](#), a free service that makes your Internet safer, faster, and smarter. [Get started today!](#)

PhishTank® Out of the Net, into the Tank.

username *****
[Register](#) | [Forgot Password](#)

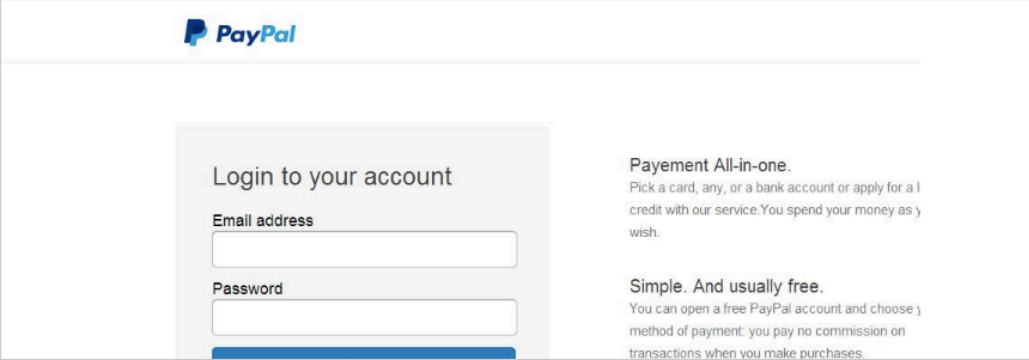
[Home](#) [Add A Phish](#) [Verify A Phish](#) [Phish Search](#) [Stats](#) [FAQ](#) [Developers](#) [Mailing Lists](#) [My Account](#)

Submission #4078151 is currently offline
Submitted May 15th 2016 8:48 PM by [paulch](#) (Current time: Sep 27th 2016 1:50 PM UTC)
<http://pyplwebapps-security.com/webapps/mpp/home/index.php>

 **Verified: Is a phish**
As verified by [SirSpamalot](#) [buava](#) [leofelix](#) [Pluto67](#)

Is a phish **100%**
Is NOT a phish 0%

[Screenshot of site](#) [View site in frame](#) [View technical details](#) [View site in new window](#)



The screenshot shows the PayPal login page. On the left, there is a "Login to your account" section with input fields for "Email address" and "Password". On the right, there is a "Payment All-in-one" section with text: "Pick a card, any, or a bank account or apply for a l credit with our service. You spend your money as y wish." Below that is a "Simple. And usually free." section with text: "You can open a free PayPal account and choose y method of payment. you pay no commission on transactions when you make purchases."

[Friends of PhishTank](#) | [Terms of Use](#) | [Privacy](#) | [Contact](#)
PhishTank is operated by [OpenDNS](#). Learn more about [PhishTank](#) or [OpenDNS](#).

Ephemerality



PhishTank is operated by [OpenDNS](#), a free service that makes your Internet safer, faster, and smarter. [Get started today!](#)

PhishTank® Out of the Net, into the Tank.


username *****
[Register](#) | [Forgot Password](#)

[Home](#) [Add A Phish](#) [Verify A Phish](#) [Phish Search](#) [Stats](#) [FAQ](#) [Developers](#) [Mailing Lists](#) [My Account](#)

Submission #4078151 is currently offline


Submitted May 13th 2016 6:46 PM by [pantia](#) (Current time: Sep 27th 2016 1:50 PM UTC)

<http://pyplwebapps-security.com/webapps/mpp/home/index.php>

 **Verified: Is a phish**
As verified by [SirSpamalot](#) [buava](#) [leofelix](#) [Pluto67](#)

Is a phish **100%**
Is NOT a phish 0%

[Screenshot of site](#) [View site in frame](#) [View technical details](#) [View site in new window](#)



Login to your account

Email address

Password

Payment All-in-one.
Pick a card, any, or a bank account or apply for a l credit with our service. You spend your money as y wish.

Simple. And usually free.
You can open a free PayPal account and choose y method of payment. you pay no commission on transactions when you make purchases.

[Friends of PhishTank](#) | [Terms of Use](#) | [Privacy](#) | [Contact](#)
PhishTank is operated by [OpenDNS](#). Learn more about [PhishTank](#) or [OpenDNS](#).

PhishTank API

- Indexes online, valid phishing sites
- Typically 25,000 to 50,000 sites per request
- Updated hourly
- Data provided in JSON, XML, CSV, serialized PHP formats
- Phish details include:
 - phish_id and url
 - target
 - submission_time and verification_time
- http://www.phishtank.com/developer_info.php

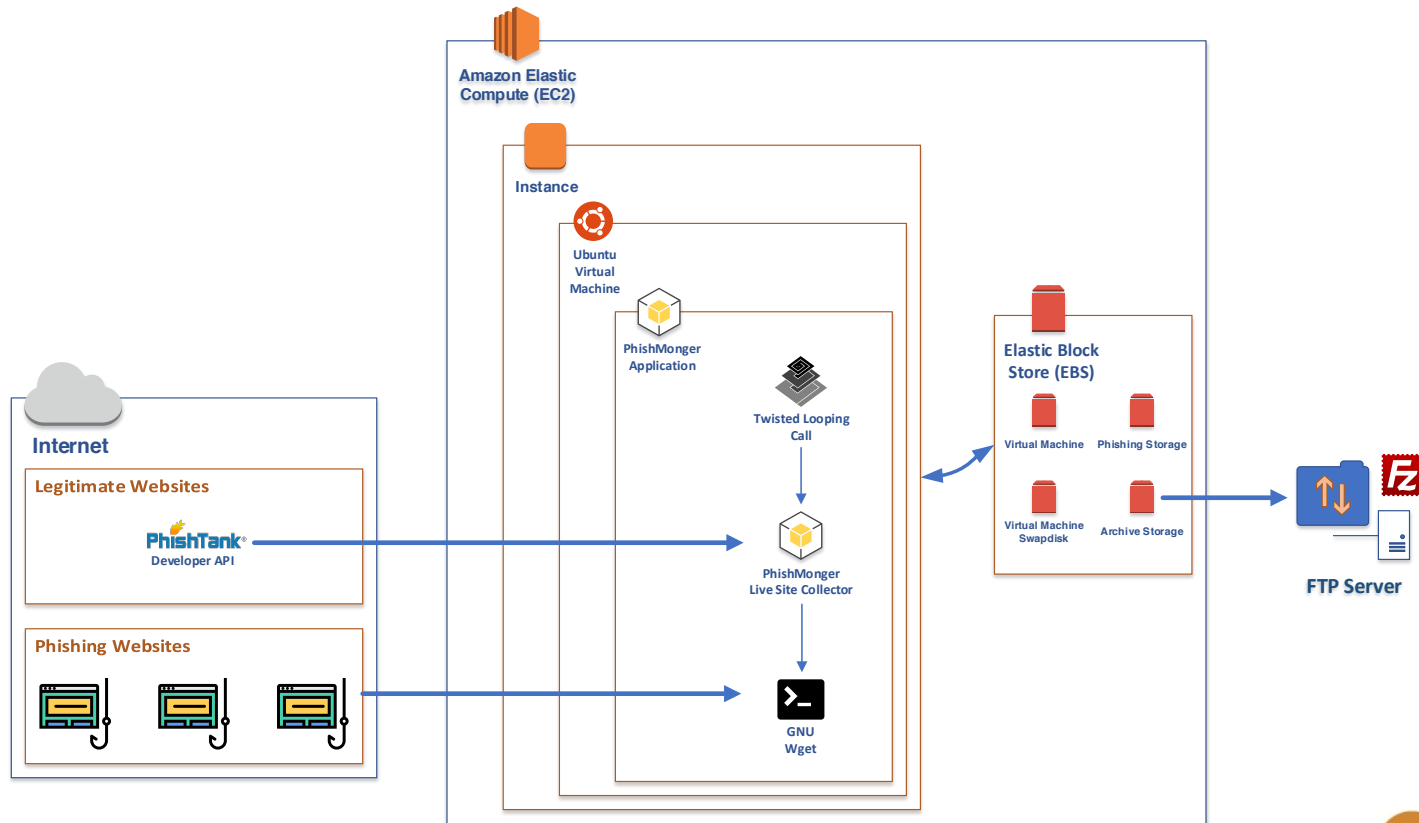
PhishMonger

- Invokes the PhishTank API hourly
- Identifies newly added phish URLs
- Fetches the new phishing websites
- Adds the captured sites to a growing corpus

About the Platform

- Leverages exclusively open-source software:
 - Ubuntu Linux, GNU Wget, Filezilla Server FTP
- Coded in Python 3.5
 - Harnesses the Twisted library for time-based scheduling
- Runs on Amazon Web Services (AWS) Elastic Compute Cloud (EC2)
- Additional statistical scripts written in *R*

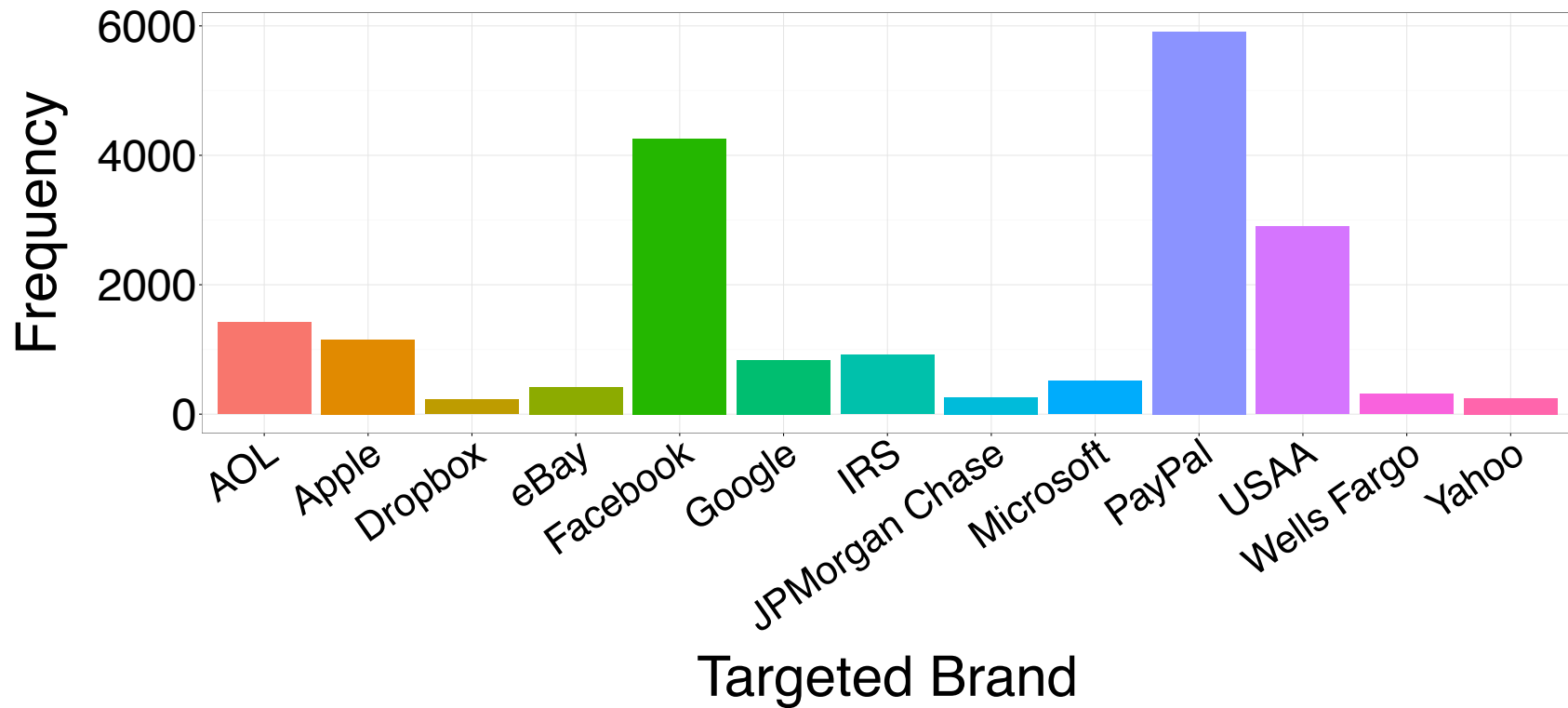
System Diagram



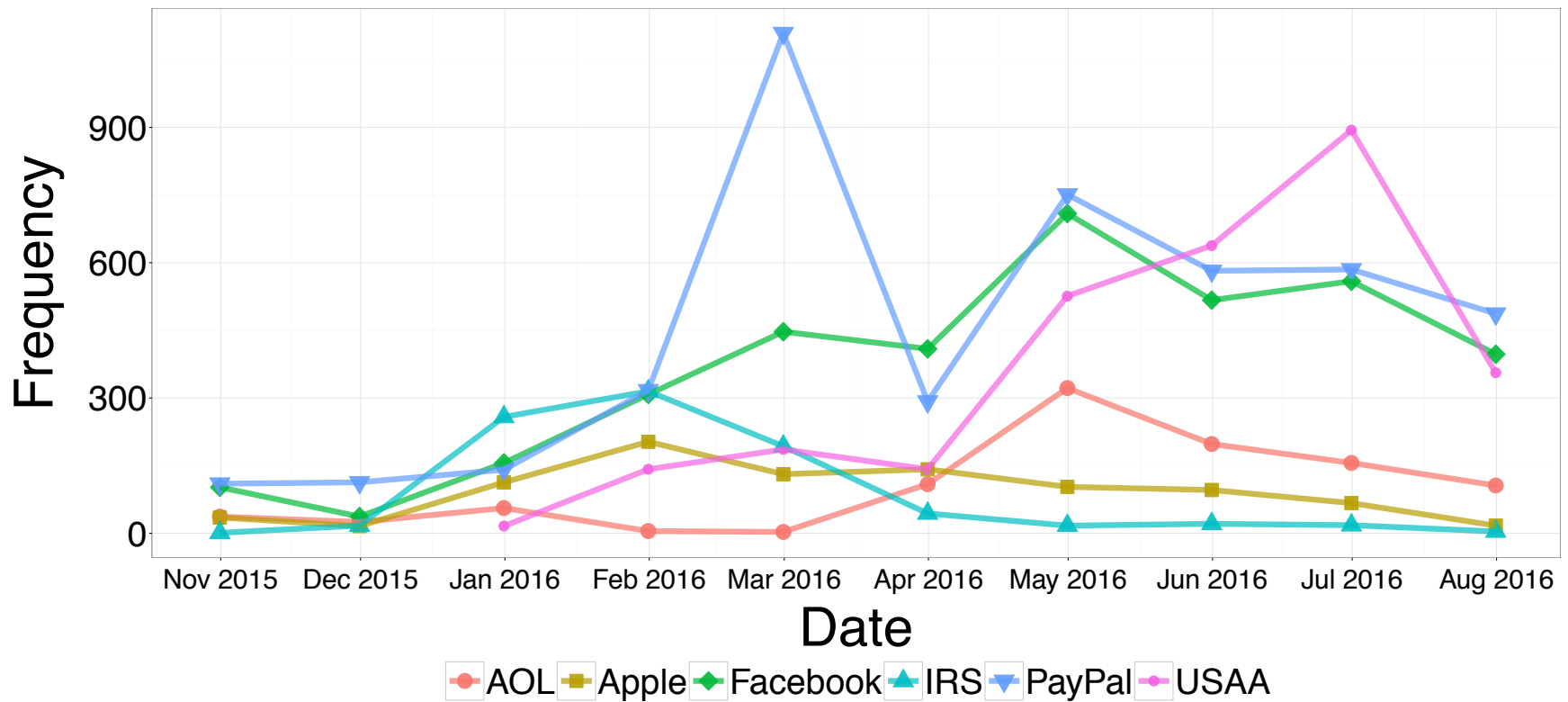
PhishMonger Corpus

- As of **September 23, 2016** and run **3649**, the collection includes:
 - **171,360** sites
 - **129** targeted brands
 - **19,690,341** files and folders
 - **200GB** of compressed storage

Targeted Brands with 200+ Sites



Top 6 Targeted Brands Over Time



Top 10 File Types in Corpus

File Extension	Description	Category	n
png	Portable Network Graphics	Graphics	3,169,772
html	HyperText Markup Language	Text	1,304,574
jpg	Joint Photographic Experts Group	Graphics	1,251,227
gif	Graphics Interchange Format	Graphics	1,208,424
js	JavaScript Code	Text	947,420
css	Cascading Style Sheet	Text	776,404
tff	True Type Font	Font	210,197
svg	Scalable Vector Graphics	Graphics	176,856
ico	Icon	Graphics	139,564
woff	Web Open Font Format	Font	134,308

Get the Data

- Available at University of Arizona Artificial Intelligence Lab:
<http://www.azsecure-data.org/phishing-websites.html>
- Contact:
 - David G. Dobolyi, dd2es@comm.virginia.edu
 - Ahmed Abbasi, abbasi@comm.virginia.edu
- Acknowledgements:
 - **US National Science Foundation ACI-1443019** “DIBBs for Intelligence and Security Informatics Research Community”
 - Data Infrastructure Buildings Blocks (DIBBs) partner institutions: **University of Arizona’s AI Lab** and collaborators at **Drexel University**, the **University of Texas-Dallas**, and the **University of Utah**



Miscellaneous Slides

Number of Targeted Brands and Percentage of Verified Phish in PhishTank Database

