



## Cloudmark's Unique Approach to Phishing

Vipul Ved Prakash, Founder and Chief Scientist, Cloudmark, Inc.  
Christopher Abad, Research Scientist, Cloudmark, Inc.  
Jamie de Guerre, CTO, Cloudmark, Inc.

## ABSTRACT

Phishing is a relatively new and sophisticated messaging security threat. Typically, phishers impersonate known and trusted financial institutions and organizations to access a user's personal account information. Phishers target qualified mailing lists, keep their attacks short lived, and quickly move among hosting sites. They also exploit software vulnerabilities to fool filtering software. For all of these reasons, traditional techniques such as Bayesian filters, IP-based black lists, and URL-based filters are not effective in stopping or filtering phishing attacks. Cloudmark uses a fingerprinting algorithm to identify each message in its system, which has proven to be particularly effective against phishing because it does not make any assumptions about the nature of the underlying message. The Cloudmark Global Threat Network™ consists of real-time users, who are themselves targets, who can distinguish a phishing attack, and mark it as such. Once enough users confirm it is an attack, any message with a matching fingerprint is moved into every user's spam folder. One fingerprint is enough to match all messages generated from the same phisher — even new generations of the same attack that may have been cosmetically altered—allowing Cloudmark to provide zero-time protection against most phishing attacks.

Phishing is a relatively new internet threat, and an insidious one at that, which is further exacerbated by its surface similarity to spam. Phishing attacks are relatively sophisticated and logistically different from spam. Many anti-spam technologies are unable to effectively protect users from phishing attacks. It is important to understand how phishing is different from spam to evaluate anti-phishing technologies.

## Sophisticated

The first important difference about phishing is that it is relatively more sophisticated when compared to spam. Not only are the messages that phishers send out carefully crafted to impersonate known, trustworthy financial institutions or organizations, but phishers also systematically exploit software vulnerabilities in web browsers, web servers, and local operating systems in order to fool filtering software, trick users, and steal as much information as possible. Their methods and techniques bear a much closer resemblance to “black hat” hackers than to spammers.

## Targeted

Unlike spammers who target large groups, lists, and user bases in an attempt to capitalize on some given advertising response rate, phishers, instead, target highly-qualified lists of e-mail addresses and individuals. They are not just trying to improve response rates. They are trying to evade less sophisticated data collection (honeypots, for example) and detection systems that might alert authorities to their fraud attempt.

## Transient

Phishing attacks are also very transient and short-lived, often occurring for only a few hours before disappearing. In contrast, spam tends to be sent in frequent and large batches. Since phishing is a well-defined criminal activity, with constraints that are quite different from spam, transient attacks are very necessary in order for the phisher to evade detection.

## Dynamic

Finally, phishing attacks, and the sites that host them, are very dynamic—they move between servers very quickly. Whereas a spammer is advertising some product or service from a known website, a phisher is redirecting users to a private web server/site specifically designed to impersonate a financial institution or organization. Phishers typically exploit software vulnerabilities in web servers or server operating systems in order to install their own content. And because of their increased sophistication, phishers are able to do this in an automated fashion, which enables them to compensate quickly and easily when a compromised site is discovered and taken offline. In fact, phishers cycle through compromised hosts quickly regardless of their discovery, simply to confuse or obfuscate the true source of the attack.

#### TRADITIONAL TECHNIQUES:

Several techniques have been developed and are used to filter spam. These include IP-based blacklists, Bayesian filters, heuristics engines, content fingerprinting schemes, and sender authentication. While all of these techniques are effective to varying degrees against spam, only some perform well against phishing. Here's a breakdown of what to expect with different techniques:

##### Bayesian Filters

Bayesian classifiers filter spam based on their semantic difference from legitimate communications. Bayesian filters are in wide-use—especially in end-point anti-spam products—where they perform best. There are considerable and easily distinguishable differences in the content of spam when compared to the content of legitimate messages. By contrast, phishing messages try to imitate legitimate messages in form and content and are difficult to distinguish using Bayesian classifiers trained to detect spam. It is rather hard for “vanilla” Bayesian classifiers to distinguish between spam and phishing.

##### IP-based Blacklists

Phishers use compromised machines to host phishing web pages and to send out phishing emails. IP- or source-based filters, created to address spam, are particularly poor at detecting phishing messages because phishing messages often originate from “good” hosts.

##### URL-based filters

Several URL-based filters look for specific IPs, domains, or URLs where known phishing web pages are hosted. These IPs, domains, and URLs are collected from reports of phishing messages gathered from email users and honeypots. As such, URL-based filters are fairly effective, but based on limited reporting, they can only represent a small sample of phishing activity at any given moment. Since phishers tend to frequently cycle through a large set of hosts, it is very difficult to have a comprehensive and updated list of bad IPs, URLs, and domains.

#### THE CLOUDMARK APPROACH TO PHISHING

The Cloudmark Network Feedback System (CNFS) with its fingerprinting detection lends itself almost perfectly to combat the phishing problem. First, the CNFS aggregates reports from real users who are themselves targets of phishing attacks. These reports, as opposed to reports from honeypots, provide a high-quality, comprehensive, and real-time dataset that represents phishing messages in propagation. Cloudmark fingerprinting algorithms are very effective at capturing and identifying phishing messages. Phishing messages generate a fingerprint, much like spam messages, or legitimate messages. Once enough trusted users of the community flag a message fingerprint as “phishing”, all messages that match the fingerprint are filtered.

In fact, this is so effective that Cloudmark has been chosen by industry leaders such as PayPal® to protect millions of their users from phishing attacks. By generating fingerprints which are flexible to small variations, Cloudmark makes it extremely difficult for spammers to make minor text or structure changes to phishing templates in order to evade previously-generated fingerprints. Therefore, the Cloudmark system can stop new phishing attacks in real-time, based on reports of previous attacks. This provides zero-time protection against the bulk of phishing attacks.