A Trend Micro Research Paper

Email Correlation and Phishing

How Big Data Analytics Identifies Malicious Messages

RungChi Chen

Contents

Introduction

Phishing is a long-running problem that has taken a turn for the worse. Phishing emails now so closely resemble legitimate ones, making it very difficult both for users and automated systems alike to tell them apart. As such, users end up clicking links embedded in phishing messages that take them to malicious sites, which directly or indirectly steal their personal information.

This research paper describes a newly developed Trend Micro methodology that correlates the format of emails with sending agents to detect phishing messages. Using actual examples, we demonstrate how we use "big data analytics" through the Trend Micro™ Smart Protection Network™ infrastructure to proactively identify phishing messages so we can protect our customers from today's more sophisticated email threats.

Phishing in 2013

Phishing in 2013 is more advanced and sophisticated than it has ever been. Phishing messages are becoming harder and harder to distinguish from legitimate ones.

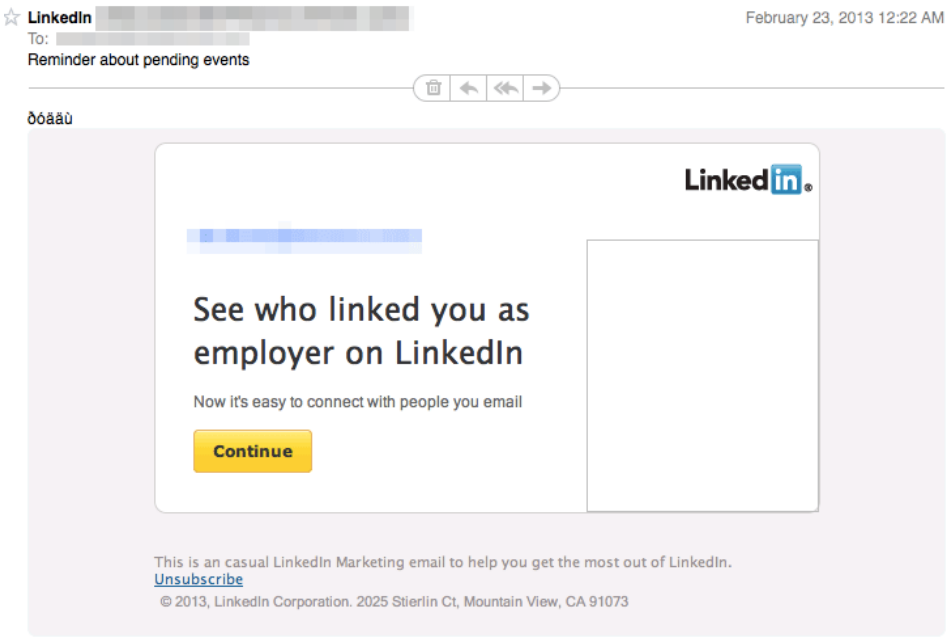

Figure 1: Sample phishing message recently obtained from LinkedIn®

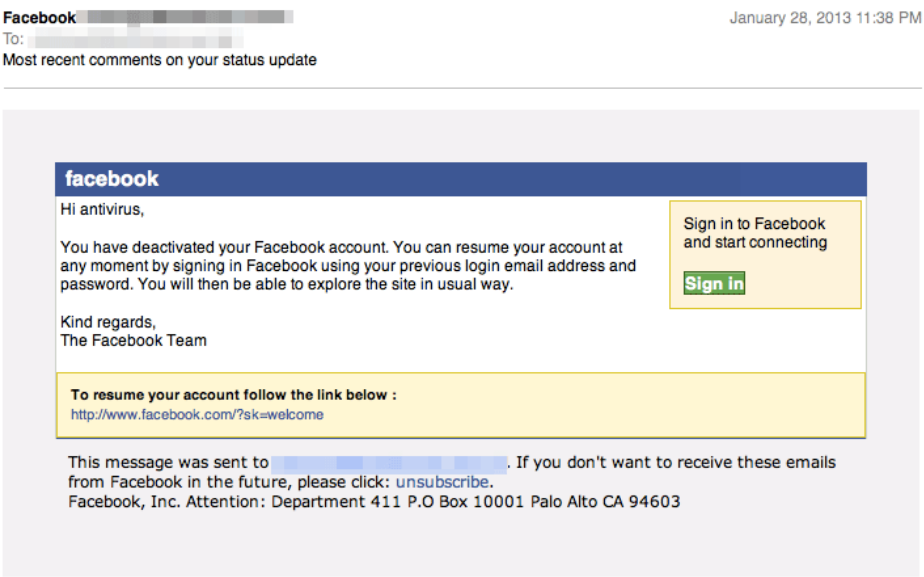

Figure 2: Sample phishing message recently obtained from Facebook

It is possible that attackers are taking legitimate messages and just replicating them with slight modifications to links. These links lead to malicious sites with exploit packages designed to compromise users' systems.

Similarities in content make it difficult for security vendors to detect and filter messages based strictly on content. Filtering such messages could result in potential false positives, as legitimate emails could be wrongly classified as "spam." Classifying messages based on embedded URLs has also become difficult as their average life span has also become extremely short.[1]

New techniques are necessary to detect more carefully prepared malicious messages. The technique discussed in this paper uses big data analytics, which correlates large sets of spam data with one another to determine their sources.[2]

[1] Rod Rasmussen and Greg Aaron. (October 2012). "Global Phishing Survey: Trends and Domain Name Use in 1H2012." Last accessed June 20, 2013, http://docs.apwg.org/reports/APWG_GlobalPhishingSurvey_1H2012.pdf.

[2] TechTarget. (January 10, 2012). *SearchBusinessAnalytics.* "Big Data Analytics." Last accessed August 5, 2013, http://searchbusinessanalytics.techtarget.com/definition/big-data-analytics.

The State of Email Authentication

Several protocols, including DomainKeys Identified Mail (DKIM) and Sender Policy Framework (SPF) have been designed and implemented to confirm the senders and integrity of emails.[3]

The two mechanisms named above are sender-authentication technologies that assist in controlling spam and improve the deliverability of legitimate messages. DKIM applies public key cryptography to emails. Senders use private keys to sign their messages and publish public keys via Domain Name System (DNS). When recipients get the emails claimed by a specific company, they will retrieve the public keys from the DNS to check if the emails were really sent by the identified company or not. Applying DKIM can help solve the phishing problem.

The mechanisms above, however, do not solve all problems related to spam and phishing due to three reasons. First, the global DKIM (35%) and SPF (63%) adoption rates are not high enough.[4] Second, DKIM is insensitive to reply emails, which means some phishing emails can be sent with valid DKIM signatures.[5] Finally, forwarded emails have reasonably high signature failure rates reaching as high as 4%.

Mail Network Basics

Our methodology is designed to correlate emails with the IP addresses responsible for sending them out. In some ways, it uses the similarities between legitimate and phishing messages in our favor.

We developed methods to identify commonly sent emails. For these methods, however, we tried to generate an identified signature for each email. The elements considered for this signature include the domain of the sender's address, the format structure, the content of email body, and whether it was authenticated or not. Emails that have been classified are then correlated with the IP addresses that sent them out.

---

3 *DKIM.org.* "DomainKeys Identified Mail (DKIM)." Last accessed August 5, 2013, http://www.dkim.org/; *Openspf.org.* "Sender Policy Framework: Project Overview." Last accessed August 5, 2013, http://www.openspf.org/.

4 Lars Eggert. (Last updated May 30, 2012). *DKIM Deployment Trends.* Last accessed August 5, 2013, http://eggert.org/meter/dkim; Lars Eggert. (Last updated May 30, 2012). *SPF Deployment Trends.* Last accessed August 5, 2013, http://eggert.org/meter/spf.

5 Douglas Otis. (June 14, 2011). *TrendLabs Security Intelligence Blog.* "Possible Phishing with DKIM." Last accessed August 5, 2013, http://blog.trendmicro.com/trendlabs-security-intelligence/possible-phishing-with-dkim/.

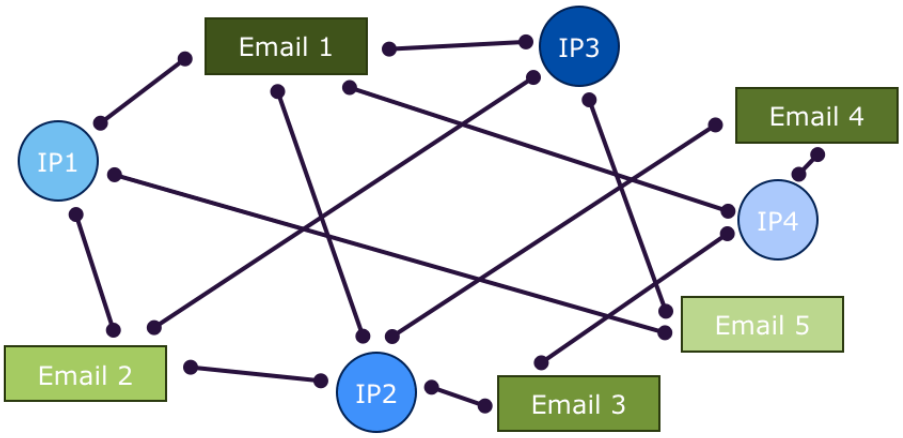

Figure 3: IP address-email correlation graph

The diagram above presents a relatively simple representation of the correlation that occurs with a very limited number of messages and IP addresses. In any actual situation, this could quickly become complicated. Another issue of interest has to with which IP addresses were used to send out similar messages. Figure 3 is a simplified diagram with email nodes that become links between IP addresses.

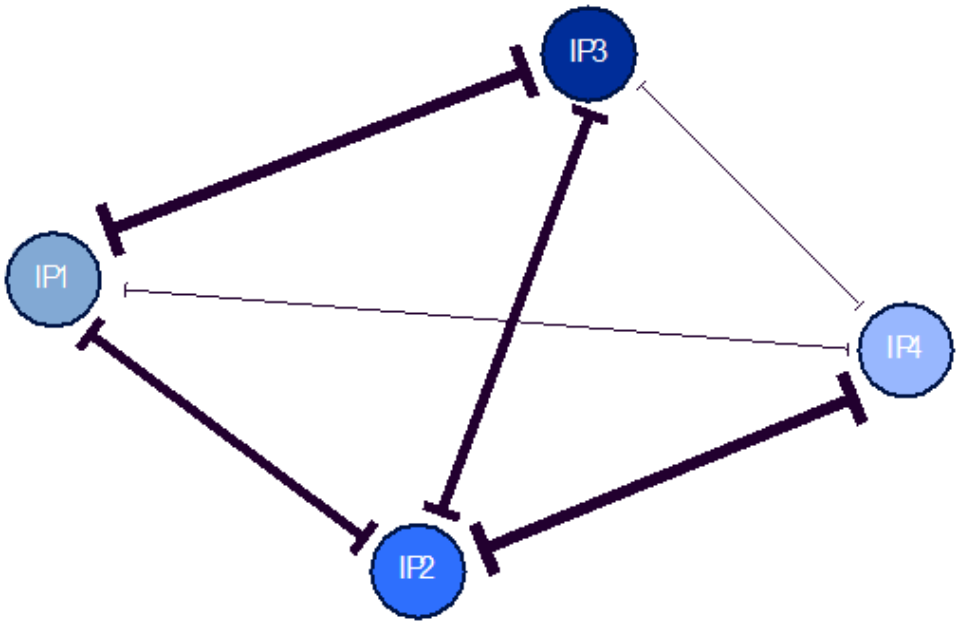

Figure 4: Simplified version of Figure 3

The distance between any two IP addresses in Figure 4 becomes shorter if they send out more similar emails. Conversely, the fewer similar emails they send out, the further apart they are.

Two concepts related to graph theory were used to create the diagram above—graph density and modularity.[6] Graph density in the diagram looks at how the nodes are connected to one another. This can be used to determine if a particular portion of the network or the network as a whole is well connected.[7] The closer a graph's density is to 1—the maximum value—the better organized and more well connected a particular "community" of nodes is.[8]

Modularity, on the other hand, determines how well a network can be divided into distinct groups.[9] Ideally, the nodes within a group or community should be well connected to one another but not necessarily to nodes in other groups.

How do we use these to detect phishing messages? Conceptually, the nodes (i.e., IP addresses) in a well-connected and "tight" community send out very similar messages. In essence, if an IP address outside this community starts sending out similar messages, these are very likely phishing messages.

6 Wikimedia Foundation, Inc. (Last updated June 7, 2013). *Wikipedia.* "Graph Theory." Last accessed August 5, 2013, http://en.wikipedia.org/wiki/Graph_theory.

7 Wikimedia Foundation, Inc. (Last updated June 6, 2013). *Wikipedia.* "Dense Graph." Last accessed August 5, 2013, https://en.wikipedia.org/wiki/Dense_graph.

8 Wikimedia Foundation, Inc. (Last updated June 12, 2013). *Wikipedia.* "Community Structure." Last accessed August 5, 2013, http://en.wikipedia.org/wiki/Community_structure.

9 Wikimedia Foundation, Inc. (Last updated May 27, 2013). *Wikipedia.* "Modularity (Networks)." Last accessed August 5, 2013, http://en.wikipedia.org/wiki/Modularity_(networks).

Actual Results

To demonstrate how effective this technique is, we took a look at some actual data Trend Micro gathered.

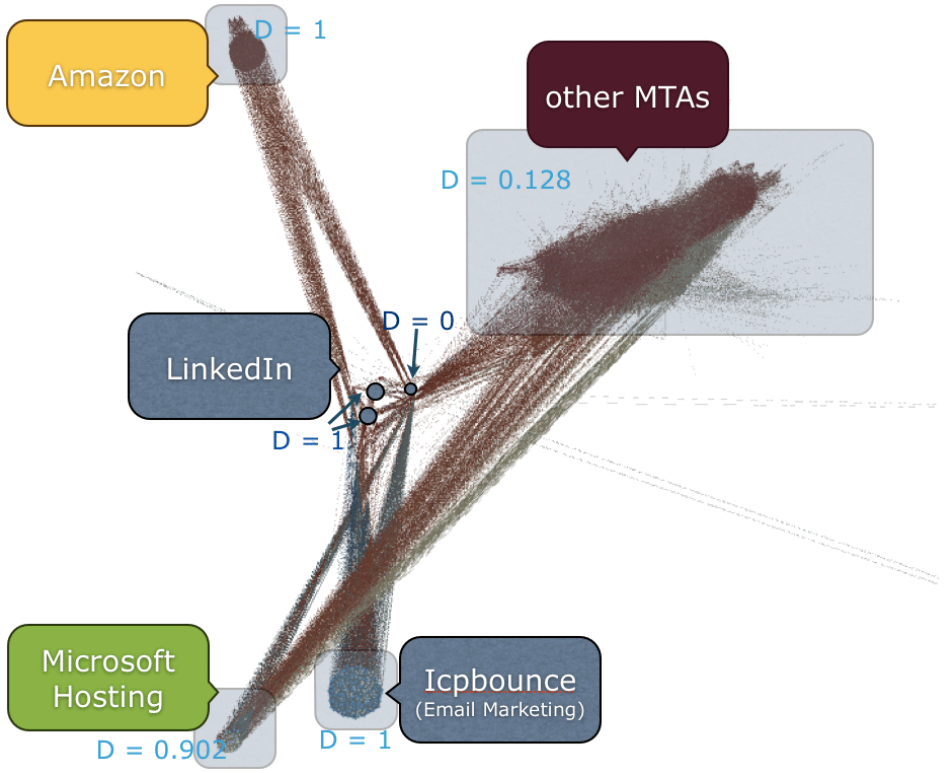

Figure 5: Part of the graph generated from Trend Micro real-world data

This technique, however, has various capabilities and limitations. For instance, it makes it easy to spot which organizational networks send out large numbers of email with very specific formats. We were able to identify four distinct subcommunities that sent out emails with very specific appearances.

To truly maximize the usefulness of such data, it should be divided into as many distinct, tight communities as possible. The end result would be a far larger version of the graph below.

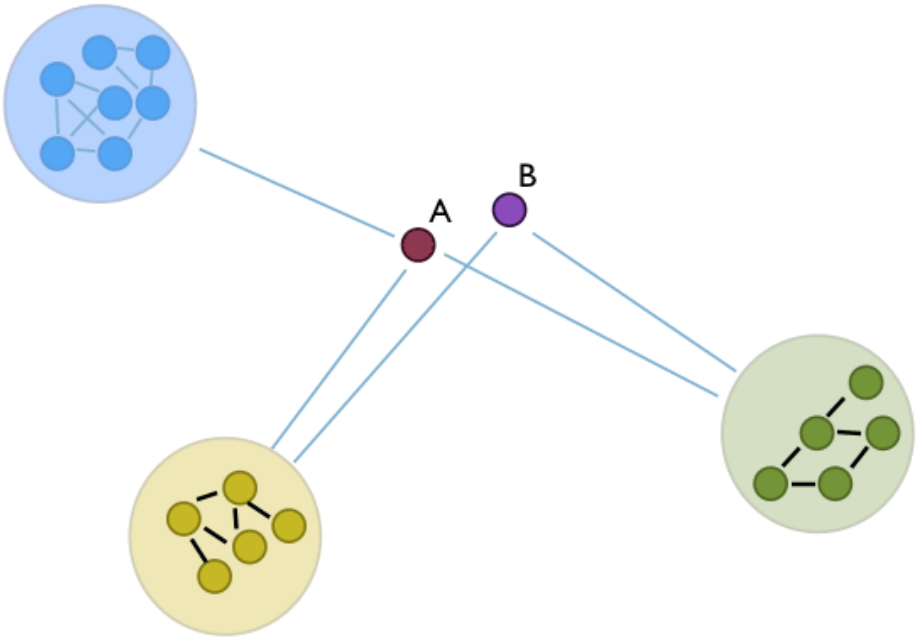

Figure 6: Simplified version of Figure 5, divided into distinct communities

Experiment Results

To gauge how effective the technique is, we chose to analyze messages from two companies—LinkedIn and QQ—to see if all of these actually originated from them.

First, we identified more than 7.8 million emails that matched LinkedIn's signature. Approximately 99% of these were sent by IP addresses associated with LinkedIn. The remaining 1%, meanwhile, may be phishing messages.

The figures were worse for QQ. Among the almost half a million messages we analyzed, more than two-thirds did not originate from QQ.

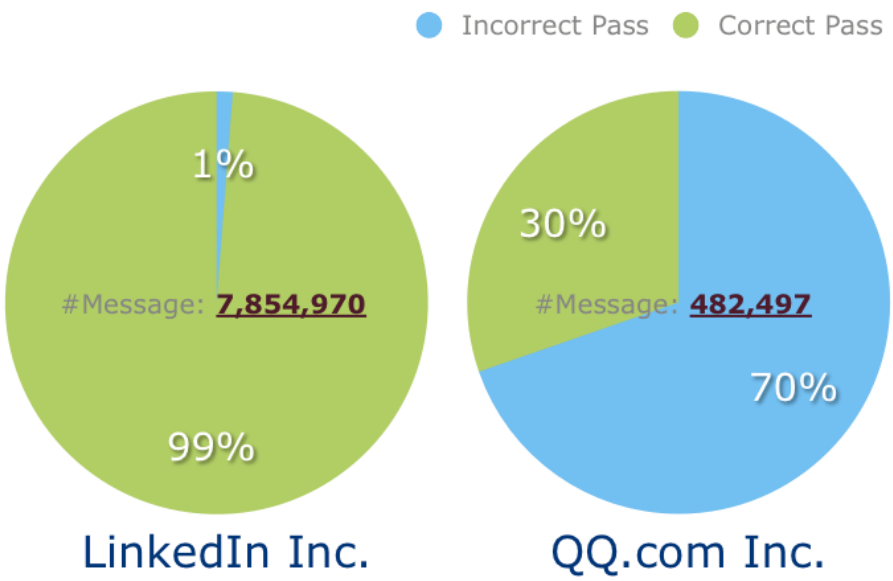

Figure 7: Percentages of identified phishing emails from LinkedIn and QQ; incorrect passes correspond to suspicious emails while correct passes correspond to legitimate messages

This newly developed technique makes it possible for us to quickly identify phishing messages, particularly more sophisticated ones, which very closely resemble legitimate emails. It allows us to turn the tables on spammers, as their attempts to closely mirror legitimate messages only make their malicious mails stand out. It is a powerful technique that is exceptionally useful against today's more sophisticated attacks.

Possible False Positive Scenario

If the mechanism proposed in this paper is adopted, false positives may arise when high-volume sites start using a new message transfer agent (MTA). An MTA, also known as a "mail transfer agent," refers to software that transfers emails from one computer to another using a client—a server application architecture.[10] Security researchers and analyze need some time to observe how the new MTA behaves. Only after receiving enough emails from the new MTA can it be added to the existing community associated with a particular high-volume site.

[10] Wikimedia Foundation, Inc. (Last updated July 17, 2013). *Wikipedia.* "Message Transfer Agent." Last accessed August 5, 2013 https://en.wikipedia.org/wiki/Message_transfer_agent.

Two methods can help reduce the number of false-positive detections. The first relies on the relationship between the email signature and the base domain instead of the original mapping between the email signature and IP address. The base domain of the new MTA usually stays the same as that of an already-existing MTA. Sandboxing technology can also be used. This ensures that any suspicious email already detected by the mechanism is analyzed in a sandbox to detect malware or malicious or deceptive content, regardless of MTA.

Conclusion

Using big data analytics to detect phishing messages was developed in response to the increased threat posed by malicious emails that closely resemble legitimate ones. This methodology not only helps detect phishing messages but also makes it easier for security companies like Trend Micro to detect such phishing messages even if they more closely mimic legitimate ones. This would, however, not be possible without knowledge of big data and previous knowledge of current threats.

Trend Micro already uses big data analytics, which makes it possible for us to better protect our customers from various email attacks.

References

- *DKIM.org*. "DomainKeys Identified Mail (DKIM)." Last accessed August 5, 2013, http://www.dkim.org/; *Openspf.org*. "Sender Policy Framework: Project Overview." Last accessed August 5, 2013, http://www.openspf.org/.
- Douglas Otis. (June 14, 2011). *TrendLabs Security Intelligence Blog*. "Possible Phishing with DKIM." Last accessed August 5, 2013, http://blog.trendmicro.com/trendlabs-security-intelligence/possible-phishing-with-dkim/.
- Lars Eggert. (Last updated May 30, 2012). *DKIM Deployment Trends*. Last accessed August 5, 2013, http://eggert.org/meter/dkim; Lars Eggert. (Last updated May 30, 2012). *SPF Deployment Trends*. Last accessed August 5, 2013, http://eggert.org/meter/spf.
- Rod Rasmussen and Greg Aaron. (October 2012). "Global Phishing Survey: Trends and Domain Name Use in 1H2012." Last accessed June 20, 2013, http://docs.apwg.org/reports/APWG_GlobalPhishingSurvey_1H2012.pdf.

- TechTarget. (January 10, 2012). *SearchBusinessAnalytics*. "Big Data Analytics." Last accessed August 5, 2013, http://searchbusinessanalytics.techtarget.com/definition/big-data-analytics.
- Wikimedia Foundation, Inc. (Last updated May 27, 2013). *Wikipedia*. "Modularity (Networks)." Last accessed August 5, 2013, http://en.wikipedia.org/wiki/Modularity_(networks).
- Wikimedia Foundation, Inc. (Last updated June 6, 2013). *Wikipedia*. "Dense Graph." Last accessed August 5, 2013, https://en.wikipedia.org/wiki/Dense_graph.
- Wikimedia Foundation, Inc. (Last updated June 7, 2013). *Wikipedia*. "Graph Theory." Last accessed August 5, 2013, http://en.wikipedia.org/wiki/Graph_theory.
- Wikimedia Foundation, Inc. (Last updated June 12, 2013). *Wikipedia*. "Community Structure." Last accessed August 5, 2013, http://en.wikipedia.org/wiki/Community_structure.
- Wikimedia Foundation, Inc. (Last updated July 17, 2013). *Wikipedia*. "Message Transfer Agent." Last accessed August 5, 2013 https://en.wikipedia.org/wiki/Message_transfer_agent.

Trend Micro Incorporated, a global leader in security software, strives to make the world safe for exchanging digital information. Our innovative solutions for consumers, businesses and governments provide layered content security to protect information on mobile devices, endpoints, gateways, servers and the cloud. All of our solutions are powered by cloud-based global threat intelligence, the Trend Micro™ Smart Protection Network™, and are supported by over 1,200 threat experts around the globe. For more information, visit www.trendmicro.com.

Securing Your Journey to the Cloud

10101 N. De Anza Blvd.
Cupertino, CA 95014

U.S. toll free: 1 +800.228.5651
Phone: 1 +408.257.1500
Fax: 1 +408.257.2003