SpamdSyslogFormat

spamd Syslog Format

SpamAssassin 2.x Format

The traditional spamd syslog format looks like this:

```
Aug 15 06:47:37 dogma spamd[22796]: identified spam (21.6/5.0) for jm:1012 in 3.1 seconds, 1098 bytes.
```

This details the time, hostname, process ID in normal syslog format; then whether the message was spam or nonspam; the points and threshold used; the username and uid; the time taken to scan the message; and the size of message in bytes. This is what a ham line looks like:

```
Aug 15 07:17:57 dogma spamd[7666]: clean message (-1.1/5.0) for jm:1007 in 0.4 seconds, 1800 bytes.
```

SpamAssassin 3.0.0 Format

As of SpamAssassin 3.0.0, spamd also produces syslog lines for each scanned message that are designed to be machine-parseable and extensible. It looks like this:

```
Aug 15 06:47:37 dogma spamd[22796]: result: Y 21 - BAYES_99,BODY_ENHANCEMENT2,DC
C_CHECK,DNS_FROM_RFC_ABUSE,DNS_FROM_RFC_POST,FORGED_RCVD_HELO,HTML_70_80,HTML_BA
CKHAIR_2,HTML_FONT_BIG,HTML_FONT_SIZE_LARGE,HTML_MESSAGE,HTML_OBFUSCATE_05_10,MI
ME_HTML_ONLY,MISSING_SUBJECT,MSGID_FROM_MTA_ID,RCVD_IN_NJABL_DUL,RCVD_IN_RFC_IPW
HOIS,RCVD_IN_SBL,RCVD_IN_SORBS_DUL,URIBL_SBL,URIBL_SC_SURBL,URIBL_WS_SURBL scant
ime=3.1,size=1098,mid=<20040815054728.0485A3104AC@amgod.boxhost.net>,bayes=1,aut
olearn=spam
```

(in this example, I've broken it into multiple lines by inserting newlines, but in syslog, it's all one long line.)

Again, the normal syslog data (time, hostname, process ID) is present. In addition, the line contains "result: ", followed by what is basically the same format used by the MassCheck tool. This means: "Y" for spam or "." for nonspam; the points, rounded to an int; a skipped field ("-"); the list of tests hit, comma-separated; then a comma-separated list of name=value pairs.

The name=value pairs may appear in *any* order, and new ones may be introduced without warning; but they'll always be comma-separated. They include the following items:

- scantime: the time in seconds taken to scan this message.
- size: the size in bytes of the message.
- mid: the Message-ID.
- bayes: the score output by the Bayes classifier (NOTE: this could be in scientific notation, e.g. '2.22044604925031e-16', or as a normal number: '0.00437882812836221', or even '1').
- autolearn: whether the message was autolearned or not (*ham*, *spam*, *no*, *failed*, *disabled*, *unavailable*).

Here's another example, this time from a ham mail:

```
Aug 16 05:40:53 dogma spamd[27160]: result: . -2 - BAYES_00,SPF_HELO_PASS,
SPF_PASS scantime=3.6,size=3959,mid=<200408161446.01084.zgampe@redhat.com
>,bayes=2.22044604925031e-16,autolearn=ham
```