4 April 2017

The Email Blocklist

The Email Blocklist (EBL) is a DNS-based blocklist that contains hashes of contact email addresses used in spam[1] emails, or “drop box” email addresses. The list primarily contains email addresses at large free web-based email services. Most listed email addresses were seen in “Nigerian” 419 Advance Fee Fraud spam or other types spam that uses email addresses as their only or primary contact points.

The list is publicly available and free during its beta test period.

Purpose

The EBL is intended to block spam that cannot otherwise be blocked without blocking significant amounts of legitimate email—causing false positives. The initial focus was on “Nigerian” 419 Advance Fee Fraud spam, which is notoriously difficult to block. As the list has developed, other types of spam were added that use similar techniques. All of them are sent though mail systems that also send a great deal of legitimate email, primarily customer outbound mailservers at ISPs and free webmail services.

IP and domain blocklists in wide use cannot list these outbound mailservers because they also have many legitimate users and send a great deal of legitimate, non-spam email. Domain-based blocklists cannot list the free webmail domains used as contact email addresses for the same reason: too many legitimate users also use email addresses at the same free webmail sites.

What the List Contains

The EBL contains cryptographic hashes of email addresses seen in spam. Most listings are generated automatically by filters that observe spam emails sent to large collections of spamtrap email addresses. These filters use algorithms to identify and list contact email addresses seen in that spam. Any email address used to receive responses to spam emails can appear on the EBL, but the focus is on listing email addresses in spam that cannot be blocked by IP- or domain-based blocklists.

The EBL lists cryptographic hashes instead of unencrypted email addresses for two reasons: a hash fits the format constraints of DNS hostnames, and a hash is not subject to the strict

[1] The MSBL defines “spam” as “unsolicited bulk email”. In other words, we use the Spamhaus definition of spam.

privacy laws that often apply to email addresses in many countries. This type of blocklist is called a HASHBL.

Email addresses, unlike IP addresses or domains, always contain one character that cannot appear in a hostname: the @ symbol. A hash is a string of alphanumeric characters, all allowed. An SHA1 hash is 40 characters long, sufficiently complex to ensure that each email address in use will have a unique hash, and short enough to be used as a section (*label*) in a hostname.

Under the laws of many countries, personally identifying information (*PII*) such as email addresses requires special protections when kept as part of a data collection. The laws are similar to those protecting social security numbers and bank account information. Cryptographic hashes are one-way. You can check an email address that you already have to see if its hash matches a blocklist entry, but you cannot use a hash to derive the email address that was used to produce it. Hashes therefore do not require strict privacy protection.

Most listings on the EBL are generated automatically by filters that process incoming spam email to large spamtrap collections. This is possible because most contact email addresses meet certain criteria:

- They are usually found in the Reply-to headers or message bodies of spam emails, or both.
- They are not usually used to send spam, so a contact email address does not usually match the email address in the From header.
- They are usually at domains belonging to large free webmail services.
- They usually appear in spam that either does not contain URIs at all or that contains only “camouflage URIs” that are used to make spam appear to be legitimate email but do not point back to the spammer.

At least half of spam messages that use email addresses instead of web URLs as the primary contact points for responses are Advance Fee Fraud spam. The remainder is a mixture of spam sent by manufacturers and merchants in China and certain other countries, dating spam, job offers, mystery shopping scams, and other types of usually fraudulent offers. Email addresses are rarely used as primary contact points in advertisements for legitimate companies except in certain narrowly-defined businesses and in certain countries, or any type of bulk email that might have been solicited by any of its recipients.

The EBL also contains a small number of manual listings that meet the criteria but are not easily detected by automated methods. Most of these email addresses are used by specific spammers or spam operations that use URLs in their spam, but change their IP addresses and domains more frequently than they do their contact email addresses. These spammers can be detected by those email addresses more reliably than by IPs or domains.

Finally, the EBL contains a test entry, the SHA1 hash of **`noemail@example.com`**.

EBL Implementation

Implementing the EBL requires some code modifications to existing MTAs, antispam appliances and spam filters. Querying the blocklist is no more difficult than querying an existing IP or domain blocklist, but requires that email addresses first be extracted, prepared, and hashed.

Email addresses need to be prepared (*canonicalized*) before hashing because mailservers “see” email addresses differently than hashing algorithms do. Mailservers ignore case in email addresses, and most ignore tags, defined as the + character followed by any alphanumeric string. Gmail, the largest free webmail site, also ignores periods in the username. However, hashing algorithms create different hashes of the same email address unless formatting variations, tags and extraneous characters are first removed. Removing the variations ensures that spammers cannot evade detection by using formatting tricks.

Canonicalizing an email address for hashing requires following the steps that the EBL’s automated spam feed filters follow.

- Convert all capital letters to lower-case.
- Remove all tags from usernames.
- Remove any periods from usernames of Gmail addresses.

NOTE: To smoothe implementation, the EBL scripts also create hashes of email addresses exactly as seen in the spam emails that they observe. However, canonicalizing an email address before hashing and checking it is the best method. Doing so guarantees that the check will succeed if that email address in any form and with any formatting was previously observed and listed.

Once you have prepared the email address, you can use any hashing tool that creates SHA1 hashes. Make sure that you hash only the email address, without leading or trailing spaces, carriage returns or linefeeds. If your preferred hashing tool produces extraneous characters around the hash, use just the hash.

How to Use the EBL

1. Extract the email addresses to be checked from the email.
2. Canonicalize the email addresses as described above.
3. Convert the email addresses into SHA1-format cryptographic hashes.
4. Append **`ebl.msbl.org`** to the hash to form the query string, just as you would combine a domain with the zone of a domain-based blocklist.

EBL Details

Zone:	`ebl.msbl.org`
Query String:	`${hash}.ebl.msbl.org`
Response Codes if Listed	
A Record:	`127.0.0.2`
TXT Record:	Brief explanation of the reason the email address was listed.

Spam Filter and MTA Support

A beta test SpamAssassin module, a procmail scirpt, and a Unix bash shell script are available to testers. Please contact the EBL Administrator ("die Eule") at **`<jaeger@msbl.org>`** to obtain a copy of any or all of these tools. As other tools become available, they will be announced on the EBL website.