



Phisherman

Real-Time Phishing Data Collection, Validation, Dissemination, and Archival

In its Phisherman project, SPARTA's Phisherman project team is developing a real-time phishing data collection, validation, dissemination, and archival system. The primary Phisherman objectives are:

- Rapidly provide reliable data from on-going phishing attacks to first responders and brand owners to reduce the impact of phishing attacks.
- Enable law enforcement to identify patterns in phishing attacks that may lead to successful prosecution, and
- Provide a comprehensive data resource for researchers to enable revolutionary advances in anti-phishing technologies.

Phishing attacks typically involve high volumes of socially-engineered emails spammed to thousands of users in a short period of time, enticing them to visit web sites that may migrate within hours of their first appearance. Phishing is

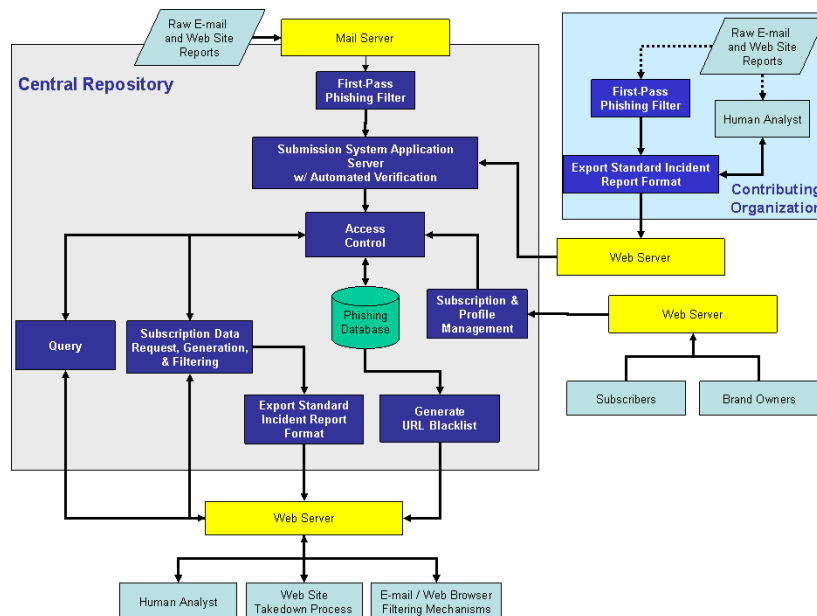
a global problem in which a single attack may access resources in multiple countries. The high volume of sophisticated attacks, combined with dynamic attack evolution, dictates that phishing responses be highly automated, extremely fast, and highly accurate.

The Phisherman project team is working with the Anti-Phishing Working Group (APWG) to enlist the support of several industry partners already actively involved in phishing data collection. Our researchers are collecting data from new sources and soliciting participation from existing sources to create a global data collection system.

The figure below shows the processing components of the Phisherman system.

Phisherman collects incident reports from multiple sources, including raw spam feeds from email services, validated reports from anti-phishing services, and the general public.

Phisherman will record all of the direct artifacts of phishing attacks, as well as related information





Phisherman

Real-Time Phishing Data Collection, Validation, Dissemination, and Archival

from external sources. The direct artifacts include email lures and URLs of phishing web sites and domains as submitted by contributors. The related information includes domain name registration data, phishing web site contents, SSL certificate attributes, malware, targeted brand names, and country of origin for the web sites. Phisherman automatically collects the related information upon receipt of a new phishing incident report. All of the direct artifacts and related information are stored in a relational database of phishing incident reports.

Phisherman uses the contents of the incident reports, heuristics derived from past experience, whitelists of known-legitimate sites, and historical attack information from the database to assess the likelihood that a new incident report is in fact phishing. Phisherman assigns a numerical confidence rating to each report and publishes the results to subscribers.

As an investigative tool, Phisherman will provide a reliable, searchable database of phishing incident reports, including e-mail content, web site content and attributes, and malware, as well as analysis and historical records of attacks. As new incidents are posted to the repository, the system will identify similarities between the new incident report and existing reports. By recording data on nearly identical attacks, the repository will help identify the scope of the attack, relationships between attacks across brands, and potential victims.

Phisherman will also provide several mechanisms for disseminating data to participating organizations:

- Subscription-based incident reports: Subscriptions are intended to meet the need for regular data delivery for incident response organizations and brand

owners. The subscription data adheres to the IODEF format being standardized by the IETF.

- URL Blacklist: The APWG currently provides this service to organizations operating spam filtering services and other anti-phishing systems to enable rapid blocking of email and web traffic to confirmed phishing sites. Phisherman is implementing a similar service.
- Generalized incident report queries: The query capability allows law enforcement, brand owners, and analysts to retrieve individual or multiple incident reports matching the search criteria.

Phisherman provides data to improve existing anti-phishing technology, enable development of new technologies, and support law enforcement in their efforts to apprehend those responsible for the attacks. Over time, other uses for a comprehensive, searchable database of phishing attack data are likely to emerge.

The Phisherman project team includes SPARTA, the Anti-Phishing Working Group, Internet Compliance Systems, Internet Identity, and Southern Methodist University. This work was supported by the United States Department of Homeland Security.

For further information on how you can support the Phisherman project or become a user, please contact:

Gregg Tally, SPARTA, Inc.

Email: Gregg.Tally@sparta.com

Phone: 443-430-8000