DnsblAccuracy082005

DNS Blocklist Accuracy Figures (as of July 2005)

Many people, whether they use SpamAssassin or not, find accuracy figures for DNSBLs to be useful. Here are accuracy figures for the DNS blocklists included in SpamAssassin 3.1.0, as measured during our July rescoring run. We use the following techniques to assure high accuracy on these figures:

- some hits are recorded from 'live' data at the time the messages were received, not post-facto testing (using 'mass-check --reuse')
- there were 9 people contributing their hit data, from a variety of geographical locations and organisational types
- both Ham and Spam hitrates are measured, and the corpora were hand-verified in advance
- the corpora use (relatively) fresh mail, received between January 2004 and July 2005

123778 spam messages and 53091 ham messages were used:

```
OVERALL%   SPAM%     HAM%     S/O    RANK   SCORE  NAME
 176869   123778    53091    0.700   0.00    0.00  (all messages)
100.000  69.9829  30.0171    0.700   0.00    0.00  (all messages as %)
```

These were randomly chosen from all contributors' logs (see below). First off, the DNS blocklists.

Note – the sorting is by mass-check's RANK metric, which puts 'better' results near the top, and the results are in HitFrequencies format. The 'S/O', 'SPAM%', and 'HAM%' columns are the most important metrics; S/O values approaching 1.0 are best.

```
 17.449  24.9285   0.0113    1.000   0.97    3.90  RCVD_IN_XBL
  3.841   5.4824   0.0132    0.998   0.88    2.16  RCVD_IN_SORBS_SOCKS
  5.865   8.3690   0.0283    0.997   0.88    3.16  RCVD_IN_SBL
  9.438  13.4652   0.0490    0.996   0.84    2.23  RCVD_IN_WHOIS_INVALID
  2.237   3.1839   0.0301    0.991   0.79    0.02  RCVD_IN_SORBS_MISC
 27.913  39.8423   0.0998    0.998   0.76    2.60  RCVD_IN_DSBL
  4.914   6.9883   0.0772    0.989   0.74    0.02  RCVD_IN_SORBS_HTTP
  0.914   1.3015   0.0113    0.991   0.72    2.77  RCVD_IN_NJABL_SPAM
  7.692  10.9486   0.0979    0.991   0.72    2.43  RCVD_IN_WHOIS_BOGONS
 22.130  31.5662   0.1300    0.996   0.71    2.05  RCVD_IN_SORBS_DUL
 10.642  15.1449   0.1450    0.991   0.67    0.72  RCVD_IN_NJABL_PROXY
 18.739  26.6946   0.1921    0.993   0.61    1.95  RCVD_IN_NJABL_DUL
  0.345   0.4888   0.0094    0.981   0.57    0.20  RCVD_IN_SORBS_SMTP
  5.309   7.5062   0.1865    0.976   0.56    1.46  RCVD_IN_SORBS_WEB
 16.300  23.1463   0.3372    0.986   0.53    1.56  RCVD_IN_BL_SPAMCOP_NET
  0.166   0.0016   0.5481    0.003   0.47   -2.20  RCVD_IN_IADB_VOUCHED
  0.161   0.0016   0.5330    0.003   0.47   -4.30  RCVD_IN_BSP_TRUSTED
  0.096   0.1365   0.0000    1.000   0.41    1.00  RCVD_IN_WHOIS_HIJACKED
  0.118   0.1656   0.0075    0.956   0.41    0.10  RCVD_IN_NJABL_RELAY
  0.040   0.0533   0.0094    0.850   0.32    0.26  RCVD_IN_SORBS_ZOMBIE
  0.000   0.0000   0.0000    0.500   0.28    0.00  RCVD_IN_SORBS_BLOCK
  0.000   0.0000   0.0000    0.500   0.28    0.00  RCVD_IN_NJABL_MULTI
  0.000   0.0000   0.0000    0.500   0.28    0.00  RCVD_IN_NJABL_CGI
```

URI blocklist lookups, against SURBL and SBL:

```
 17.882  25.5522   0.0000    1.000   1.00    4.50  URIBL_SC_SURBL
  9.684  13.8369   0.0019    1.000   0.98    3.81  URIBL_AB_SURBL
 34.260  48.9497   0.0132    1.000   0.98    4.09  URIBL_JP_SURBL
 36.356  51.9317   0.0414    0.999   0.90    3.01  URIBL_OB_SURBL
 30.956  44.1605   0.1695    0.996   0.66    2.14  URIBL_WS_SURBL
  0.266   0.3805   0.0000    1.000   0.56    2.80  URIBL_PH_SURBL
 22.415  31.8425   0.4370    0.986   0.49    1.64  URIBL_SBL
```

SPF lookups:

```
  3.437   4.8942   0.0396    0.992   0.80    1.38  SPF_SOFTFAIL
  1.006   1.4292   0.0207    0.986   0.71    2.43  SPF_HELO_SOFTFAIL
  2.550   3.5717   0.1676    0.955   0.53    1.14  SPF_FAIL
  2.297   3.2090   0.1695    0.950   0.52    1.07  SPF_NEUTRAL
  1.796   2.5029   0.1488    0.944   0.51    0.00  SPF_HELO_FAIL
  0.935   1.2724   0.1488    0.895   0.43    0.00  SPF_HELO_NEUTRAL
  5.334   2.5925  11.7252    0.181   0.21   -0.00  SPF_HELO_PASS
  3.267   2.6241   4.7654    0.355   0.10   -0.00  SPF_PASS
```

RFC-ignorant, testing against the envelope sender's domain:

```
  3.038   4.3352   0.0132    0.997   0.86    2.60  DNS_FROM_RFC_DSN
  1.174   1.6715   0.0151    0.991   0.75    1.94  DNS_FROM_RFC_BOGUSMX
  3.590   5.0607   0.1620    0.969   0.57    1.45  DNS_FROM_RFC_WHOIS
 13.930  19.7071   0.4615    0.977   0.47    1.71  DNS_FROM_RFC_POST
 12.120  16.7154   1.4051    0.922   0.34    0.20  DNS_FROM_RFC_ABUSE
```

other network rules:

```
  1.898   2.7081   0.0094    0.997   0.82    3.20  NO_DNS_FOR_FROM
  1.449   2.0593   0.0245    0.988   0.74    1.51  DNS_FROM_SECURITYSAGE
  7.200  10.0898   0.4615    0.956   0.44    0.23  DNS_FROM_AHBL_RHSBL
```

More details of the source mass-check log files and test procedure can be read in SpamAssassin bug 4505. the full list of freqs can be found in the STATISTICS-set3.txt file in the 3.1.0 release. Here's a list of the data files used. Note that only a randomly-chosen one tenth of each file was used.

Use of --reuse for real-time network results: confirmed on: 4 users (bmenschel, jm, parker, cthielen); confirmed off: 1 user (duncf); unknown: 4 users (bzoetekouw, misak, quinlan, theo).

```
bash-3.00$ ls -l /home/corpus-rsync/corpus/submit/
total 2839184
-r--r--r--   1 rsync    rsync    7967268 Jul 16 18:18 ham-bayes-net-bzoetekouw.log
-r--r--r--   1 rsync    rsync    1987090 Jul 16 14:49 ham-bayes-net-cthielen.log
-r--r--r--   1 rsync    rsync    23284450 Jul 24 08:04 ham-bayes-net-daf.log
-r--r--r--   1 rsync    rsync    51469171 Jul 19 02:26 ham-bayes-net-jm.log
-r--r--r--   1 rsync    rsync    45026386 Jul 19 02:27 ham-bayes-net-jm2.log
-r--r--r--   1 rsync    rsync     294744 Jul 25 18:57 ham-bayes-net-misak.log
-r--r--r--   1 rsync    rsync    22130676 Jul 27 04:17 ham-bayes-net-parkerm.log
-r--r--r--   1 rsync    rsync    14056970 Jul 27 19:37 ham-bayes-net-quinlan.log
-r--r--r--   1 rsync    rsync    8603737 Jul 27 17:01 ham-bayes-net-rod.log
-r--r--r--   1 rsync    rsync    28410747 Jul 27 02:34 ham-bayes-net-theo.log
-r--r--r--   1 rsync    rsync    62685697 Jul 16 18:22 spam-bayes-net-bzoetekouw.log
-r--r--r--   1 rsync    rsync    11891366 Jul 16 14:50 spam-bayes-net-cthielen.log
-r--r--r--   1 rsync    rsync    96553037 Jul 24 08:09 spam-bayes-net-daf.log
-r--r--r--   1 rsync    rsync    28662170 Jul 19 02:28 spam-bayes-net-jm.log
-r--r--r--   1 rsync    rsync    209202453 Jul 19 02:34 spam-bayes-net-jm2.log
-r--r--r--   1 rsync    rsync     243487 Jul 25 18:57 spam-bayes-net-misak.log
-r--r--r--   1 rsync    rsync    39357821 Jul 27 04:19 spam-bayes-net-parkerm.log
-r--r--r--   1 rsync    rsync    41987897 Jul 27 19:39 spam-bayes-net-quinlan.log
-r--r--r--   1 rsync    rsync    97404262 Jul 27 17:03 spam-bayes-net-rod.log
-r--r--r--   1 rsync    rsync    358576609 Jul 27 02:34 spam-bayes-net-theo.log
```