Journal of Digital Forensic Practice, 1:115–129, 2006

ISSN: 1556-7281 print / 1556-7346 online
DOI: 10.1080/15567280601015564

Taylor & Francis
Taylor & Francis Group

ARTICLE

The Phisherman Project: Creating a Comprehensive Data Collection to Combat Phishing Attacks

Gregg Tally
SPARTA, Inc., 7110 Samuel Morse Dr., Columbia, Maryland 21046, USA

David Sames
SPARTA, Inc., 7110 Samuel Morse Dr., Columbia, Maryland 21046, USA

Thomas Chen
Southern Methodist University P.O. Box 750338 Dallas, Texas 75275-0338, USA

Chris Colleran
Internet Compliance Systems, LLC, 1750 112th Ave., NE, Suite D-155, Bellevue, Washington, 98004, USA

David Jevans
Anti-Phishing Working Group 5150 El Camino Real, Suite A20 Los Altos, California 94022, USA

Kevin Omiliak
Internet Compliance Systems, LLC, 1750 112th Ave., NE, Suite D-155, Bellevue, Washington 98004, USA

Rod Rasmussen
Internet Identity Box 1295 Tacoma, Washington 98402, USA

ABSTRACT The Phisherman project is developing a real-time phishing data collection, validation, dissemination, and archival system. The objectives are to rapidly provide reliable data from on-going attacks to first responders and brand owners and to collect comprehensive data for researchers and law enforcement. Working with the Anti-Phishing Working Group, the project is enlisting the support of several industry partners already actively involved in phishing data collection. There are several proprietary and specialized data collection efforts underway. The Phisherman project is working to merge the existing data collections into a global data collection system.

As an investigative tool, Phisherman will provide a reliable, searchable database of phishing incident reports, including email, data collection sites, and malware, as well as analysis and historical records of attacks. All data entering the Phisherman system will pass through two stages of automated verification to provide a confidence rating on the incident report. As new incidents are posted to the repository, the system will identify similarities between the new incident report and existing reports. A relational database will link potentially related attacks through common characteristics of attack methods. By recording data on nearly identical attacks, the repository will help identify the scope of the attack and potential victims.

KEYWORDS phishing, identity theft, IODEF, malware, incident response

INTRODUCTION

Phishing is a widespread and rapidly evolving form of electronic identity theft that attempts to gather personally identifying information from unwitting victims, often through social engineering. The "classic" phishing attack starts with a "lure" (unsolicited email) that directs recipients to a fraudulent web site that appears to be that of a well-known company, government agency, or other organization. The lures and web sites often bear a very close resemblance to those from the legitimate organizations. The attacks tarnish the public image of the hijacked brand (financial institutions, online sellers, government

agencies, and other organizations) and reduce the public trust in online commerce. "Phishers" attempt to evade tracking by law enforcement by rapidly relocating their email servers and Web servers. Phishing web site domain names are typically registered shortly before an attack, and fraudulent web sites stay online for only five days on average.[1] Speed is of the essence for any response to be effective. Anti-phishing technologies must be constantly updated to meet the changing nature of the threat.

Phishing attack techniques constantly evolve as the public becomes more educated about the threat and as new defensive tools emerge. Early phishing attacks were relatively untargeted, hoping for overall success by targeting millions of potential victims with a small probability of a response from any single recipient. A 1% response rate on one million emails yields 10,000 responses at very low cost to the phisher, each one providing information that could be worth tens or hundreds of dollars to the phisher when the information is sold. While the untargeted attacks are still common, new forms of personalized or targeted attacks are starting to emerge in which the phisher customizes the lure for each intended victim, trading reduced volume for a higher response rate.

Phishing data are collected by numerous organizations, resulting in many disjoint collections of varying quality and content. Consumers often report phishing incidents to the companies involved, generally via email forwarding. Additionally, because phishing involves the crime of identity theft, incidents are also reported to government organizations such as the U.S. Internet Crime Complaint Center (www.ic3.gov) and U.S. Federal Trade Commission (www.consumer.gov/idtheft). However, the current state of tracking phishing attacks is somewhat fragmented. There is no single resource available that provides ready access to on-going and historical phishing attacks for first-responders, brand owners, researchers, and law enforcement. The goal of the Phisherman project is to create that single information resource.

There are several business and technical challenges to creating the Phisherman data repository. One business challenge is to preserve the revenue models that fund the data collection efforts. Some companies that identify phishing web sites are paid by brand owners. Other companies are paid by the consumers that are protected from the attacks. Each has different concerns regarding privacy and the time value of the data. On the technical side, the challenges include rapid automated validation to cope with the increasing scale of the attacks, and privacy protection, particularly when attacks are targeted at specific individuals. The Phisherman project is working to address all of these issues.

Phisherman's ultimate success depends on support and participation from the global anti-phishing community. To ensure that success, the project team is working closely with prospective participants to ensure that Phisherman's data collection and dissemination models meet their technical and business requirements. Project members also support existing phishing data collection and repository efforts and are applying that experience to the Phisherman design and implementation. Phisherman is a "work in progress," with the initial implementation currently in development. The goal of the first implementation is to demonstrate the effectiveness of the Phisherman architecture and gain the support of anti-phishing organizations.

This article describes the emerging Phisherman system. It discusses other phishing data collection efforts, provides an overview of the Phisherman system design, describes the solutions being implemented for the key technical and business challenges, discusses the potential application of the Phisherman repository to law enforcement, and identifies potential future enhancements. Contact information for the project is available at the end of the article.

RELATED WORK

The idea for the Phisherman project originated in the Anti-Phishing Working Group. The APWG is an association of industrial, research, and law enforcement organizations dedicated to eliminating fraud and identity theft caused by phishing, pharming, and email spoofing. In 2003, several member companies of the APWG found that they shared similar needs for real-time information on in-progress phishing attacks in order to update signatures in anti-phishing products. These companies started to make bilateral relationships to share data among themselves. As the usefulness of this information became evident, the APWG formed a Phishing Repository, Data Streams, and Alerts subgroup to create "the policies and procedures around sharing and using phishing repository data."[2] The APWG currently collects phishing attack reports, archives them in a relational database, and provides data to APWG members. The group is also working to establish data

interchange format standards through the Internet Engineering Task Force (IETF). Since the APWG is primarily a volunteer organization, resources for development and operation of the repository are limited.

The APWG phishing repository is one of several currently in operation. The APWG system takes reports of phishing from the public, anti-phishing projects, brand owners, and Internet service providers (ISPs). The reports are typically phishing email messages that are forwarded by consumers. Several other working groups send raw phishing emails to the APWG including the SANS Institute, various incident response organizations, APACS (representing the UK financial services industry), Adbusters, PIRT and others. In addition to the APWG repository project, other groups conduct phishing data collection activities, each with its own objectives and participation requirements. These data collection projects include:

- **Digital PhishNet (DPN).**[3] The DPN project is an effort spearheaded by Microsoft Corporation to provide law enforcement agencies with a view into the phishing world. The essence of the system is a searchable database of phishing URLs and related information. The export format is compatible with commercial spreadsheets. Several law enforcement members have been trained on the system; however, it does not appear to have many data sharing relationships, so it is not clear how effective the database is, in supporting effective law enforcement activities. Microsoft is a member of the APWG.
- **Phishing Incident Response Team (PIRT).**[4] PIRT is a virtual volunteer organization based in the United Kingdom. This group operates in a way similar to that of the SANS incident handlers. People from around the world process phishing reports and verify the sites. PIRT members also try to work with ISPs to have sites taken down. PIRT shares data with the APWG.
- **Phish Report Network (PRN).**[5] The PRN is a commercial data sharing network operated by Symantec, Inc. PRN takes data feeds from several data sources and sells the data to subscribers. The PRN Data Provider Agreement[6] places the burden of verifying the phishing incident data on the data provider. PRN is a commercial offering and as of this time does not share data with other systems.
- **National-Level Incident Response Teams.** The various incident response teams around the world collect phishing reports and endeavor to have the sites taken down by ISPs in the countries where the response teams operate. Some of the response teams exchange data with the APWG.
- **Cloudmark**[7] has created a network of users to collect, verify, and submit reports of phishing activity to support its anti-phishing products. Cloudmark is a for-profit company and member of the APWG.

The Phisherman project was formed to expand and improve data collection and distribution for the entire anti-phishing community, including "first responders," brand owners, researchers, and law enforcement. Phisherman seeks to unite existing data collection efforts so that all members of the anti-phishing community and their customers benefit. First responders include organizations providing spam filtering and web site takedown services. Rapid data delivery will minimize the damage from an attack by reducing the number of people who access a phishing email or web site. Brand owners will benefit through a reduced impact of the attacks on their customers. Likewise, researchers and analysts will be able to use the comprehensive data collection as a resource for analyzing the evolving nature of phishing attacks and for testing new defense mechanisms. Finally, law enforcement is expected to be one of the major beneficiaries from the Phisherman project. The database will identify common elements across multiple attacks, such as email and web site content, malware, email servers, and web servers. We expect that it will be possible to identify the evolution of a series of attacks as the attackers shift their methods but reuse prior attack components.

OVERVIEW OF THE SYSTEM

The Phisherman processing overview is shown in Figure 1. At the highest level, the system consists of three major processes: data inputs into the repository; storage of phishing data in the repository; and outputs from the repository to its users. Each of the processes consists of several elements.

The majority of the Phisherman system is dedicated to receiving and processing incident reports. Phisherman will receive input from a number of public and private sources, including raw emails, suspected URLs, and verified incident reports. First-pass verification will sift out the potential phishing-related

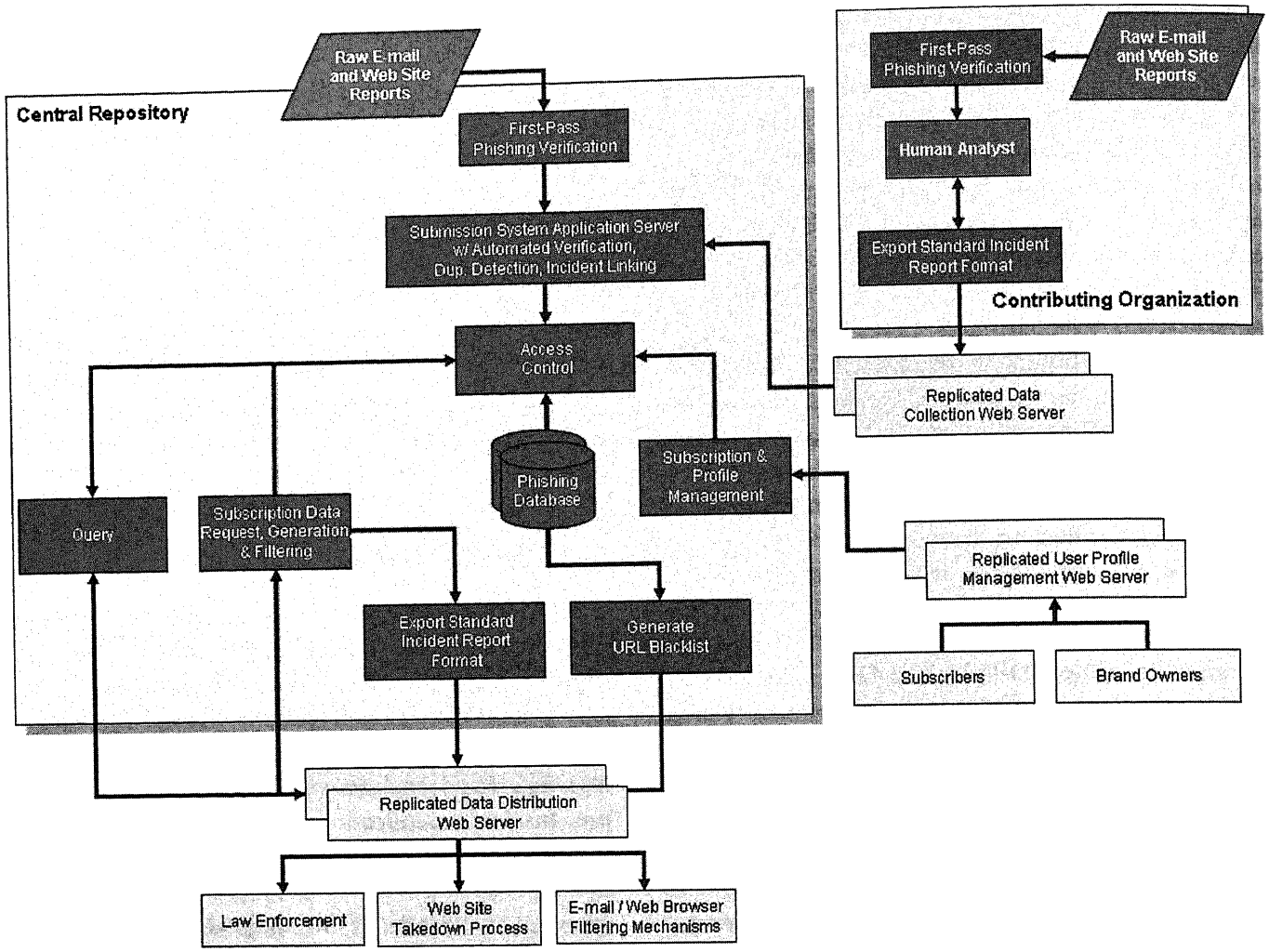

FIGURE 1 **Phisherman Processing Overview.**

data from data that is highly unlikely to be phishing. After verification, the email or URL report will be converted to a standard incident report format for secondary verification and posting to the database. Secondary verification will provide a confidence rating that an incident report is in fact phishing. The final step of receiving an incident report is to post it to the database.

Phisherman will also provide several mechanisms for disseminating data to participating organizations. Phisherman will provide a web-based user interface for subscription-based incident reports and responses to queries. Subscriptions are intended to meet the need for regular data delivery for incident response organizations and brand owners. A generalized query capability will allow law enforcement, brand owners, and other analysts to retrieve individual or multiple incident reports matching the search criteria. Subscriptions and queries are the primary means of disseminating data to Phisherman users.

Repository Inputs

Phisherman will collect incident report data from multiple sources through both web forms and email, but will convert all of the inputs to a standard incident report format. First, Phisherman will receive high-volume spam feeds from organizations (e.g., ISPs and businesses). Second, commercial organizations continually search for fraudulent web sites by looking at domain name registries for possible brand infringement and crawl the web for fraudulent Web sites. The commercial organizations may submit the domain names through email or a Web form. Third, phishing incidents or phishing email will be reported by individuals or organizations that encounter fraud incidents. The data

may be submitted by email or a web form. Phisherman will record information on the source of the data in the associated incident report. The incident report source is one factor in determining the releasability of the data. Incident reports from brand owners or their delegates may be under greater restrictions with regard to disclosure of the targeted brand name than data from other sources. Similarly, ISPs may require restrictions on the targeted victim of an email as a condition of receiving the spam feed. Phisherman will provide unique software adaptors to collect the data from the various sources and convert the input to a standardized format for further processing.

Input from public sources, such as email reports from the spam feeds or by individuals, will be verified twice. The first-pass verification process is unique to each source and input type (email, URL, or other future type) and is intended to separate the obvious non-phishing inputs from the potentially valid phishing reports. The details of the first-pass verification process are described later in this article. Phisherman will perform more time-consuming, higher reliability secondary verification on those reports that pass the first-pass verification process. The first-pass verification prevents the system from being overwhelmed with "junk" reports while the secondary verification process attempts to ensure a very low overall false positive rate.

The following sections describe the primary data sources from high-volume spam feeds, domain name and URL submissions, and individual incident reports.

High-Volume Spam Feeds

Phisherman will receive several spam feeds from both watchdog groups and the mail-filtering systems of large ISPs. The feeds will be delivered to the Phisherman project through email forwarding. The Phisherman project will provide several email accounts around the Internet to which the feeds may be directed. The delivery location of the email accounts may be specified by the spam feed originator or, more likely, will be left up to the project to determine. The primary requirements from the spam feed provider perspective are that the delivery locations remain constant and that Phisherman never reject mail or generate "bounce" messages back to the source of the feed. Phisherman will monitor inbound servers to ensure that capacity limitations are not exceeded and will not bounce messages.

The concept of identifying phishing emails in streams of spam email is not a new one. Phisherman project team members currently receive the spam feeds via email forwarding. The biggest technical challenge to date has been distinguishing the original email after it has been forwarded multiple times, and sometimes altered through edits to the original email. The team has developed an adaptor designed specifically for forwarded email. If other intake sources are required, the project will develop additional intake adaptors to accept the email.

A long-term project goal of is to be able to receive one gigabyte of email per day per email source. By distributing the delivery across the Internet we will be able to avoid any bottlenecks for the large amount of email that we receive from any one source. Distribution and replication of servers should allow the project to meet the throughput goals.

Domain Names Searches

An important source for early identification of Phishing attacks is domain name registrations. Contributors to the Phisherman project currently monitor the registration of new domains under the generic top level domains (gTLDs) and scan them for potential Phishing sites. Those domain names that trip the contributors' scanning filters will be submitted to Phisherman via email for further verification. Generally, the contributors will run their own verification process prior to submitting the URL or domain name to the Phisherman system. In such cases, the contributor may submit a more complete incident report containing the suspect URL or domain name.

Individual Incident Reports

Email and web submissions provide low-volume inputs to the Phisherman repository that enable individuals, corporations, and other organizations to submit phishing attacks. Individual incident reports will be sent through a similar ingest process as the submissions from the high-volume spam feed. The ability to accept incident reports via email from the general public allows Phisherman to collect incident reports from around the world, but also opens the system to potential denial of service attacks through flooding. To minimize this risk, the first-pass verification process will be able to operate independently of the repository,

relying only on information in the raw incident report. This approach allows the system to handle increased input volumes through replication of the email processing servers.

Posting Reports to the Repository

The process of posting incident reports to the repository will include first-pass verification, conversion of raw artifacts (email, URL, or domain names) to a standard incident report format, secondary verification of the incident report, and linking the new incident report to others already in the database.

Phisherman will create reliable, complete incident reports using data from unknown sources that will have inconsistent format and quality, and some of which will be intentionally false. Phisherman will perform a number of processing steps on input data with the goal of discarding non-phishing inputs as early in the process as possible to maximize potential throughput. The system will perform first-pass verification on all submissions of raw email (spam feed or individual reports) to separate potential phish from generic spam and other non-phishing email. Similarly, first-pass verification on URL submissions will detect known-good and known-bad URLs and discard those reports. All raw reports that pass first-pass verification will be converted to Incident Object Description and Exchange Format (IODEF)[8] for further processing, including secondary verification.

IODEF is an emerging standard developed within the Incident Handling (INCH) Working Group[9] of the IETF and is the standard representation of an incident report in the Phisherman project. The IODEF format is based on XML and supports extensions for additional incident data specific to classes of incidents, including the Phraud Extensions,[10] which are also being standardized within the IETF. IODEF and the Phraud Extensions are described in more detail later. By using the standard IODEF formats, the Phisherman project expects to improve participation from the international anti-phishing community.

All incident reports are automatically evaluated in a second-pass verification process to minimize false positives and assign a confidence factor to the phishing identification. In existing repository projects, verification is usually a manual process, but an important feature of Phisherman is the automation of this process. The goal is to ensure that data accepted into the repository is highly accurate without the delays caused by human reviews. Verification also protects the system from deliberate attacks from phishers to poison the repository with misinformation. Automated secondary verification is a key distinguishing feature of the Phisherman architecture from existing data collection systems.

In addition to these verification and conversion steps, Phisherman will reduce data volumes and support incident linking through duplicate checking of both complete reports and components of the incident reports.

First-Pass Verification

First-pass verification will be performed on individual incident reports and high-volume spam feeds submitted to the Phisherman repository. The process is unique to each form of input, currently either email, a domain name, or a URL. First-pass verification will separate potential phishing inputs from inputs that are highly unlikely to be phishing, greatly reducing the data input and processing loads on the Phisherman system.

The first-pass email verification implementation will rely on work by the Institute for Security Technology Studies (ISTS) at Dartmouth University in their project on Automated Analysis of Spam-Vectored Phishware. The technique uses Bayesian filtering which is the same technique that is effectively used in commercial anti-spam mailbox tools.[11]

The Phisherman implementation of the Bayesian filtering will vary slightly from the Dartmouth implementation in that Phisherman requires a very low false negative rate on first-pass verification and can tolerate a limited number of false positives. False negatives (verification that erroneously concludes that an incident report *is not* phishing) causes incident reports to be dropped that should be recorded in the repository. There is no recovery from a false negative unless and until the incident is reported by another source. False positives (verification that erroneously concludes that an incident report *is* phishing) are less serious in the first-pass verification phase since secondary verification will re-examine the incident report and should discard the false-positive incident reports. The Dartmouth implementation uses the "bogofilter" that has a very low false-positive rate, moderate false-negative rate, and is highly efficient. Phisherman is investigating other Bayesian filters, including CRM114 that is reported to have more compatible false-positive and false-negative rates.

Conversion to IODEF Format, Decomposition, Privacy Protection

Phisherman supports the IODEF format for incident report submission, internal processing, and incident report dissemination. The IODEF format is a flexible, XML-based reporting mechanism standardized by the IETF to facilitate data sharing between different organizations (such as law enforcement, incident response teams, and ISPs). Importantly, IODEF has been adopted by some national incident response organizations, and the format is extensible to include reporting of malware, spam, e-crime, and other criminal incidents. The IODEF format reduces the need for customized incident report formats both internally to Phisherman and externally.

Phisherman uses the IODEF format internally to simplify and improve the secondary verification process and other phases of the incident report submission process. This standardized XML format allows all incident reports to be handled in the same way, regardless of the source of the report or the type of report (suspicious email, domain name, URL, or other future extensions). Phisherman will convert the initial incident report to IODEF and then attempt to collect additional information related to the report to enhance the incident report. The additional information includes associated Web page content, and domain name and IP address registration information. Once converted to IODEF format, all incident reports receive the same processing.

The data conversion process takes a raw incident report input, such as an email or URL, as the initial input. Emails will be decomposed into their component parts, including the header, subject, and body, and stored in an IODEF record. URLs and domain names will also stored in the IODEF record. The email, web site content, and other data that was generated by the attacker are collectively referred to as *artifacts*.

Upon receipt of a suspicious email report, Phisherman will identify the embedded links within the email, record them in the IODEF incident report, and retrieve the associated web page content for each link, if the sites are still active. All properly formatted URLs in a suspected email will be extracted for crawling. The system will crawl both the listed URL as well as the base URL. The base URL is defined as the URL starting with the protocol and ending with the port number or the first "/" after the "//" protocol enders. Phisherman will crawl these URLs but will not navigate to any URLs that lead away from the base URL. Since the information that can identify the site as a phishing site may not be contained on the initial landing page, Phisherman will crawl up to ten pages for each URL and base URL in an effort to extract the information required to identify phishing sites. The results of the embedded link crawling will be stored in the incident report containing the original email report.

The data conversion process will also populate other elements of the IODEF record that may be useful for secondary verification or future investigations. Phisherman will retrieve "WhoIs" information on the domain name registration and IP address ownership for all suspect URLs in the IODEF report. If the incident report originates as a suspicious URL or domain name, the data conversion process will identify all emails recorded in the database containing the URL or domain name. All of this information will provide a more complete picture of the suspected incident.

Secondary Verification

Phisherman will automatically verify all incident reports submitted to the repository, regardless of the source. Secondary verification will use all available information to make a determination as to the probability that the incident report is in fact phishing. One source of information will be the brand owner profile. Each brand owner will have the opportunity to identify the mail servers and web servers used by their organization, including those used by their contractors. The concept is similar to white lists of known-good web sites and mail servers. Phisherman will also be able to incorporate domain keys and SenderID Framework information, similar to the content of the brand owner profile. Many organizations will have a difficult time creating such a profile, but for those that can create and maintain an accurate profile, Phisherman should be able to provide a very low false-positive rate.

Secondary verification will be performed through a combination of heuristic rules applied to the information available in the incident report. These tests are both individually and cumulatively non-deterministic. They will not confirm or deny that a given incident report is phishing, but they are good indicators that can be used as part of a scoring algorithm. Below are the tests that Phisherman will use:

- Have elements of this incident report been marked before as used in a phishing attack?
- Are there links in the email?
- Are any URLs in the email an IP address?
- Is the DNS name new?
- Is the site using a valid SSL certificate?
- Is the site on a shared environment, such as GeoCities or Yahoo?
- Does the root of the site correspond with data on the form?
- Is it referencing a site on the same network as the bank?
- Is the mail server correct?
- Is the email sender different than alias?
- Does a reverse PTR exist?
- Does an SPF record exist?
- Is the email source on a spam list, such as Orbs or Spamhause?
- Are embedded URLs on the same IP range as known phishing sites?

By using these tests in addition to the first-pass Bayesian filter, Phisherman will be able to determine with a high level of accuracy whether or not a specific incident report is indeed phishing.

Duplicate Checking and Incident Linking

Incident reports and components of incident reports will be checked for duplication, and exact duplicates of incident reports will be eliminated for efficient storage. An incident report is a duplicate if and only if the exact same incident report was received previously. All aspects of the incident report must be identical to a previously reported incident in order to be considered a duplicate, including items such as the targeted victim in the "To" address of an email. If any component of the incident report varies, a new incident report will be recorded.

Phisherman will reduce storage requirements and improve detection of related incidents by storing only one copy of an artifact component in the database. The situation is complicated by the possibility of similar but not identical incident reports. Comparison of similar artifacts may reveal only small differences. In these cases, static and variable parts of artifacts are identified in the repository. A single instance of the static part is stored, while multiple instances of the variable part are stored.

Because a phishing incident can cause multiple related artifacts, Phisherman performs incident linking. Related artifacts may be multiple emails sent to different recipients, multiple independent discoveries of the same fraudulent web site, or multiple artifacts related to different phases of the same attack. Incidents can be linked based on targeted brand, email attributes, web site attributes, or malware artifacts.

Data Distribution

Phisherman will provide incident report information through subscriptions and queries. Subscriptions are periodic, pre-defined reports in which the content is determined by the subscriber's role and organization. Queries will provide a flexible mechanism for retrieving incident reports that meet specified criteria. Recipients will have the option to receive query output as an XML file of IODEF reports or visually displayed on their web browser.

Subscriptions

Subscriptions provide periodic data to organizations that require frequent updates on new incident reports. Real-time subscriptions provide data to first responders (organizations requiring the most up-to-date information to mitigate the effects of a phishing attack). Such organizations include take-down services, ISPs, national incident response teams, and other organizations in position to do something about the reported email or sites. Since time is critical for response, these organizations require up-to-the-minute information on new incident reports. Other organizations require periodic reports, but on a less urgent basis. Examples include brand owners that might receive daily reports of incidents related to their brands. The organization's role will determine the data to be generated by the subscription.

One form of real-time subscription is the URL blacklist. ISPs and other email filtering service providers will receive a blacklist of suspected phishing URLs and domain names. To preserve the privacy of the targeted brands, the blacklist will use the same hashing process employed in the existing APWG blacklist. The hash allows the filtering organization to identify emails containing phishing URLs without revealing the targeted brand names. Due to the nature of the input and automated verification of phishing reports, the blacklist

will include a confidence factor for each URL. If a false positive is detected, the confidence factor will be set to zero for future versions of the blacklist.

Except for the URL blacklist, other subscriptions that provide incident report data will use the IODEF format for the output. Brand owners, national incident response teams, and some service providers will receive periodic reports related to their brand or geographic region. These subscriptions may occur on a delayed basis for subscribers with less urgent requirements. The output of these subscriptions will be an XML file of IODEF incident reports.

Queries

The Phisherman repository will be implemented as a relational database based on the content of the IODEF draft standard with the Phraud Extensions. Stored data will include the full incident report, the artifacts associated with the report, and the actions taken in response to reports. The database schema closely resembles the IODEF format, with the addition of links between related artifacts and incident reports. The schema is designed to allow law enforcement and researchers to identify attack patterns across multiple incident reports.

The database schema supports a number of requirements:

- Fields to record artifacts from the entire life cycle of a phishing attack
- Inclusion of both sanitized and unsanitized versions of artifacts
- Inclusion of metadata about the fraud activity report and artifacts
- Inclusion of a confidence rating assigned during data submission
- Inclusion of responses taken to the incident
- Possible updates or revocation of fraud activity reports
- Support for Unicode character sets.

The database supports queries for many types of users, including brand owners, researchers, and local law enforcement, with appropriate restrictions imposed by the user's role and identity. These queries will be initiated by authorized users through a web form interface. This interface will allow the user to retrieve and review the entire content of one or more incident reports, including related incidents. Access control policy will restrict the incident reports provided in response a query.

Access Control Policies

Secure access to the repository will be provided through a access control policy enforced in the application server that acts as a "gatekeeper" for the database. The Phisherman access control model is a combination of role-based access control and identity-based access control. When a user logs into the repository, the user's role will be determined by his organizational affiliation. Organizational roles will include law enforcement, incident response teams, brand owners, administrators, and researchers. The identity-based component of the access control model will be driven by the user's parent organization. For example, a user in the role of "brand owner" will have a specific set of brands owned by their organization. The user will be restricted to incident reports related to those brands. The law enforcement role will have few restrictions since an investigation may cross multiple brands and countries. Law enforcement will also be allowed to update reports as they acquire more information about the incidents. The access control policy will be applied to all subscriptions, queries, and incident report updates.

KEY CHALLENGES

There are several critical technical and non-technical challenges to implementing the Phisherman system. Among the technical challenges are effective automated verification, linking related incidents, handling the high volume of legitimate and false incident reports, minimizing denial of service attacks, and protecting the privacy of both brand owners and individuals. The primary business challenge is to ensure that the success of the project does not endanger the businesses upon which Phisherman will rely to contribute incident report data.

Automated Verification

The Phisherman project is attempting to automate a process in which even human experts occasionally make mistakes. We do not expect, or require, that the automated verification process will be perfect, but will constantly strive to improve the verification process during the initial implementation and in later iterations of

the system. The project will measure the false-positive and false-negative rates against known data sets to assess the performance of the automated verification process. As new heuristics are developed, the project will reassess the performance of the automated verification process.

When the automated verification process is found to be in error, human experts will be able to override the confidence value assigned to an incident report. In the case of a false positive, the human experts will be able to assign a confidence factor of "0," thereby marking the incident report as a false report.

Incident Linking

Some potential users of the Phisherman repository may require that every incident report be recorded, even if the majority of the incident report is a duplicate of previously reported information. For example, law enforcement may need to determine the number of reported victims of a particular attack and perhaps each individual's email address. Therefore, Phisherman assumes that an email submission itself is not a duplicate unless it is 100% identical to a previous submission. However, parts of the email (URLs, images, attachments, from addresses, message bodies, etc.) may be duplicated from email to email. Phisherman will create a hash of each component of the email and save it only once to the database. As subsequent incident reports are confirmed as phishing, Phisherman will hash the artifact components and compare them to the artifact components already in the database. The posting process will create a unique record in the database for each incident report. If a specific artifact component already exists in the database, the new incident report will only record the relationship to the artifact component, not the artifact itself. The relationship is a fully indexed associative table. The space needed to maintain these relationships is relatively small, even under large numbers and assuming that a BIGINT field is used. For example, assume that 10,000 incident reports a day are confirmed with an average of 20 components and that 98% of those components are duplicated, the amount of space necessary to save the relationships is approximately 90 megabytes a year.

There are two challenges for linking incidents reported to the system. First, there are multiple artifacts and copies of artifacts all relating to a single phishing attack that should be linked to form a comprehensive picture of a single incident. Second, separate incidents can be linked together in various ways to create patterns across attacks, including attempting to link methods, means, and likely attackers.

Incident Linking within an Attack

The Phisherman system is likely to collect multiple instances of phishing attack artifacts related to a single attack. These artifacts include:

- Multiple instances of an email sent to different recipients
- Multiple, independent discoveries of a web site on a single host
- Multiple instances of the same web site on different hosts
- Various phases of the same attack involving multiple artifacts, such as an email, URL, and malware sample

The majority of incidents in a single attack will be linked via the target URL for the phishing site itself, as this is (usually) a unique value. Using this as a base, email lures, hosting locations (for portable URLs), and phases of an attack can all be linked together. Duplicate report culling can be handled by comparing the physical attributes of each report on a given URL. A single URL may not link an entire event, however – use of URL forwarding, DNS aliasing, data submission proxies and other techniques can add several independent URLs to a single incident. Future enhancements to the Phisherman query capability should allow for automated discovery of such links as well as manual updates to link such artifacts together.

Multiple Incident Linking

One of the intended uses of the Phisherman system is to link related attacks, tracing the history as the attacker shifts attack methods and targets. The goal is to enable linking of related incident reports and their artifacts so that the evolution of a series of attacks, or the reuse of attacker resources, can be identified. Patterns of attack can greatly help investigators anticipate future attacks, and perhaps track down actual perpetrators. Such linkages can also help law enforcement scope attacks better and justify applying more resources as the overall level of fraud warrants. Unfortunately, such linkages – beyond the basics – are difficult to find and

maintain automatically. Thus the system needs to allow for various linking attributes that can be added and updated by investigators.

There are four major attack vectors for linking: phishing lures (emails), phishing site characteristics (e.g., kits, content, delivery mechanisms, exploits, malware), phishing locations (providers, countries, etc.), and targeted brands.

Lure classifications can lead to dozens of potential link points. Example key commonalities include content (language/text, included malware, other body contents), routing (origin, forwarding), technology/methods used (mass mailers, known spam sources, spam avoidance techniques), addressing (particular email addresses, addressing techniques, common recipient targeting lists), and geographic distribution of any providers involved. Phisherman will record these attributes of the lure to enable identification of related incidents.

Phishing sites themselves provide a wealth of information for linking events – the challenge is seeing the forest for the trees. Web site content and style are key for tracking groups; some candidates here include: kits used (content), programs for processing data, techniques for grabbing real site content, fingerprints in the source (e.g., comments within the tags), destination addresses of stolen credential drops, types of site forwarding or obfuscation, malware/exploits present, and credentials asked for in the form. Hosting locations and providers for the sites, DNS, and any involved domain names are useful for tracking problematic providers as well as particular groups. This can include IP addresses, autonomous system (AS) numbers, domain registrars, domain registration information (e.g., owner, admin email), DNS servers, and the countries or localities for these providers. All of these attributes of the data collection site may be useful in identifying phishing patterns.

Phisherman can play a role in developing the "big picture" of how and why phishing attacks are carried out against particular brands. Phisherman will be able to identify which brands are being targeted, when the attacks occurred, and how the attacks were implemented. If this information can be combined with other information concerning the targeted organizations (e.g., a lack of ATM Track 2 validation by banks or credit unions, or sharing a common hosting location with another target), an investigator may be able to "connect the dots" to reveal motivations for certain perpetrators or vulnerabilities which the targeted brand owner can then attempt to resolve. Marrying up the real-world information like this with the incident reports tracked by Phisherman is likely to be a significant task.

Handling High Volumes

The major challenges of handling high data volumes are rapid verification and the process of posting incident reports to the database. We believe that automated verification is essential as phishing attack volume grows. While the problem seems similar to those faced by the anti-virus industry, the scale is much greater. Anti-virus companies typically deal with 50–75 new malware incidents per day. The APWG reports that as of April 2006,[1] phishing attack volumes are running about 10 times that volume at approximately 600 unique emails per day. Automated verification is essential for any organization attempting to handle that type of volume at reasonable labor costs. Therefore, addressing this issue is a significant technical component of the Phisherman project. The Phisherman approach is to process incident reports in stages, with the goal of discarding false reports as early in the process as possible without generating excessive false negatives. Each succeeding stage of verification will provide greater accuracy but at greater cost in terms of processing resources.

The other aspect of high volume, the process of posting incident reports to the database, is addressed by separating the incident "parsing" and verification functions from the database host. Volume is driven by both the number of unique attacks and the number of times an attack is reported to the repository. As volume increases, Phisherman supports replications of application servers that perform the incident report processing prior to posting the report to the database. The actual posting of reports to the database should not be a serious issue with respect to handling reporting volumes.

Protecting against Denial of Service Attacks

Denial of service attacks are a potential risk based on the past history of spam blacklists and similar projects. Identity thieves may try to disrupt the data collection and distribution functions of the repository, particularly if it is having an impact on their criminal activities.

These attacks could take a variety of forms, including flooding the public email addresses with email, or network flooding attacks on the publicly accessible web servers.

Phisherman will protect against distributed denial of service attacks by isolating the database and application servers from direct access by data submitters and recipients. All external access to the repository will pass through replicated and distributed web servers. Users will connect to one of several web servers for queries, data submission, and data subscription requests. If an attacker performers a denial of service attack on one or more web servers, other replicas will take the load and provide continuous service. Since web servers are essentially stateless, new replicas can be easily created as needed to handle the attack volume. Communications between the web servers and backend systems will use encrypted virtual private networks to exclude attack data from the internal communications.

Privacy Issues

There are two significant concerns with respect to privacy: brand names and individuals. Many brand owners wish to minimize the public association of their brand names with phishing attacks. Brand owners fund, and therefore control, many of the domain name discovery services that we expect will provide the best early warning of phishing attacks. To encourage brand owner participation in the Phisherman project, the system will avoid unnecessary exposure of brand names in the data distribution. One such data distribution is the URL blacklist, discussed earlier, which is used by spam filtering services to identify phishing email with embedded links to known phishing web sites. Another method of protecting brand owner privacy is through the access control system. Researchers generally do not need to know the brand names in the incident reports they analyze or use for testing. The data distribution for users in the "researcher" role will omit the brand name from the data. The combination of URL hashing and access control should eliminate unnecessary distribution of brand names from incident reports.

Individual privacy is an equally important issue addressed by Phisherman. Individuals may be identified by the "To" address in an email, or by the body of the email in the case of a targeted attack. In a targeted attack, the individual may be identified in the email by name, account number, or other information. Most data distributions, except for those to law enforcement, will not contain the "To" addresses for email artifacts embedded in incident reports. If an incident is identified as being a targeted attack, it will be completely omitted from data distributed to brand owners and researchers until such time as automated methods have been implemented for identifying the personalization in the email body. So far, these types of targeted attacks have been infrequent.

Business Issues

As described earlier, Phisherman will rely on existing data collection systems operated by for-profit companies that currently provide data to paying customers. The customers of the data collecting companies are usually either companies protecting their brands, or people who wish to be protected from phishing attacks. Phisherman will take data from these data collection companies and other sources, and then distribute it to organizations which may be customers or potential customers of the data collecting companies. Since Phisherman requires data from the existing data collection systems, it is vital that Phisherman not conflict with the business models of the for-profit data collection companies that will provide incident reports to Phisherman. Phisherman will provide mechanisms that will preserve or enhance the value of the data collected by the existing for-profit companies, even though it may appear that "free" data would destroy the companies.

Phisherman does not remove the need for brand owners to continue to purchase Web and domain name searching services and preserves the business models of domain name searching companies. Phisherman itself does not actively monitor new domain name registrations or search for new phishing web sites. Instead, it relies on the companies currently providing that service to brand owners to provide the data to Phisherman. If a brand owner does not use a search company and does not perform the searches themselves, it is highly unlikely that newly registered domain names would be entered into the Phisherman repository prior to the discovery of the first related phishing email. If a brand owner allows their domain search service to submit phishing domain name data to Phisherman or they submit the data themselves, the brand owner's customers will benefit by improved protection from the existing spam filtering services and other systems

that inhibit phishing attacks. By improving the utility of the domain name search data, Phisherman enhances the value of the service to the search companies' customers.

For-profit anti-phishing services that derive their revenue from the end users that they protect do have a more significant concern over the real-time data distribution aspects of Phisherman. These anti-phishing companies view the timeliness and accuracy of their data to be an essential business asset. They generally collect, verify, and distribute the data at their own expense. Therefore, they are not expected to share their data with Phisherman in real-time. However, based on interactions with one such company, it should be feasible for the anti-phishing services to provide slightly aged data to Phisherman after the time value of the data has diminished. Since this data will arrive some time after the initial discovery of the attack, it will not be very useful for first responders. However, the additional incident report will be useful to law enforcement and researchers.

LAW ENFORCEMENT USES

Phisherman will enable law enforcement to identify relationships between incident reports, and with externally obtained data. As described earlier, Phisherman's incident report submission process will create links between new reports and existing reports with common artifacts or artifact components. The submission process automatically creates links between reports sharing the following elements in common:

- Parsed email components, including headers and bodies
- Hashes of binary large objects, such as images
- Common identifiers for reused data collection sites
- Hashes of normalized URLs
- Domain name registration information
- Malware names as determined by anti-virus scanning

In addition, there are elements of incident reports that are easily included in a standard database search. Those elements include:

- URL or domain
- Data collection site (by IP address or domain)
- Email sources
- Malware
- Embedded images (e-mail body or web site)
- Brand name
- Targeted victim

Phisherman will allow law enforcement to track the evolution of phishing attacks across brands, methods, and servers by identifying incident reports containing common elements. The following example shows how Phisherman could be used in one hypothetical investigation.

Example

Assume that a brand owner, XYZ Corp., has submitted a complaint to law enforcement regarding a series of phishing attacks against the XYZ brand that occurred in a two-week period. XYZ Corp. is a relatively small company and not a frequent target for phishing attacks. An investigator needs to determine (1) if the incidents would appear to be related and (2) if the incidents are related to attacks on other brands. Each attack against XYZ Corp. included an email lure that directed the recipient to enter personal information on web site imitating XYZ's web site. The investigator plans to use Phisherman's query capability to examine the characteristics of the recent phishing incidents involving XYZ and search the database for other incident reports with similar characteristics.

The investigator starts by submitting a query to Phisherman's web server for all attacks against XYZ Corp. for the two-week period in question. The investigator can choose to view them individually online, via the web browser, or retrieve a file containing an XML representation of the incident reports. In this case, the investigator finds that there are five incident reports and chooses to view them online. Each incident report provides full content of the email lure, the URL of the data collection site, the web pages of the collection site, domain name and IP address registration of the data collection site, and the date and time that each incident was reported. The investigator notices that in three of the five incidents, the body of the email lure is identical, except for the IP address of the embedded link to the data collection site. According to the timestamps of the reports, these three incidents are the first three incidents in the series. The remaining two incidents have distinctly different email bodies and reference different data collection servers. However, the data collection servers for all five incident reports are located in the same country. Furthermore,

the content of the servers is identical. The investigator concludes that the attacks are likely to be related, answering his first question.

Two answer his second question, the investigator submits a query to determine if the IP address of any of the data collection servers in the first five incident reports was used in another incident report over a time period starting two weeks prior to the attacks on XYZ Corp, and ending two weeks after. The search results in no incident reports being found. The investigator then shifts his attention the emails used in the attacks on XYZ. Each of the first three emails, which shared a common email body, also used the same subject line. The investigator submits a query to determine if the same subject line appears in other incident reports. The results reveal nine other incident reports against two relatively small companies in the same line of business as XYZ Corp. Further analysis of the bodies of the nine incident reports reveals similar, but not identical, content to the attacks on XYZ Corp. The language of the email bodies contain identical warnings and instructions to the recipient, but the specific information related to the brand name and URL varies with each. Each of these nine emails also used an IP address for the data collection server. Furthermore, the registration information for the IP address indicated the same country as the initial five incident reports on XYZ Corp. The investigator concludes that based on the similarity of the social engineering aspects of the emails and the common country for the data collection sites, all fourteen incident reports may be related.

FUTURE WORK

Phisherman is currently being developed as version 1.0 to support the essential requirements for phishing data collection, verification, archival, and distribution. The project team has identified other features that would be useful in future versions of the system. Laboratory testing and field testing will likely reveal other necessary enhancements.

Emerging Attack Methods

As phishing attack methods evolve, Phisherman's data formats and database will need to evolve. The current data formats emerging from the IETF working groups contain explicit support for the existing attack methods, particularly email lures, web sites as data collectors, and malware. Phishing attack methods are likely to continue to evolve; the data collection and storage methods will need to evolve with them. Future enhancements to Phisherman should include support for instant messaging and voice-over-IP ("vishing")[12] attack methods. It is likely that the data interchange standards, database schema, and verification methods will also evolve to support the emerging attack methods.

Enhanced Privacy

It is essential for any data collection system to preserve the privacy of individuals and institutions if it is to gain public and commercial support. The initial Phisherman system will provide source-based access control policies, URL hashing, and limited sanitization capabilities for email headers as mechanisms for protecting the privacy of both individuals and brand owners. The access control policy will allow or deny access to the entire incident report based on the source of the report. A future version of the Phisherman system should provide access to portions of an incident report, including some artifacts, which may not be sensitive. Additionally, Phisherman should provide automated sanitization of personalized emails from targeted attacks as an alternative to prohibiting distribution of the personalized emails in their entirety.

Expanded Participation

Ultimately, the project's goal is to receive phishing incident report data from diverse sources around the world in order to provide a comprehensive view of the phishing problem, and provide the most complete investigative tool to combat phishing. Phishing attacks do not respect international boundaries, so the project will to solicit participation from industry, academia, and quasi-public incident response organizations around the world. As with most software systems, the beta test phase will limit the data sources and recipients. After the testing phases have been completed, the project will work with the APWG to greatly expand participation.

CONTACT INFORMATION

The Phisherman project is a work in progress and is soliciting input from potential users and contributors, including data collectors, first responders, brand owners, researchers, and law enforcement. The initial

implementation is currently under way. To learn more about the project, find out how your organization can participate, or make suggestions for future capabilities, please contact one of the following people:

Gregg Tally, Principal Investigator, Gregg.Tally@sparta.com
David Jevans, Chairman Anti-Phishing Working Group, Dave.Jevans@antiphishing.org

SUMMARY

The Phisherman project is building a phishing data collection, verification, archival, and dissemination system for use by first responders, researchers, brand owners, and law enforcement. Phisherman leverages the extensive collection capabilities that have already been developed by the anti-phishing community and expands those capabilities with automated verification, standards-based data formats, privacy protection mechanisms, and incident linking capabilities. As a law enforcement tool, Phisherman will provide a comprehensive archive of phishing attack data collected from sources around the world. The database will create easily searchable data fields to identify common features in multiple incidents, allowing investigators to chart the evolution of a series of attacks across time.

ACKNOWLEDGEMENTS

Several additional people on the Phisherman project team have contributed to the ideas presented in this article. The authors acknowledge Brian Appel (SPARTA, Inc.), Patrick Cain (Cooper-Cain Associates), Vijaya Ramamurthi (SPARTA, Inc.), and Brandon Shalton (Internet Compliance Systems) for their contributions to the design of the Phisherman system.

NOTES

1. Anti-Phishing Working Group, "Phishing Activity Trends Report," http://www.antiphishing.org/reports/apwf_report_April_2006.pdf (no longer available). Current report available at Antiphishing,org Accessed April 2006.
2. Anti-Phishing Working Group, "Charter for Phishing Repository, Data Stream and Alerts Subgroup," http://antiphishing.kavi.com/apps/org/workgroup/repository/description.php (Members-only site). Accessed August 13, 2006.
3. Microsoft Corporation, "Industry, Law Enforcement Team to Launch Digital PhishNet," Microsoft Press Release, December 8, 2004.
4. Joris Evers, "Neighborhood Watch for Phishing Launches," CNET News.com, March 28, 2006.
5. Symantec Corporation, "Phish Report Network," Phish Report Network Brochure, 2005.
6. Symantec Corporation, "Phish Report Network Data Provider Agreement," http://www.phishreport.net/phish_report_network_data_provider_agreement_online_06_04_27.pdf. Accessed April 27, 2006.
7. Vipul Ved Prakash, Christopher Abad, and Jamie de Guerre, *'Cloudmark's Unique Approach to Phishing'* (Cloudmark, Inc., 2006) http://www.cloudmark.com/releases/docs/wp_unique_ approach_10550406.pdf.
8. R. Danyliw, J. Meijer, and Y. Demchenko, "The Incident Object Description Exchange Format Data Model and XML Implementation," Internet Engineering Task Force Extended Incident Handling Working Group Draft, June 8, 2006. http://tools.ietf.org/wg/inch/draft-ietf-inch-phishingextns/draft-ietf-inch-phishingextns-03.txt.
9. Patrick Cain, "Using XML to Support Robust Information Sharing: An IODEF Automated Approach" (AntiPhishing Working Group 2006 eCrime Research Summit, November 2006).
10. Patrick Cain and David Jevans, "Extension to IODEF-Document Class for Phishing, Fraud, and Other Non-Network Layer Reports," Internet Engineering Task Force INCH Working Group Draft, June 14, 2006.
11. Paul Graham, Hackers and Painters: Big Ideas for the Computer Age p. 121. Sebastapol, CA: O'Reilly.
12. Andy Patrizio, "Vishing Joins Phishing as Security Threat", InternetNews.com, July 11, 2006.