

Windows Eudora Junk Mail Architecture

February 16, 2006

1 Introduction to Eudora Junk Mail Handling

Eudora implements spam or junk mail handling by allowing any number of EMSAPI plugins to assign junk scores to email messages. Messages that are scored high enough to be considered junk are stored in a special mailbox named “Junk”. All of this functionality is controllable through user settings.

Junk handling is not available in sponsored or light mode, though certain artifacts of this feature will appear even in these modes. This document refers to how the feature works in paid mode. At the end of the document we discuss the differences in sponsored and light modes.

2 Message Data

There are three crucial pieces of data that Eudora must keep track of for purposes of junk handling. The CSummary class keeps track of the junk score assigned to a message (in the m_ucJunkScore data field) as well as the ID of the plugin that generated that score (in m_ulJunkPluginID). It also keeps track of whether the junk score was assigned by a plugin or was manually specified (in m_bManuallyJunked).

3 The Junk Mailbox

Eudora uses a special mailbox named “Junk” to hold messages that are considered junk.

3.1 POP Junk Mailbox

The local (POP) junk mailbox has the same basic properties as the other special local mailboxes (“In”, “Out” and “Trash”), namely that it must always exist, it cannot be deleted or renamed and it always remains in a fixed position at the top of the mailbox tree. Internally its special state is noted in the CTocDoc class by setting its m_Type field to MBT_JUNK.

3.2 IMAP Junk Mailbox

Each IMAP account has its own “Junk” mailbox on the server which behaves in much the same way as the “Trash” mailbox does for accounts which use “fancy trash.” As with the “Trash” mailbox there are fewer special qualities for these mailboxes as compared to

their POP counterparts. In the end they are just IMAP mailboxes and thus can be renamed or deleted on the server which will cause Eudora to recreate them as needed. There is no data field to indicate that an IMAP mailbox is in any way special (the `m_Type` field is set to `MBT_IMAP_MAILBOX` for all IMAP mailboxes). Eudora determines whether an IMAP mailbox is special based on the name and location of the mailbox.

4 Junk Scoring

Junk scoring is performed by any junk handling EMSAPI plugins (or “translators”) Eudora finds when it launches. Eudora accesses the junk translators by calling `CTranslatorManager::CallJunkTranslators()`. The single `CEudoraApp` object contains a reference to a `CTranslatorManager`. The code making the call to `CallJunkTranslators()` generally takes the form: `GetTransMan()->CallJunkTranslators()`.

When a message is passed through each junk translator for scoring, Eudora keeps track of the highest score and which plugin gave that score. When finished, Eudora assigns the message the highest score and the plugin ID.

The context parameter of `CallJunkTranslators()` determines the way in which the translators will be used. The following four values can be used:

- `EMSFJUNK_SCORE_ON_ARRIVAL` – Indicates the specified message is being scored on message arrival.
- `EMSFJUNK_RESCORE` – Indicates that the specified message is being rescored.
- `EMSFJUNK_MARK_IS_JUNK` – Indicates that the specified message is being declared as junk.
- `EMSFJUNK_MARK_NOT_JUNK` – Indicates that the specified message is being declared as not junk.

Additionally, the bit flag `EMSFJUNK_USER_INITIATED` can be logically OR’ed with the above to indicate that the action was user initiated.

5 Processing Mail

All incoming mail is passed through the junk processor before being passed through the filters. The user also has the ability to recalculate the junk score at a later time and to manually declare a message to be either junk or not junk.

5.1 Incoming POP Mail

Once incoming mail has been downloaded it is processed in the method `CPOP::GetMailFromSpool()`. This method first calls `DoProcessJunkMail_()` to do any junk processing. Messages which are considered junk are processed in this method and

are not passed through the filters. If there are messages which were not considered junk CPOP::GetMailFromSpool() then calls DoFilterMail_() to perform filtering on those messages.

5.2 Incoming IMAP Mail

The basic junk handling process for IMAP is essentially the same as POP with the post-download processing being done in the method CImapChecker::PostProcessSummaries(). This method calls CImapChecker::DoJunkProcessing() to do any junk processing then calls CImapChecker::DoPostFiltering() to filter any messages not considered junk.

There are, however, a couple key differences with IMAP.

Proper junk handling requires that the IMAP server supports UIDPLUS. Without UIDPLUS there is no way to preserve the junk score information when a message is transferred between mailboxes. For example, when a message is downloaded to the Inbox and is scored as junk it will be moved to the Junk mailbox where the user would expect the junk score to be preserved. UIDPLUS is necessary to correctly associate the copy of the message in the source mailbox with the copy in the destination mailbox.

Another issue is the option to download varying portions of an IMAP message. If a message is downloaded headers only then junk scoring will be based on only the header data that has been downloaded. If the full message is downloaded at a later time then the message will be rescored using the entire message. This processing happens in CImapDownloader::CheckJunkScore() which is called from a variety of locations within CImapDownloader.

A third issue arises if a message appears in the Junk mailbox by a method other than by a transfer on the local machine. For example, if the message was placed in the Junk mailbox on another machine it will appear in the Junk mailbox with no junk score. In this case, Eudora runs the message through the junk scoring mechanism. It should be noted that the junk score might be completely different than on a different machine since the score is determined by plugins and different installations of Eudora might have different sets of plugins or the same plugins might have been set up or “trained” differently and might produce different scores on different machines. If a message appears in the Junk mailbox and is given a score below the junk threshold score then the message is assigned a score of 100, indicating that the message is indeed junk by virtue of it having been placed in the Junk mailbox on a different machine.

5.3 Recalculating the Junk Score

As well as processing incoming mail for junk Eudora supports user-initiated junk scoring. This is useful for messages that were downloaded with a version of Eudora that did not support junk scoring or for doing recalculations if there was a change to one of more of

the junk scoring plugins. This is accomplished by calling the static method `CJunkMail::ReprocessOne()` and passing the `CSummary` object for the message to be rescored.

5.4 Whitelisting

Eudora supports the option of whitelisting any email messages whose senders are in the address book. Any messages that are whitelisted are given a score of 0. For POP messages whitelisting is all performed inside the main junk processing method `CJunkMail::ProcessOne()`. For IMAP whitelisting is performed by directly calling one of the three static `CJunkMail::IsWhitelisted()` functions.

6 Manually Changing Message Status

When the user manually changes a message's junk status the code calls `CJunkMail::DeclareJunk()` which takes a boolean parameter to indicate whether the message is being declared as junk or not junk. Declaring a message to be junk assigns it a score of 100 and moves it to the Junk mailbox. Declaring a message to be not junk assigns it a user-defined score (default of 0), moves the message to the In or Inbox then passes it through the incoming filters.

7 Aging of Junk Messages

Messages in the Junk mailbox are eventually aged out of the mailbox, assuming the user has not turned off that option. On every quit the function `TrimJunk()` is called. This function checks the logic to see if it is an appropriate time to check for junk to be trimmed. If the timing is correct it calls `CTocDoc::TrimJunk()` on the POP Junk mailbox then calls the function `TrimImapJunk()` to trim junk from any IMAP Junk mailboxes.

For users who quit Eudora infrequently `TrimJunk()` is called before mail checks. When called in this way the timing logic is much less stringent. It will only trim junk when three times the desired interval has elapsed. For example, if the user specifies that junk should be trimmed once a day, junk will be trimmed on a mail check every three days if the user has not quit Eudora in the meantime.

Trimming of messages in the Junk mailbox(es) can also be user initiated. It can be initiated from a menu but it also happens when the user compacts the Junk mailbox. User initiated trimming also calls the function `TrimJunk()` with a parameter of true to indicate that the trimming is requested by the user so the timing logic is bypassed.

8 Sponsored and Light Modes

The Junk mailbox is created even if Eudora is not running in paid mode and the user is free to filter or manually move messages into the Junk mailbox. No junk scoring will be performed. The one junk feature that remains available in sponsored and light modes is aging off of old junk messages from the Junk mailbox.

9 Future Considerations

- Junk column
 - We will likely represent the junk score with icons for scale.
 - What should the icon be for the column label? (We might use the same icon as the icons for scale.)
- The status column will display an icon indicating that a given junk message will soon be aged off (to avoid the addition of yet another column for such a specific use).
 - The age icon will be displayed instead of the normal status icon, and the user will be able to sort by it. However the user will not be able to change the status of such a message (probably the "Change status" menu will be displayable but disabled).
 - When sorted by status, aged items will be grouped together
 - Will the user be able to change the status of such a message? What would changing status do?
 - One possibility would be to prevent them from being aged off.
 - Another possibility would be to reset the age clock.
 - A third possibility would be simply to change the status, but not give any visible indication.
 - A fourth possibility would be to make the delete indication something that could be combined with the status indicator. I could perhaps ask Sid for possibilities.
 - What icon do we use? An X, a trash can, or something else?
 - Given that we won't be likely to fit in the number of days until deletion in this amount of space:
 - How many days ahead of time do we display it?

Do we use different icons based on the number of days (e.g. changing color to red for imminent deletion or a little tiny "hangman" that slowly gets a new body part added as each day passes ;-).

- We might want to be careful of giving a plugin credit merely for scoring higher than other plugins. After all if half the plugins in use recognize something as junk - why give preferential treatment to the highest scoring plugin just because it assigns bigger numbers? To avoid this which plugin is assigned credit might be randomly chosen from among plugins that have assigned a score greater than the threshold.
- Stats for spam - including breaking down the stats by plugin.
- "Show Junk Score(s)" command in the "Message" menu (on Windows also in right click context menu) - allows the user to see exactly what each plugin says (I purposely used the word "says" we could re-ask the plugins when the user chooses this option via ON_RESCORE calls).
- Expose the hidden setting that allows the user to specify another mailbox to age off to (with a special value of "-", which will mean nuke messages when aging them off). Probably use radio buttons:
 - Age old junk mail to [Trash] mailbox where [Trash] is a button that brings up a dialog that allows the selection of "Trash" and any other non-special mailbox (i.e. no In, Out, or Junk).

- Delete old junk mail PERMANENTLY.
- What do we do when the user raises or lowers the Junk threshold?
 - User raises Junk threshold - do we run through the junk mailbox and move messages out that no longer qualify, or does the user choose the "Recalculate Junk Score" menu item?
 - User lowers Junk threshold. We do nothing. User can select messages and choose the "Recalculate Junk Score" menu item.
- There has been discussion of adding "Junk"/"Not Junk" to the message popup.
- Steve thinks we need a button on message windows, about where Queue normally goes, that says "Junk" or "Not Junk" (see "Junk Command and Preview Pane" later in this document for placement with the preview pane). If the user is using the "Junk" mailbox, it says "Junk" if not in the Junk mailbox, "Not Junk" if it is. If the user is not using the Junk mailbox, the decision of what to name the button could be made on the Junk threshold. Shift/option clicking this button could let the user assign a specific score. Unclear what assigning a specific score means to plugins?
- Rescoring might be via a button in the mailbox window next to compact (see the first option under "Junk Command and Preview Pane" later in this document for possible placement).
- Ask spam authors for an appropriate interval? E.g. ticks at 25/50/75, and we recommend they return values like 12/36/64/86). Tick labels could be worded to indicate the tradeoff between false positives and missed junk mail.
- Another possibility is a general setting that says if I have a stub that is older than x days - download and delete it.
- Another possible hidden setting would be a threshold value that indicates how junky partially fetched mail needs to be before it would be automatically deleted from the server.