Bayesian Analysis Inbound

https://campus.barracuda.com/doc/3866703/

For outbound mail, see Bayesian Analysis Outbound.

How Bayesian Analysis Works

Bayesian Analysis is a linguistic algorithm that profiles language used in both spam messages and legitimate email for any particular user or organization. To determine the likelihood that a new email is spam, Bayesian Analysis compares the words and phrases used in the new email against the corpus of previously identified email. Note that Bayesian training works only on messages with 11 words or more. The Barracuda Email Security Gateway only uses Bayesian Analysis after administrators or users classify at least 200 legitimate messages and 200 spam messages.

Global Bayesian Filtering Versus Per-User

The administrator can configure a global Bayesian database, per-user Bayesian databases or disable Bayesian altogether. With the global setting, which is configured on the **BASIC > Spam Checking** page, the administrator trains and maintains one Bayesian database for all users. With the per-user configuration, users must train and manage their own Bayesian databases, which they access from their **PREFERENCES > Spam Checking** page. There are pros and cons to each configuration.

A global Bayesian database is typically more effective than per-user databases because the administrator can maintain and reset it for all to use, thereby providing a more reliable source of Bayesian management. If, however, the Barracuda Email Security Gateway is filtering mail for many domains, the users of which expect to receive different types of email, it could be either difficult or impossible to train the global Bayesian database to identify spam for all users. For example, if one domain for a medical organization typically receives email regarding medical topics, while another domain for a political organization tends to receive political emails and yet another domain is an entertainment site, then what is spam to one domain may be valid email for another on the same Barracuda Email Security Gateway. In this case, per-user Bayesian filtering would make more sense than global.

In most cases, however, it is not practical to enable Bayesian at the user level because maintaining an accurate Bayesian database requires that users to understand the concept of how Bayesian analysis works and how to use it as an effective tool. That said, while sophisticated users may be trained and savvy enough to initially train their own Bayesian database, they may not have the time to spend in their regular work schedule to effectively maintain their Bayesian databases.

Because spammers frequently change tactics and content, Bayesian data can quickly become "stale" if the database is not reset from time to time and new messages consistently classified as spam or not spam in equal numbers. Without this maintenance the users may see false positives resulting in the blocking of good email.

Getting the Best Accuracy From the Bayesian Database

All Bayesian systems rely on the fact that messages classified are not much different than new messages arriving. Over time however, spam messages change drastically and the Bayesian system – while initially able to compensate for the new format – gradually declines in its effectiveness. When this happens new classifications are needed to update the Bayesian database. To keep a Bayesian database accurate:

- For a global Bayesian database, the administrator should periodically (every 6 months or so) clear it out by resetting it from the **BASIC > Spam Checking** page, then, from the **BASIC > Message Log** page, marking at least 200 messages as either Spam or Not spam using the buttons on the page. Bayesian filtering will NOT take effect until 200 or more of each spam and not-spam messages are marked as such.
- For each per-user database, the user should reset their own Bayesian database and follow up with marking 200 or more messages as spam or not spam, either in their quarantine inbox (**QUARANTINE > Quarantine Inbox** page) or from their regular email client if they have installed the Barracuda Outlook add-in (see below).

When to Use Bayesian Analysis

Barracuda Networks does not recommend using Bayesian filtering in most circumstances. With Energize Updates constantly updating the Barracuda Email Security Gateway with protection against the latest spam and virus threats, spam accuracy should not be an issue for most organizations.

A case for using Bayesian Analysis would depend on the following:

- You are using global Bayesian as opposed to per-user, and the users in the organization tend to be a homogenous population with regard to the kind of content considered to be ‘valid’ email versus spam. This situation would make it easier for an administrator to "train" the global Bayesian database as to what is spam and what is not spam for the organization.
- Your organization requires a very high granularity of accuracy for identifying spam.
- If enabling Bayesian at the per-user level, users are sophisticated and can be trained to properly identify ‘valid’ messages versus spam so as to train the Bayesian database, and are willing to consistently mark BOTH ‘valid’ messages and spam messages in equal numbers so as

to maintain the Bayesian database.
- The administrator and/or users are disciplined about resetting the Bayesian database(s) on a regular basis and re-initializing with 200 each of marked spam and not spam messages to ‘keep current’ with new spam techniques over time.

Barracuda Outlook Add-in

If both per-user quarantine and per-user Bayesian are enabled, on the Barracuda Email Security Gateway 300 and higher, the administrator can choose to allow users to download an add-in that allows messages to be classified as Spam or Not Spam directly from their email client. Users must have a quarantine account on the Barracuda Email Security Gateway to use the add-in. For information about automatically or manually creating quarantine accounts for users, see Creating and Managing Accounts. For more information about the Barracuda Outlook Add-in, see the Barracuda Outlook Add-In Deployment Guide.

Bayesian Analysis on Clustered Systems

When the Barracuda Email Security Gateway is clustered, resetting the Bayesian database must be done on each system individually. However, messages classified as *SPAM* or *NOT SPAM* will synchronize across the clustered systems.

Bayesian Poisoning

Some spammers will insert content in messages intended to bypass spam rules, such as excerpts of text from books or other content that may look "legitimate" in order to fool spam filtering algorithms. This tactic is called Bayesian Poisoning and could reduce the effectiveness of a Bayesian database if many of these messages are marked as either spam or not spam. The Barracuda Networks Bayesian engine is, however, very sophisticated and protects against Bayesian Poisoning if administrators or users consistently maintain their databases.