



# DCC client come plugin di SpamAssassin

Versione 1.1 – 10 Dicembre 2004

Autori: Fulvia Costa, Michele Michelotto – INFN Sezione di Padova<sup>1</sup>

## **Introduzione**

SpamAssassin è un progetto open source che cerca di identificare nei mail trattati da un mail server i messaggi di tipo SPAM. SpamAssassin assegna ad ogni mail un punteggio dato dalla somma dei punteggi ottenuti confrontando il messaggio con un insieme di regole. Se un messaggio ottiene un punteggio superiore ad una certa soglia questo viene marchiato come Spam. L'amministratore del sistema può decidere se consegnare o meno il messaggio marchiato, può decidere di consegnarlo in un folder speciale o infine semplicemente consegnarlo cambiandone l'header in modo da facilitare all'utente le operazioni di filtro.

Le regole sono, per la maggior parte, di tipo euristico, cioè assegnano un punteggio (positivo o negativo) se il mail contiene nell'header o nel body alcune caratteristiche tipiche di uno spam

Altre regole molto usate sono le statistiche bayesiane. Ogni sito raccoglie, un certo numero di mail di tipo spam e di tipo non-spam (ham) da cui viene ricavato una tabella ottimizzata per quel sito di parole (in realtà token) presenti con più probabilità in uno spam o in mail regolare. Il filtro bayesiano poi assegna un punteggio aggiuntivo al mail calcolando la presenza in un certo mail dei diversi tipi di token

In questo documento vogliamo introdurre un metodo aggiuntivo chiamato DCC, Distributed Checksum Clearinghouse. Questo è un sistema di migliaia di client e centinaia di server che contano dei checksum relativi a oltre 100 milioni di messaggi al giorno. Il conteggio viene usato per identificare e marciare gli spam di tipo bulk (UBE, Unsolicited Bulk Email).

L'idea alla base di DCC è che se i destinatari possono confrontare tra di loro i mail che ricevono allora possono discriminare quelli inviati in modo bulk, cioè lo stesso messaggio a migliaia di destinatari.

---

<sup>1</sup> Questo documento è stato redatto nell'ambito del working group GARR chiamato wg-sec-mail

## ***Prerequisiti***

In questo documento viene descritta l'installazione su un server SMTP su cui sono installati **sendmail** e **SpamAssassin 2.6x**

Nel nostro caso SpamAssassin viene chiamato da MimeDefanger.

## ***Dove si trova il software***

Il software si prende all'indirizzo del progetto: <http://www.rhyolite.com/anti-spam/dcc> alla seguente URL: <http://www.rhyolite.com/anti-spam/dcc/dcc-tree/INSTALL.html> . Bisogna scaricare il file **dcc-dccproc.tar.Z** che contiene tra l'altro le interfacce procmail e alcune utilities come cdcc, le pagine del manuale in linea e la documentazione.

## ***Installazione***

Per prima cosa si deve scompattare l'archivio in una directory:

```
% tar zvfz dccproc.tar.Z  
  
dcc-dccproc-1.2.57/CHANGES  
dcc-dccproc-1.2.57/FAQ.html  
dcc-dccproc-1.2.57/FAQ.txt  
...  
dcc-dccproc-1.2.57/win32.makinc1  
dcc-dccproc-1.2.57/win32makinc2  
dcc-dccproc-1.2.57/configure  
%
```

Si entra nella directory :

```
% cd dcc-dccproc-1.2.57
```

Se non si usa l'interfaccia DCC-sendmail presente nel pacchetto **dccm** ora dobbiamo configurare, costruire e installare i programmi DCC. Per la configurazione migliore per il vostro sito leggere i consigli di installazione nella man page di DCC (Installation Considerations)

Nel nostro caso la configurazione scelta è la seguente:

```
%./configure --disable-server --disable-dccm  
creating cache ./config.cache  
Rhyolite Software DCC 1.2.5702 dccproc  
checking for cc... cc  
checking for gcc... (cached) cc
```

```
checking wether the C compiler (cc ) works... yes
...
checking for IPv6... no getipnodebyname()
updating cache ./config.cache
creating ./config.status
```

con la quale richiedo l'installazione senza server e senza interfaccia **dccm**

Ora posso passare alla fase build e all'installazione:

```
% make
```

con questo comando compilo i diversi file sorgenti e mi creo gli archivi e i file eseguibili

```
%make install
```

con questo comando invece installo nelle directories prescelte i files che ho appena costruito.

Nel mio caso i file di configurazione si trovano in **/var/dcc**, i binari in **/usr/local/bin** e **/var/dcc/libexec**.

C'è una lista predefinita di servers che può essere consultata con il comando:

```
% cdcc info
```

Questa lista predefinita contiene i servers DCC messi a disposizione dal progetto DCC per l'accesso anonimo. Si possono aggiungere altri server ed è consigliabile aggiungere che nella topologia della rete TCP/IP siano più vicini al proprio mail server per esempio sulla rete GARR. Inoltre conviene contattare i gestori di questi siti per chiedere di avere accesso non anonimo (protetto da password) per avere nelle richieste DCC priorità più elevata.

## ***Integrazione in SpamAssassin***

Si aggiungono le seguenti linee al file di configurazione di SpamAssassin:

```
use_dcc 1
dcc_path /usr/local/bin/dccproc
dcc_home /var/dcc
dcc_options -Rw whiteclnt
```

Con la prima linea comunichiamo a SpamAssassin che vogliamo usare il plugin DCC

Con la seconda linea e terza linea forniamo a SpamAssassin il path dell'eseguibile e dei files di configurazione.

Infine la quarta riga riporta le opzioni di **dccproc**. In particolare "**w whiteclnt**" abilita l'uso della whitelist nel file **whiteclnt**

## ***Il daemon dccifd***

Per evitare di lanciare ad ogni mail il processo **dccproc** è consigliabile abilitare l'uso del daemon **dccifd** modificando i file di configurazione di SpamAssassin:

```
use_dcc 1
dcc_dccifd_path /var/dcc/dccifd
dcc_home /var/dcc
```

Con la prima linea comunichiamo a SpamAssassin che vogliamo usare il plugin DCC  
Con la seconda linea e terza linea forniamo a SpamAssassin il path dell'eseguibile e dei files di configurazione.

In questo file c'è un settore dedicato a **dccifd**, in particolare deve essere:

```
DCCIFD_ENABLE=on
```

Poi si possono inserire le opzioni; nella nostra configurazione le opzioni sono:

```
DCCIFD_LOGDIR=
DCCIFD_WHITECLNT="$DCCM_WHITECLNT"
DCCIFD_USERDIRS=
DCCIFD_LOG_AT=
```

Per lanciare il daemon si usa lo script:

```
/var/dcc/libexec/start-dccifd che crea il socket /var/dcc/dccifd
```

## ***Modifica della lista dei server:***

Si fa con con il comando **cdcc** con la seguente sintassi:

```
% cdcc "add dcc.to.infn.it RTT-1000 ms anon"
```

Con cui, per esempio, si aggiunge il server DCC dell'INFN di Torino:

Se abbiamo avuto dall'amministratore dell'INFN di Torino un id e password per il server al posto di **anon** si mettono **id** e **password**

```
% cdcc "add dcc.to.infn.it RTT-1000 ms 32768 2449xxxx680"
```

DCC decide dinamicamente quale server usare misurando il Round Trip Time verso i diversi server presenti nella sua mappa.

L'opzione **RTT-1000 ms** permette di pilotare la scelta del server, togliendo 1000 millisecondi al Round Trip Time misurato per quel server.

## ***Colloquio con il server***

Il client DCC deve comunicare con il server DCC quindi è necessario che la porta UDP 6277 sia aperta in entrambe le direzioni tra le due macchine. La porta 6277 è quella di default e può essere cambiata.

## ***White list***

Il file che contiene le configurazioni si trova in **/var/dcc**, nel nostro caso si chiama **whiteclnt**. È stato modificato per alcuni server di mailing list che noi consideriamo fidati (o comunque legittimi), per esempio:

```
ok env_from owner-press-release-other4@lists.hp.nasa.gov
```

il che vuol dire che vengono accettati tutti i mail che hanno nell'envelope l'indirizzo riportato.

Nota Bene: se la white list funziona deve comparire il file

```
/var/dcc/whiteclnt.dccw
```

che è una hash table aggiornata molto frequentemente.

## ***Vantaggi nell'uso di DCC***

L'uso di DCC non cambia radicalmente l'efficacia di SpamAssasin. Abbiamo osservato che tra il 40% e il 60% dei mail vengono riconosciuti come UBE dal server DCC. Tuttavia con i punteggi di default di SpamAssasin solo nel 3% dei casi il punteggio di DCC è stato determinante per decidere passare la soglia (che di default è 5.0).

Se gli spam non identificati sono attorno al 10% questo vuol dire comunque che è possibile passare dal 10% al 7% di inefficienza.

Naturalmente è possibile aumentare il peso del test DCC se si è molto confidenti nel fatto che DCC non fornisca falsi positivi. In questo caso è importante che venga messe nella whitelist tutte le mailing list ad alto volume (per es. Bugtraq) che potrebbero essere segnalate come UBE.

