INFN ISTITUTO NAZIONALE DI FISICA NUCLEARE

Sezione di Genova

INFN/TC-07-02
31 Gennaio 2007

CONFIGURAZIONE DEI SERVIZI DI POSTA ELETTRONICA PER LA SEZIONE INFN DI GENOVA

Alessandro Brunengo[1], Mirko Corosu[1]

[1)]*INFN-Sezione di Genova, Dip. Di Fisica Università di Genova, I-16146 Genova, Italy*

Abstract

Viene di seguito descritta la scelta degli applicativi e delle configurazioni hardware e software per dotare la sezione INFN di Genova di un servizio di posta elettronica affidabile e dotato delle funzionalita' ritenute idonee al lavoro dell'utenza del Servizio Calcolo locale.

La realizzazione dei vari servizi descritti sono il risultato di approssimazioni successive che ha raggiunto un livello di stabilita' e di prestazioni idoneo alle funzionalita' richieste, e potenzialmente scalabile entro un ordine di grandezza senza sostanziali rivoluzioni della architettura. Come vedremo i requisiti imposti al sistema, in particolare quello di indipendenza delle macchine coinvolte da altri servizi che non fossero l'infrastruttura di rete, ha reso necessaria una configurazione opportuna anche di servizi di rete potenzialmente indipendenti dalla posta elettronica, quali DNS ed authentication service.

Le scelte operate non sono le uniche possibili, ma sono tra quelle ritenute idonee a dotare la sezione di un servizio che offre le funzionalita' desiderate, trovando un equilibrio tra le esigenze di ridondanza e quella di non eccedere nella complessita' del sistema, sia per contenere i costi che per limitare il man power necessario alla gestione hardware e software.

Published by ***SIS–Pubblicazioni***
Laboratori Nazionali di Frascati

1 INTRODUZIONE

Il servizio di posta elettronica ha assunto nel corso degli ultimi anni una importanza fondamentale per l'attivita' di ricerca dell'INFN, divenendo uno strumento di lavoro irrinunciabile per la comunicazione tra i ricercatori.

Nel corso degli anni e' anche aumentata la necessita' di fornire servizi aggiuntivi al semplice trasporto dei messaggi di posta, quali servizi automatizzati di archiviazione, reiezione ed avvisi di assenza, servizi di gestione decentrata di mailing list ed archiviazione dei messaggi inviati alle liste. Oltre a questo, il crescente traffico di messaggi indesiderati (SPAM) ha reso necessaria l'implementazione di filtri efficienti, cosi' come la diffusione dei virus trasmessi per posta elettronica richiede l'implementazione di filtri antivirus opportuni. Infine, le politiche di sicurezza che proteggono l'interno della rete locale da attivita' illecite e potenzialmente dannose provenienti dall'esterno impongono la configurazione di opportuni metodi di accesso al servizio di posta elettronica qualora l'utente si trovi connesso tramite una rete esterna alla LAN di sezione.

La progettazione di un servizio di posta elettronica idoneo deve prevedere per quanto possibile la ridondanza degli apparati, in modo da garantire un servizio affidabile, e procedure operative di recovery adeguate in occasione di perdita dei dati.

2 DEFINIZIONE DEI REQUISITI

La fase progettuale richiede la definizione dei requisiti che si desiderano soddisfare. Questi coprono diversi aspetti:

- definizione di piattaforma hardware e sistema operativo: questa scelta deve essere idonea all'implementazione dei servizi desiderati, garantendo allo stesso tempo stabilita' per alcuni anni, e non deve costituire un aggravio di lavoro al Servizio Calcolo locale per lo sviluppo di competenze nella gestione del sistema scelto;
- definizione dei servizi: oltre alle funzionalita' minime (servizio di mail relay, servizio di archiviazione delle mail ed accesso alle stesse), e' stato ritenuto opportuno dotare la sezione di servizi accessori quali filtro anti-virus, filtro anti-spam, accesso alle mail via interfaccia Web (WebMail), servizio di archiviazione automatizzata in folder diversi dall'inbox in base a pattern specifici configurabili dall'utente, servizio di list server con delega della gestione delle liste, servizio di relay autenticato; si richiede anche che il filtro anti spam sia configurabile a livello di utente nei parametri di selettivita', azione da intraprendere, white-list e black-list;
- availability: per fornire un servizio fruibile con continuita' si e' deciso di rendere gli host dedicati a queste funzioni indipendenti da servizi di rete forniti da altri host, limitando alla funzionalita' della rete locale e geografica l'unica

dipendenza effettiva; ove possibile, devono essere realizzate configurazioni ridondanti; infine, si richiede che gli host dedicati ai servizi non permettano il login degli utenti per motivi di sicurezza e per impedire che le attivita' di calcolo interattivo o batch degli stessi competano con i processi del servizio di posta nell'utilizzo delle risorse dei calcolatori;

- osservanza degli standard: i protocolli utilizzati devono essere tutti integralmente compatibili con gli standard che definiscono le specifiche a livello internazionale;
- sicurezza: questo requisito impone la configurazione di soli protocolli criptati per l'autenticazione degli utenti;
- monitoraggio ed allarmistica: il progetto deve comprendere il setup di quanto necessario per tenere sotto controllo le funzionalita' del sistema, assieme ad un meccanismo di allarmistica che segnali tempestivamente ogni malfunzionamento; si dovranno anche produrre statistiche di utilizzo delle risorse al fine di programmare opportunamente l'aumento delle stesse; inoltre si dovra' realizzare l'archiviazione dei log file necessari a ricostruire l'accaduto in conseguenza di potenziali problemi;
- backup: in aggiunta alle misure di ridondanza, si dovranno definire dettagliate procedure per il recovery in caso di failure di tutto il sistema (procedure di restart, recovery dei dati, recovery delle configurazioni), o per il recupero di dati persi, sia a seguito di failure hardware, che per erronea cancellazione da parte dell'utenza;
- risorse: il traffico di mail corrente al momento della progettazione e' pari a qualche migliaio di mail al giorno, per un throughput medio inferiore a 100 Kb/s. Risulta evidente che le risorse di rete necessarie saranno determinate dalle esigenze di prestazioni di picco piuttosto che alle esigenze di throughput medio, pur prevedendo un aumento di due ordini di grandezza nel corso dei successivi quattro anni. Per le dimensioni dello storage l'occupazione era allora pari a circa 20 GB. In fase progettuale sara' necessario destinare allo storage delle mail almeno un fattore dieci in piu', cercando se possibile soluzioni scalabili con continuita'.

3 PIATTAFORMA HARDWARE E SISTEMA OPERATIVO

La piattaforma scelta e' x86/linux.

L'utilizzo di CPU x86 e' motivato dal fatto che, a fronte del supporto a sisitemi operativi idonei alla gestione dei servizi richiesti (linux, BDS, Solaris, MS Windows), i costi nei confronti di processori differenti sono generalmente piu' contenuti. Questo permette di pianificare uno svecchiamento dell'hardware a cadenza di tre o quattro anni senza dover affrontare spese eccessive.

Data la natura critica del servizio da offrire, l'hardware deve essere dotato delle seguenti caratteristiche:

- server rack mountable, per essere ospitato nella Sala Macchine della Sezione, in modo da godere di un ambiente climatizzato e controllato ed alimentazione assistita tramite UPS;
- server dotati di doppia alimentazione, alimentati attraverso interruttori differenti del quadro elettrico, per garantire continuita' di funzionamento in caso di failure dell'alimentatore o di sgancio dell'interruttore sul quadro elettrico;
- doppio processore, per garantire maggiore performance;
- le partizoni di sistema devono essere ospitate su doppio disco SCSI serviti da controller RAID in configurazione RAID1; il disco SCSI garantisce maggiore affidabilita' rispetto a soluzioni IDE o SATA, e la configurazione in mirroring protegge dalla failure del disco stesso;
- interfaccia Gigabit Ethernet; l'analisi del traffico medio atteso e' lontano dal saturare il Gigabit Ethernet, ma la prestazione di picco e' sicuramente funzionale a migliorare le prestazioni; inoltre le attuali schede madri in commercio sono quasi tutte dotate di doppia scheda di rete Gigabit Ethenret, quindi questo requisito non comporta alcuna spesa aggiuntiva;
- almeno due Gigabyte di RAM, in modo da garantire ai processi coinvolti idoneo spazio di memoria, e mantenere spazio in RAM per il caching del file system

Il sistema operativo scelto e' una versione ricompilata di RHEL, nota come SLC (Scientific Linux CERN), ampiamente diffusa nell'ambiente della fisica delle alte energie.

Benche' scelte differenti, quali ad esempio BSD, possano offrire a priori migliori garanzie di sicurezza contro accessi non autorizzati, il costante aggiornamento del sistema alle ultime versioni dei pacchetti software installati fornisce soddisfacenti garanzie in questo senso; il meccanismo di aggiornamento puo' essere automatizzato utilizzando tool quali *apt per rpm* o *yum* (software evoluti di gestione dei pacchetti installati via rpm) che gestiscono correttamente pacchetti e dipendenze. Questa scelta inoltre minimizza l'impatto di lavoro sul Servizio Calcolo, il cui personale e' gia' competente sulla gestione dello stesso, e permette di omogeneizzarne il management con quello dedicato alla maggior parte degli altri calcolatori linux presenti in Sezione.

4 SERVIZIO DI MAIL RELAY

Il servizio di mail relay e' quello preposto a ricevere ed inoltrare i messaggi di posta elettronica, sia quelli provenienti dall'esterno e destinati all'utenza locale, sia quelli inviati dall'utenza locale con qualunque destinazione.

Per garantire la compatibilita' con il resto di Internet, il software che implementa questo servizio deve soddisfare le specifiche del protocollo SMTP definito negli standard RFC [1)] (rfc821, rfc974, rfc1123 e l'aggiornamento del rfc2821).

Dal punto di vista logico si possono distinguere due funzionalita': quella di mail relay, dedicata a ricevere tutte le mail provenienti dall'esterno e destinate a caselle di posta locali, e quello di outgoing mail server, preposta a ricevere tutti i messaggi di posta elettronica inviati dalla utenza locale, indipendentemente dalla loro destinazione.

La funzionalita' di mail relay viene realizzata configurando opportunamente una o piu' macchine dedicate ad inoltrare a destinazione tutte le mail indirizzate a ***<destinatario>@ge.infn.it***; tale calcolatore viene reso noto al mondo registrando un opportuno record nel DNS, di tipo MX, che associa al dominio di destinazione della mail l'indirizzo del mail relay. Questo record viene letto da tutti i mail relay del mondo quando devono inoltrare un messaggio a ***<destinatario>@ge.infn.it***. Il protocollo SMTP prevede che si possano associare allo stesso dominio diversi record di tipo MX, eventualmente con differenti priorita', per poter configurare soluzioni di backup o di load balancing.

La funzionalita' di outgoing mail server si realizza configurando esplicitamente sui client di posta elettronica l'indirizzo del mail relay preposto ad inoltrare le proprie mail; la rete locale deve avere una macchina destinata a questa funzione, in quanto per motivi di sicurezza non e' permesso all'utenza locale di accedere direttamente ai relay esterni alla rete locale. Per non inserire nella configurazione dei client una dipendenza dal nome o dall'indirizzo di una macchina, si puo' inserire nella configurazione un nome generico che, tramite la registrazione opportuna di un record di tipo CNAME nel DNS, si possa tradurre nell'indirizzo di una macchina; all'occorrenza, il CNAME puo' essere modificato in modo trasparente all'utenza.

La scelta operata e' quella di utilizzare la stessa macchina per entrambe le funzioni, in quanto le due funzionalita' si distinguono solo per la configurazione dell'interfaccia (cioe' per come viene reso noto al mondo quale calcolatore sia destinato alla funzione) e non per le operazioni che il calcolatore svolge.

4.1 Pacchetto software: sendmail

Il software piu' diffuso al mondo che realizza questo servizio su piattaforma linux/unix e' sendmail [2)], prodotto open source che ha il vantaggio di essere estremamente flessibile nella sua configurabilita' e di essere dotato di un supporto che nel passato si e' sempre rivelato di qualita' eccellente e superiore a tutti i prodotti concorrenti; a dispetto della sua complessita' e' stato scelto questo applicativo per il fatto che garantisce a tutt'oggi la piena rispondenza agli standard definiti negli rfc pertinenti. La sua piena compatibilita' con i software aggiuntivi necessari alla completa configurazione del sistema (filtri anti-spam/anti-virus) non ha creato alcun tipo di problema.

4.2 Configurazione di sendmail

Il file che contiene la configurazione in linguaggio *m4* per il mail relay (vedi Appendice B) e' stato sviluppato a partire dal template messo a punto dal gruppo mail di CCR. I punti cardine della configurazione di Sendmail possono essere cosi' riassunti:

- Sicurezza: e' importante definire quali domini debbano essere considerati come locali, in modo da inoltrare verso l'esterno solo i messaggi inviati da tali domini. Una configurazione errata potrebbe rendere il proprio server veicolo di spam o virus e si correrebbe il rischio di essere inseriti in black list.
- Meccanismo di local delivery: il messaggio destinato ad un dominio riconosciuto come locale deve essere rediretto verso il servizio di mailbox. Viene utilizzato a questo scopo il file di configurazione userdb di sendmail. L'utilizzo di questo file permette di definire una convenzione per l'indirizzo degli utenti locali: l'indirizzo ufficiale di posta elettronica e' definito come ***<nome.cognome>@ge.infn.it***, ma viene garantito anche il recapito delle mail indirizzate a ***<username>@ge.infn.it***. Il file userdb istruisce sendmail a reindirizzare in entrambi i casi i messaggi verso l'indirizzo esplicito della mailbox dell'utente: ***<mailbox_name>@<mailbox_server>.ge.infn.it***.
- Filtri: nel file di configurazione viene definita la modalita' di connessione ai milter (message filter), cioe' programmi esterni che compiono operazioni di analisi sui messaggi in transito (filtri antispam, antivirus, content filtering, ecc..).

4.2.1 Autenticazione SMTP

Sendmail puo' utilizzare opzionalmente un framework di autenticazione ed autorizzazione chiamato SASL (Simple Authentication and Security Layer) o verificare i dati di un certificato x509. Questo permette di accettare ed inoltrare mail provenienti da un host connesso ad un dominio esterno e destinate ovunque, a patto che il client si sia in qualche modo autenticato dimostrando di essere un utente locale.

I meccanismi utilizzabili per SASL sono: GSSAPI, DIGEST-MD5, CRAM-MD5, LOGIN. Si e' deciso di autenticare tramite LOGIN (cioe' username e password) e via certificato. In particolare, per quanto riguarda LOGIN, le librerie SASL sono state configurate per utilizzare PAM (Pluggable Authentication Method) ed essere cosi compatibili con il sistema di autenticazione centralizzato (nel nostro caso NIS). La sessione di autenticazione avviene in modalita' cifrata tramite un meccanismo di handshaking TLS. Nel caso di autenticazione via x509 si e' deciso di accettare solo connessioni che presentassero un certificato emesso dall'authority dell'INFN, recante nel campo "Organization" la stringa "INFN". Questo per permettere potenzialmente a tutti gli associati INFN l'utilizzo del server SMTP.

4.3 Outgoing mail server

La scelta operata e' stata quella di configurare l'outgoing mail server di tutti i client di posta elettronica locali utilizzando un nome generico, smtp.ge.infn.it, a cui corrisponde nel DNS un record di tipo CNAME che punta al mail relay principale.

4.4 Dipendenze

Il servizio di mail relay, date le sue caratteristiche, dipende da alcune funzionalita' esterne di seguito elencate. Vedremo come siano state operate configurazioni che, mantenendo il requisito di dedicare alle funzioni di mail relay una macchina non pubblica, si sia garantito il funzionamento del servizio stesso senza introdurre dipendenze da altri calcolatori.

4.4.1 Dipendenza dal Domain Name System Service

Il mail relay utilizza il servizio DNS per identificare chi lo contatta, al fine di accertare l'eventuale appartenenza del mittente ai domini locali, e quindi di accettare di inoltrare i messaggi verso altri domini. In particolare, quindi, l'inoltro di messaggi inviati da calcolatori locali richiede che il mail relay possa accedere alle zone DNS locali.

Per evitare una dipendenza da altri calcolatori, si e' deciso di configurare il mail relay anche come DNS server delle zone locali. Questa soluzione pare opportuna in quanto il DNS service richiede risorse minimali, non entra in competizione con il servizio di posta e sarebbe un inutile spreco di risorse collocarlo su un calcolatore diverso e dedicato; inoltre il servizio DNS non ha la necessita' di operare su un public login server, quindi ben si adatta ad essere configurato sullo stesso calcolatore che svolge funzioni di mail relay.

4.4.2 Dipendenza dal servizio di autenticazione locale

La funzionalita' di mail relay con autenticazione richiede che il mail relay possa accedere al servizio di autenticazione locale; poiche' tale servizio e' attualmente fornito tramite NIS il mail relay deve necessariamente essere client del dominio NIS centrale di Sezione, per accedere allo user database del NIS server.

Benche' la mancanza di accesso ad un NIS server provochi solo un parziale malfunzionamento del servizio, si e' comunque ritenuto di risolvere anche questa interdipendenza, configurando il mail relay come NIS server sullo del dominio NIS locale.

Il servizio NIS, operato attraverso il demone ypbind, ha infatti le stesse caratteristiche del DNS service: ha un minimo impatto sulle risorse del calcolatore, ed e' idoneo alla sua utilizzazione su un calcolatore non pubblico.

4.5 Configurazione di ridondanza

Una attenzione particolare e' stata posta sulla ridondanza dei servizi: l'obiettivo e' quello di poter affrontare la perdita di funzionalita' di un calcolatore, per motivi di failure o per manutenzione, senza che l'utenza soffra disservizi.

E' stato quindi deciso di configurare due calcolatori in modo identico, cioe' che svolgano funzioni di mail relay, DNS server e authentication server.

Per i servizi DNS e NIS, il software dei client stesso si occupa di cercare da solo il server responsivo in caso di failure del primario.

4.5.1 Failover per la funzionalita' di mail relay

Per la ridondanza del servizio di mail relay e' necessario configurare un record di tipo MX che punti al secondo server; la scelta e' stata quella di non configurare un load balancing ed utilizzare un record a priorita' piu' bassa, in modo da operare, in condizioni di piena funzionalita', attraverso un solo server: le risorse di una singola macchina sono sufficienti, e se ne avvantaggiano le procedure di tracciamento dei problemi, dovendo ricercare le informazioni di log su un solo calcolatore.

Tuttavia la configurazione della ridondanza richiede un ulteriore accorgimento: le funzionalita' di recapito locale vengono svolte attraverso l'utilizzo del file di configurazione dello user database, che deve essere sincronizzato sui entrambi i mail relay; per fare cio' e' stato modificato il makefile di configurazione di sendmail ed e' stato creato uno script che copia i database ed il file sendmail.cf sul server secondario ogni volta che questi vengono creati o modificati.

In aggiunta, per tutelarsi da un malfunzionamento prolungato della rete geografica, si fa uso di un terzo mail relay collocato esternamente alla LAN (situato attualmente al CNAF) registrato nel DNS come MX record a priorita' inferiore.

4.5.2 Failover per la funzionalita' di outgoing mail server

Questo aspetto presenta difficolta' che non permettono, nella nostra configurazione, un failover automatico, in quanto gli applicativi client non permettono di configurare un outgoing mail server di failover.

Una possibilita' e' quella di definire nel DNS due record corrispondenti al nome generico smtp.ge.infn.it, e che puntino a due indirizzi differenti. Tuttavia questa soluzione comporta, in caso di failure di uno dei due server, il fatto che la meta' delle volte i client riceveranno come risoluzione per il nome generico l'indirizzo corrispondente al calcolatore inaccessibile, con tempi di risposta e messaggi di errore che si desiderano evitare.

Si e' quindi ritenuto di operare manualmente la modifica del record CNAME del DNS corrispondente al nome generico in occasione di failure. Affinche' questo sia possibile, e' necessario che il mail relay principale ed il DNS server primario risiedano su macchine diverse: in questo modo, in occasione di una failure del mail server primario, il DNS server primario sara' accessibile e sara' possibile modificare il record DNS.

5 SERVIZIO DI MAILBOX

L'accesso ai messaggi di posta elettronica e' garantito dal servizio di mailbox. Un'implementazione di questo sistema deve essere in grado di:

- ricevere, contenere e rendere disponibili tutti i messaggi entranti destinati ai fruitori del servizio. Deve soddifare quindi requisiti minimi di storage, filesystem, connessione di rete, cpu ed efficenza del software;

- rendere semplice, sicuro e continuativo l'accesso agli utenti e permettere loro di compiere autonomamente le varie personalizzazioni possibili (risposte automatiche, filtro e smistamento dei messaggi in arrivo);
- permettere il piu' semplicemente possibile all'amministratore la gestione delle caselle di posta (creazione, eliminazione, permessi e quote) ed il backup/restore dei dati contenuti;
- essere compatibile con gli standard vigenti [1] (rfc3501, rfc2033)

Nella configurazione preesistente questo servizio era svolto dal calcolatore di public login, e l'archiviazione delle mail era gestita dal software imap sviluppato dalla Washington University, con i file contenenti le mail direttamente accessibili agli utenti. Cio' ha comportato in diverse occasioni la corruzione dei file delle mail conseguente ad accessi diretti da parte degli utenti. Si e' quindi scelto di demandare a questa funzione un calcolatore dedicato, senza accesso pubblico, che rendesse disponibili le mail esclusivamente tramite un protocollo di rete. La scelta di utilizzare un calcolatore dedicato richiede naturalmente che questo server svolga anche funzioni di mail relay con la sola funzione di delivery locale, per permettere al mail relay principale di consegnare i messaggi verso la destinazione finale.

Rispetto ai requisiti posti per il servizio di mail relay (vedi cap. 2), vanno aggiunte considerazioni sullo storage necessario alla gestione delle mail degli utenti, che deve essere ospitato su dischi affidabili ed in configurazione ridondante.

5.1 Scelta dell'hardware e del software di sistema

Le considerazioni fatte nel cap. 3 sulla scelta di CPU, disco, memoria RAM, scheda di rete e sistema operativo sono valide anche in questo caso. Per lo storage si e' deciso di utilizzare un sistema di dischi SCSI in configurazione ridondata RAID5 per assicurare la continuita' del servizio e non incorrere in perdita di dati in caso di rottura di un disco.
Al momento del progetto lo spazio occupato dalle mailbox degli utenti era di circa 20 GB. L'hardware scelto e' capace di ospitare 420 GB netti di area dati, che e' stata ritenuta idonea a sopportare una crescita di esigenza di storage per i quattro anni successivi, senza la necessita' di configurare quote.

L'utilizzo della soluzione software in seguito descritta (Cyrus IMAP server) richiede una scelta opportuna del filesystem: questo dovra' gestire numerosissimi file di piccole dimensioni (attualmente le 300 mailbox locali contengono circa 2.900.000 file). Reiserfs e' sembrato la scelta piu' opportuna in quanto appositamente progettato per gestire in modo ottimale questa tipologia di storage [3].

5.2 Protocollo di accesso al servizio: IMAPS

Se si desidera fare in modo che la maggioranza dei client di posta elettronica sia in grado di connettersi al mailbox server, la scelta del protocollo di accesso alle caselle di

posta e' ristretta a due candidati: IMAP e POP3. Il primo presenta alcuni vantaggi che ci hanno fatto propendere per esso. IMAP consente infatti di mantenere tutta la struttura di una mailbox (mail e subfolder) all'interno di un server centrale, cio' significa che:

- l'utente puo' accedere alla sua casella di posta elettronica da macchine diverse ritrovando intatta la struttura della stessa;
- l'utente non si deve preoccupare della gestione dei backup dei suoi messaggi in quanto questo puo' essere centralizzato;
- il controllo dei messaggi in arrivo avviene scaricando solo gli header, evitando cosi' utilizzo di banda inutile (dovuta ad esempio grossi allegati).

In realta' il protocollo realmente utilizzato per l'accesso alle caselle di posta e' IMAPS (IMAP over SSL), una variante di IMAP che utilizza SSL per cifrare il traffico rendendo le connessioni piu' sicure.

5.3 Scelta del pacchetto software: cyrus

Cyrus e' l'IMAP server scritto e mantenuto dalla Carnegie Mellon University. A differenza di altri software dello stesso genere, Cyrus e' pensato per gestire i messaggi di posta elettronica su un file system non accessibile direttamente dai destinatari.

5.3.1 Componenti del software

Il processo principale e' il *master*. Questo puo' essere configurato per ascoltare sulle porte TCP relative ai seguenti protocolli: *imap, imaps, pop3, pop3s, lmtp, sieve*.
Le connessioni vengono poi gestite dai componenti secondari di Cyrus, attivati dal master: *imapd* per imap e imaps, *pop3d* per pop3 e pop3s, *lmtpd* per lmtp (protocollo di archiviazione delle mail nelle mailbox) e *timsieved* per sieve (protocollo di configurazione di funzionalita' quali archiviazione automatizzata, messaggi di vacation, reiezione, etc.).

5.3.2 Gestione dei messaggi

I messaggi identificati dal MTA come "locali" vengono trasmessi a Cyrus che li gestisce attraverso il processo lmtpd (local mail transport protocol daemon). La funzione di lmtpd e' quella di applicare al messaggio le regole di analisi e smistamento personalizzato (configurabile attraverso il *sieve scripting language*, rfc3028 [1)]) e di conseguenza salvarlo nella mailbox del destinatario. Ogni messaggio recapitato sara' registrato da lmtpd sul database degli indici (vedi §5.3.3)

5.3.3 Struttura delle mailbox

Cyrus e' capace di gestire folder imap che contengono contemporaneamente mail e subfolder; su disco Cyrus utilizza una struttura in cui ogni folder IMAP corrisponde ad una directory, i subfolder sono subdirectory ed ogni mail contenuta in un folder imap viene archiviata come singolo file all'interno della directory corrispondente. La struttura della

mailbox di un utente e’ quindi un sottoalbero la cui radice rappresenta il folder della INBOX: tutti i folder imap creati dall’utente appaiono come subfolder della INBOX.

Le mailbox sono contenute in una o piu' directory principali, che vengono chiamate *partition*: ogni directory nella parition contiene la maibox di un utente. Diverse partition possono risiedere su dischi fisici differenti: questo rende possibile espandere dinamicamente l’area dedicata allo storage delle mail aggiungendo nuove partition senza richiedere un ridimensionamento delle partizioni fisiche su disco.

I dati relativi a tutta la struttura delle mailbox sono racchiusi nel database “mailbox.db” che si trova nella directory definita nel file di configurazione come *configdirectory*. In aggiunta a cio’, ogni folder contiene un file di indice che garantisce un accesso rapido alle mail.

5.3.4 Autenticazione e autorizzazione

Cyrus utilizza un framework di autenticazione chiamato SASL (Simple Authentication and Security Layer) attraverso il quale e' possibile definire i seguenti meccanismi di autenticazione/autorizzazione: PLAIN, DIGEST-MD5, CRAM-MD5, KERBEROS_V4, GSSAPI.

Nel caso dell'installazione in oggetto si e' deciso di utilizzare PLAIN (cioe' username e password) e di verificare le credenziali via PAM (Pluggable Authentication Method). In questo modo l'installazione puo' essere resa compatibile con un numero molto ampio di sistemi di autenticazione, sia distribuiti che locali. Poiche' Cyrus non e' configurabile per interagire direttamente con le PAM, nel file imapd.conf (vedi Appendice B) e' necessario definire un demone intermedio di autenticazione, *saslauthd*, che dovra' essere configurato a sua volta per utilizzare PAM (cioe' essere eseguito con il flag “-a pam”).

5.3.5 Motivi della scelta di cyrus

La struttura delle mailbox e la corrispondenza biunivoca tra messaggi e file offre numerosi vantaggi: l'efficienza nell'utilizzo della memoria e’ maggiore rispetto ai sistemi che associano un file a ogni mailbox, la dimensione dei backup incrementali giornalieri e’ piu' contenuta, la struttura dei subfolder e’ a piu' livelli.

Cyrus dispone di un meccanismo di gestione delle quote e di ACL sui folder: benche’ si sia deciso di non configurare queste funzionalita’ nella fase iniziale, si ritiene importante poterne disporre per il futuro.

Un altro vantaggio dell'installazione di Cyrus risiede nel fatto che l'utente puo' configurare a piacere lo smistamento dei messaggi e le risposte automatiche (vedi §5.5). Entrambe le operazioni sono eseguite senza dover fare login interattivo sul mailbox server.

Cyrus puo' essere installato seguendo uno schema chiamato Murder o Imap Aggregator [4)] che garantisce un'ottima scalabilita' in funzione del carico degli accessi. L'aspetto centrale del Murder e' la distribuzione delle mailbox su diversi server di backend e la separazione delle funzioni di local delivery e di gestione degli indici. Le prestazioni del sistema possono cosi' essere accresciute aumentando il numero delle macchine che

compongono il sistema. Esistono in produzione configurazioni di notevoli dimensioni, quale quella della Carnegie Mellon University, che gestisce piu' di 26000 mailbox [5)].

Infine il sistema di autenticazione dell'utente si puo' integrare semplicemente con l'infrastruttura locale.

5.4 Configurazione di cyrus

I file di configurazione da prendere in considerazione sono *cyrus.conf* e *imapd.conf* (vedi Appendice B).

5.4.1 Il file cyrus.conf

Il file cyrus.conf viene letto dal processo master e contiene tre sezioni: START, SERVICES e EVENTS. La sezione START contiene i comandi che il processo master esegue alla partenza; in particolare deve essere sempre presente il controllo dei database che contengono la struttura della mailbox (vedi Appendice B).

Il cuore della configurazione del master e' la sezione SERVICES. Essa contiene la tipologia di servizi che il server dovra' gestire ed il modo in cui dovranno essere lanciati i demoni associati. Almeno un servizio lmtp deve essere presente per consentire il local delivery, sia esso attuato attraverso socket tcp o socket unix.

Infine la sezione EVENTS contiene la lista dei processi che devono essere eseguiti dal master ad intervalli di regolari.

5.4.2 Il file imapd.conf

Il file *imapd.conf* contiene la configurazione di tutti i servizi Cyrus disponibili. (vedi Appendice B). In particolare vanno inseriti in *imapd.conf* tutti i percorsi relativi ai database degli indici ed alle directory delle mailbox.

Atraverso il file *imapd.conf* e' possibile configurare varie funzioni di ogni altro servizio (sieve, pop3, pop3s, lmtpd).

5.5 Servizi accessori all'inbox service: local delivery, reiezione, vacation

Come descritto in precedeza il processo lmtpd si occupa di consegnare i messaggi nelle mailbox degli utenti, avendo cura di eseguire le operazioni di smistamento automatico personalizzato. Il processo lmtpd e' capace di interpretare un linguaggio di scripting chiamato ***sieve*** che definisce alcune azioni di base da eseguire sui messaggi in relazione a caratterisctiche dei messaggi stessi. E' possibile spostare un messaggio in una folder definita, rigettarlo, modificarne uno o piu' header, inoltrarlo ad un indirizzo e-mail. Esiste anche un funzione di "vacation" cioe' una risposta automatica che avverte il mittente di una eventuale assenza prolungata del destinatario. Ogni utente ha la possibilita' di definire il proprio sieve script (vedi Appendice A) collegandosi attraverso un client remoto a linea di comando (sieveshell), ma in modo piu' generale via http tramite una web application (Websieve) che offre un'interfaccia piu' semplice ed immediata.

Il componente di Cyrus che risponde alle connessioni dei sieve client e' il processo timsieved.

5.6 Dipendenze da altri servizi

Il funzionamento del servizio di mailbox ha due dipendenze fondamentali: il DNS e il servizio di autenticazione (NIS nella configurazione locale). Per rendere il server il piu' possibile indipendente dalla raggiungibilita' di altre macchine si e' provveduto a installare su di esso un NIS slave server, configurando ypbind in modo da utilizzare “localhost” come NIS server. Questo server viene utilizzato esclusivamente dal client locale e non fornisce servizio ai NIS client del dominio di sezione.

5.7 Migrazione

Dopo avere installato e configurato il nuovo server IMAP e' stato necessario trovare un modo per migrare la posta elettronica immagazzinata sul sistema precedente. Nel caso locale la migrazione e’ stata operata a partire dal sistema preesistente, basato su WU IMAP. Vengono di seguito descritte la tecnica e gli accorgimento adottati.

Lo scenario antecedente alla migrazione era il seguente:

- messaggi: 350.000
- storage: 20 GB
- mailbox: 250

5.7.1 Gli strumenti

Per trasferire i messaggi e' stato utilizzato un tool scritto in perl chiamato Imapmigrate [6)]. Il tool e’ necessario per automatizzare la migrazione della struttura di subfolder, in quanto il protocollo IMAP non permette di trasferire foder interi, ma solo (gruppi di) messaggi; inoltre la struttura di destinazione richiede che i folder siano tutti subfolder di INBOX, a differenza della struttura di partenza in cui i folder sono separati dalla INBOX.

Per installare Imapmigrate e’ necessario l’interprete perl ed il suo modulo Mail-IMAPClient.

Imapmigrate deve essere eseguito sul server Cyrus IMAP. In pratica utilizza il modulo IMAPClient di perl per connettersi ad entrambi gli IMAP server e trasferire i messaggi dall'uno all'altro. L'utilizzo e' piuttosto semplice:

- si deve creare un file contenente gli username e le password IMAP in chiaro degli utenti da migrare, ogni coppia username/password su una linea e divisi da uno spazio
- si esegue il comando imapmigrate userlist dove "userlist" e' il nome del file suddetto.

5.7.2 Modalita' di attuazione

Come visto e’ necessario conoscere le password degli utenti. Nel nostro caso, l'autenticazione avviene attraverso mappa NIS dei file passwd/shadow; e' possibile salvare il file shadow esportato dal NIS server e modificarlo, assegnando a tutti gli utenti una password predefinita. Questo richiede naturalmente uno shutdown del servizio di autenticazione centrale, e di tutti i servizi connessi.

Il passo successivo e' la disabilitazione dell'applicazione deputata ad eseguire il local delivery (sendmail nel caso in oggetto) in modo che non possano giungere messaggi durante la migrazione. E’ possibile configurare un SMTP server secondario allo scopo di tenere i mail in arrivo per il tempo necessario alla migrazione; noi abbiamo utilizzato il servizio fornito dal CNAF (infngw).

Per quanto riguarda il server Cyrus IMAP, prima di iniziare la migrazione, e' necessario creare tutte le mailbox degli utenti utilizzando l'apposita utility cyradm (o in alternativa scrivendo uno script in Perl utilizzando il modulo Cyrus::IMAP::Admin, come fatto localmente). A questo punto si puo' procedere all'esecuzione di Imapmigrate.

5.7.3 Tempo impiegato

Una mailbox contenente un numero di messaggi pari a 500 MB e' stata processata in un tempo di circa 15 min. In totale e' stata impiegata un'intera giornata per migrare tutta la struttura. In una ventina di casi si sono verificati errori di migrazione conseguenti a file sorgenti corrotti, che hanno richiesto un intervento manuale. I messaggi migrati saranno contrassegnati dal server Cyrus come nuovi (unseen) e la data di creazione del file sara’ quella della migrazione; a causa di cio’ alcuni client di posta elettronica (ad esempio Microsoft Outlook) mostreranno ogni messaggio come appena ricevuto. Un workaround a questo problema si ottiene modificando opportunamente il modulo Mail::IMAP::Client [7)]

6 FILTRI ANTISPAM/ANTIVIRUS

Il riconoscimento automatico di messaggi indesiderati e' un aspetto di primaria importanza nella gestione di un sistema di posta elettronica. L'assenza di un sistema simile puo' rendere praticamente inutilizzabile una casella e-mail.

La funzione di un filtro antispam e' quella di analizzare i messaggi entranti e, sulla base di alcune regole predefinite, assegnare ad ognuno di essi un punteggio. Se il punteggio dovesse superare una soglia (configurabile) il messaggio sarebbe interpretato come “spam” e verrebbe trattato come tale.

I filtri antivirus confrontano il contenuto dei messaggi con alcuni pattern (che vengono aggiornati automaticamente) e, se i dati corrispondono, riconoscono l'e-mail come infetto.

In entrambi i casi i programmi che eseguono i filtri permettono di definire il modo di trattare i messaggi interpretati come SPAM (eventuali inserimento di una tag, modifica di subject o body, rimozione della mail) o virus (eventuale rimozione dell’attachment ed archiviazione in zona di quarantena).

Oltre all'efficienza (cioe' al numero di spam riconosciuti sul totale di quelli arrivati) e al numero di falsi positivi, che deve essere contenuto al minimo possibile, abbiamo ritenuto di fondamentale importanza offrire all'utente la possibilita' di configurare filtri personalizzati, in modo da adattarli alle proprie esigenze. Come vedremo e' stata adottata una configurazione con doppio filtro in cascata, il primo (Sophos PureMessage) operante a livello di mail relay, il secondo (spamassassin) operante a livello di mailbox server.

6.1 Pacchetti software (1): Spamassassin

Spamassassin [8)] è un sistema antispam opensource scritto in Perl. Il pacchetto comprende, oltre agli script veri e propri, una serie di file contenenti la descrizione dei vari test ed i punteggi associati ad ognuno di essi.

6.1.1 Struttura di Spamassassin

Spamassassin puo' essere utilizzato in modalita' client-server o standalone. La prima è piu' efficiente e quindi indicata per sistemi di gestione dello spam centralizzati (come nel caso in esame), mentre la seconda è piu' semplice e adatta ad un uso personale.

Il demone *spamd* riceve il messaggio completo di header dal client *spamc* che a sua volta lo ha ricevuto da Sendmail attraverso un processo chiamato *spamass-milter*, che fa uso delle librerie *milter* (mail-filter) di Sendmail (vedi §4.2). Quando *spamd* ha terminato l'analisi del messaggio e le eventuali modifiche agli header, lo ripassa al client *spamc* e quindi a Sendmail.

6.1.2 Configurazione di Spamassassin

Spamassassin e' stato configurato (vedi appendice B) per riconoscere un messaggio di spam se il punteggio supera il valore di soglia di 3 punti. Tale soglia costituisce un valore di default personalizzabile, ed e' stata decisa dopo un periodo di prova di circa due mesi del sistema in produzione. Nel caso il punteggio superi il valore di soglia, il subject del messaggio viene modificato aggiungendo una stringa opportuna, in modo che ogni utente possa decidere come agire configurando opportunamente il sistema di smistamento (vedi §5.5).

Al fine di migliorare l'efficienza del sistema si e' fatto uso di alcuni test che presuppongono il collegamento con server remoti:

- RBL/DNSBL [9)]: Spamassassin controlla che i relay attraverso i quali sono giunti i messaggi non siano elencati in speciali server DNS (dnsbl) che raccolgono indirizzi noti per essere veicolo di spam. A seconda dell'attendibilita' del server rbl e' possibile assegnare un punteggio al test.
- Razor2 [10)]: e' un programma che Spamassassin utilizza per collegarsi ad un sistema di riconoscimento dei messaggi di spam che si basa sul contributo volontario degli utenti.
- Pyzor [11)]: simile a razor2.

- DCC [12]: e' un sistema centralizzato che basa il riconoscimento degli spam sul numero di destinatari che hanno ricevuto lo stesso messaggio.

E' importante inoltre cercare di discriminare le fonti “fidate” in modo da diminuire il rischio di falsi positivi. A tal proposito è stato utile configurare un elenco di indirizzi affidabili e conferire punteggi negativi a messaggi provenienti da utenti di sezione autenticati o dal webmail di sezione (vedi §7.1). Infine viene assegnato un punteggio alto ai messaggi che sono stati gia' riconosciuti come spam dal precedente filtro (Sophos Puremessage) attraverso un header inserito da quest'ultimo (Vedi §6.3)

6.1.3 Vantaggi e svantaggi

Vantaggi:

- E' possibile in modo semplice e sicuro permettere agli utenti una gestione personalizzata del filtro (indirizzi in whitelist, subject tag, soglia)
- Elasticita' di configurazione.

Svantaggi:

- La configurazione personalizzata funziona correttamente solo se ogni messaggio contiene un solo destinatario. E' necessario percio' suddividere ogni messaggio con destinatari multipli in piu' messaggi con un singolo destinatario, con conseguente aumento di carico del sistema.
- Appena installato, Spamassassin ha un'efficienza piuttosto bassa e richiede un lungo lavoro di calibrazione della configurazione per aumentare la qualita' del filtro.
- I controlli con server remoti possono far aumentare il tempo di analisi del singolo messaggio nel caso ci fossero problemi di rete.
- In alcuni casi l'utilizzo di RBL provoca falsi positivi dovuti alla presenza in black list di indirizzi dinamici assegnati da provider in dial-up o ADSL.

6.2 Pacchetti software (2): Sophos Puremessage

Sophos Puremessage [13] e' un prodotto antispam e antivirus commerciale scritto in Perl. E' configurabile attraverso un'interfaccia web ed interagisce con Sendmail attraverso un milter proprietario. Per funzionare correttamente ha bisogno di un database Postresql che contiene la statistica e lo storico del sistema e viene installato automaticamente. I test dell'antispam ed i pattern dell'antivirus vengono aggiornati giornalmente dal sito del produttore. Le policy di analisi e gestione dei messaggi sono scritte in linguaggio Sieve e sono anch'esse configurabili via web.

6.2.1 Configurazione di Puremessage

Il sistema è configurato per agire in modo differente a seconda che un messaggio provenga da una macchina interna o esterna alla rete di sezione. Nel primo caso vengono

analizzati solo gli eventuali virus presenti nel messaggio e, in caso di esito positivo, il virus viene estratto ed il subject modificato. Nel secondo caso Puremessage attiva sia i test antispam che antivirus. Nell'eventualita' il messaggio sia positivo all'antivirus, il filtro agisce come in precedenza, se fosse positivo all'antispam aggiunge un header in modo che venga riconoscuto dal filtro successivo (Spamassassin, vedi §6.3). La configurazione asimmetrica permette di filtrare virus trasmessi verso l'esterno ma di evitare di identificare messaggi uscenti come SPAM, demandando questa funzione al sito di destinazione.

6.2.2 Vantaggi e svantaggi

Vantaggi:

- Comoda e semplice interfaccia di configurazione
- Configurazione server group in ridondanza (vedi §6.3)
- Statistiche web di transito messaggi per categoria
- Aggiornamento automatico del sistema
- Buon rapporto efficienza/tempo di manutenzione

Svantaggi:

- Insufficiente livello di configurazione personalizzata per gli utenti; inoltre il sistema archivia spam e virus in una area di quarantena e richiede che l'utente acceda a quest'area per controllare eventuali falsi positivi con una azione indipendente ed estranea alla lettura delle mail.
- Non si integra con i sistemi di autenticazione locali: l'eventuale accesso alla quarantena ed alla configurazione personalizzata richiede la creazione di un accesso specifico per ogni utente, con password dedicata.

6.3 Configurazione del sistema

Le caratteristiche dei due software antispam sono da molti punti di vista complementari; si e' pensato percio' di integrare in due sistemi in modo da giovare dei vantaggi di entrambi. Puremessage e' stato installato sui due mail relay (vedi §4.5.1) in configurazione di server group. In pratica su uno dei due server (quello il cui indirizzo corrisponde al record MX a piu' alta priorita') è presente l'installazione di Puremessage che ha funzione di master, cioe' possiede fisicamente il database Postgresql. Ogni modifica di configurazione attuata sul master viene automaticamente trasferita sullo slave anche se entrambi i server possono funzionare autonomamente.

La policy da attuare in caso di riconoscimento di un messaggio spam si limita ad aggiungere un opportuno header al messaggio stesso, in modo che esso possa essere trattato successivamente.

Spamassassin e' installato sull'inbox server e configurato per attuare i test di rete (vedi §6.1.2) e per assegnare un punteggio elevato ai messaggi gia' riconosciuti da Puremessage. Il ruolo giocato nel sistema da Spamassassin e' duplice: aumenta l'efficienza antispam (vedi §6.4) e permette agli utenti di configurare a piacere il proprio filtro senza dover interagire direttamente con Sophos.

L'unico inconveniente risiede nel fatto che, per questioni di sicurezza, non e' permesso il login degli utenti sul server delle mailbox, dove si trovano i file di configurazione personali di Spamassassin. A tal fine e' stata sviluppata localmente un'applicazione web (Webspam) che risiede sul server e rende possibile ad ogni utente (previa autenticazione criptata via https) la modifica del comportamento del proprio filtro. Webspam e' scritto in Perl [14)].

6.4 Prestazioni

Al fine di verificare le prestazioni del sistema antispam si e' deciso di duplicare le e-mail di un limitato numero di utenti appartenenti a diversi gruppi sperimentali o servizi, in modo da avere un campione che fosse il piu' possibile significativo. Tali utenti hanno successivamente controllato i messaggi sulle mailbox duplicate, avendo cura smistarli in ham (mail buoni), spam, falsi positivi e falsi negativi. I risultati sono riportati in tabella 6.1

TAB. 6.1: statistiche di efficienza filtri antispam

Messaggi		Numero di messaggi	Percentuale
Totale		12321	
Ham		3444	27.95%
Spam		8877	72.05 %
Falsi positive		47	1.36 %
Spam identificati	Sophos	8428	94.94 %
	Spamassassin	6917	77.92 %
	Doppio filtro	8601	96.89 %

L'efficienza, calcolata come il rapporto tra numero di spam riconosciuti correttamente e numero di spam ricevuti, e' circa il 96.9%. Da notare come ognuno dei due sistemi preso singolarmente avrebbe avuto una efficienza minore.

Va considerato che in questo test l’identificazione di falsi positivi non ha potuto godere dell’incremento di prestazioni legata alla personalizzazione del filtro tramite white list, in quanto le utenze di test erano configurate utilizzando i parametri di default, senza ottimizzazioni specifiche. Il test si e’ protratto per circa tre settimane, ed e’ finalizzato solo a verificare l’entita’ dell’aumento di prestazioni legate all’utilizzo del doppio filtro.

7 SERVIZI ACCESSORI

Il servizio di posta elettronica e’ stato integrato con due servizi accessori che ne completano la funzionalita’.

7.1 Webmail

Poiche' può capitare che un utente abbia necessita' di accedere alle proprie mail da un calcolatore sprovvisto di un client IMAP, si e' configurato un servizio di accesso basato su protocollo http. L'applicativo utilizzato e' Horde [15)], un software di pubblico dominio che fornisce una interfaccia al server IMAP largamente configurabile dall'utente, il cui sistema di autenticazione e' integrabile con il sistema di autenticazione centrale tramite PAM. La configurazione del software e' praticamente quella di default, fatto salvo che, per motivi di sicurezza, si impone l'utilizzo di protocollo criptato https.

Horde permette anche di inviare mail senza dover configurare sul client un mail relay opportuno, in quanto l'autenticazione conseguente all'accesso rende disponibile in modo trasparente l'utilizzo del mail relay di sezione.

Horde e'stato installato sull'inbox server.

7.1 List server

Gran parte dell'utenza si trova spesso nella necessita' di inviare mail ad uno stesso gruppo di persone. Benche' la maggior parte dei client permetta di configurare internamente liste di destinatari, risulta molto utile poter disporre di indirizzi definiti per gruppi di destinatari organizzati in un unico database ed accessibili a piu' utenti.

A questo scopo e' stato installato un software di pubblico dominio, Mailman [16)], che gestisce liste di utenti assegnando a ciascuna lista un indirizzo locale di posta.

Le caratteristiche di Mailman sono una ampia configurabilita' delle liste, che possono essere aperte (tutti possono inviare mail), chiuse (solo gli iscritti possono inviare mail), o moderate (la decisione di inoltrare una mail alla lista e' presa da un moderatore che viene avvisato via mail ed opera manualmente per l'inoltro), ed hanno diversi parametri di configurazione che ciascun utente può modificare. E'anche possibile disporre di archivi per salvare tutte le mail inviate e rendere tali archivi disponibili agli utenti. Inoltre ciascuna lista può essere gestita da uno o piu' utenti che non devono necessariamente essere gli amministratori del sistema, rendendo cosi' distribuito il carico di lavoro di management. Gli amministratori debbono intervenire obbligatoriamente solo per creare una lista, in quanto la creazione non implica solo la definizione della lista nel database gestito da Mailman, ma anche l'inserimento di opportuni record nel file userdb gestito dal sendmail dei mail relay.

L'unico aspetto negativo di questo pacchetto software risiede nel fatto che il meccanismo di autenticazione necessario alla gestione delle liste non si integra con i sistemi di autenticazione centrali, e richiede l'utilizzo di username/password dedicate. Per questo motivo la soluzione adottata e' da ritenersi temporanea in attesa di un pacchetto software che superi questa limitazione.

Anche il list server di Mailman e'installato e configurato sul mailbox server centrale.

8 ALLARMISTICA

Le potenziali cause di malfunzionamento del sistema sono:

- problemi hardware che non pregiudicano il funzionamento del sistema (rottura di un relay server o di una parte dello stesso, rottura di un disco del mailbox server o di un elemento hardware ridondato);
- problemi hardware che pregiudicano il funzionamento del sistema (rottura di entrambi i relay server, rottura del mailbox server);
- problemi software che non pregiudicano il funzionamento del sistema (interruzione di sendmail su un relay server, interruzione dei filtri antispam, crash del sistema operativo su un relay server);
- problemi software che pregiudicano il funzionamento del sistema (interruzione di sendmail su entrambi i relay server, interruzione di sendmail sul mailbox server, interruzione di cyrus sul mailbox server).

Sono stati adottati sistemi di allarmistica differenti a seconda del tipo di problema da diagnosticare. Per verificare la raggiungibilita' via rete dei calcolatori e lo stato dei vari servizi (IMAP, SMTP, relay) e' stato installato un server di monitoring su cui si e' configurato NAGIOS, un sistema centralizzato che contatta periodicamente i server e le applicazioni desiderati e ne mostra lo stato attraverso una pagina web. Ogni qualvolta occorra un problema, NAGIOS invia un messaggio di posta elettronica che avverte gli amministratori del malfunzionamento. Lo stato dei dischi e degli altri elementi hardware e' tenuto sotto controllo da applicazioni che risiedono sui server e sfruttano le informazioni che i driver delle periferiche mettono a disposizione. Anch'essi inviano un e-mail quando rilevano un guasto.

Questo meccanismo di allarmistica ha lo scopo di allertare in tempo reale il personale in caso di eventi che pregiudichino la ridondanza hardware o software del sistema. Se il funzionamento del sistema di posta elettronica fosse completamente pregiudicato, nessun messaggio potrebbe giungere agli amministratori e solo l'allarmistica via web sarebbe disponibile, con conseguente impossibilita' a un pronto intervento qualora il guasto si verificasse in assenza del personale del calcolo. Per questo motivo in futuro si pensa di introdurre un meccanismo di notifica dei malfunzionamenti via SMS.

9 STATISTICHE

In fig. 9.1 e' mostrato l'aumento del numero di mail in conseguenza dello splitting dei messaggi indirizzati a piu' destinatari, necessario per l'applicazione delle configurazioni personalizzate dei filtri antispam (vedi §6.1.3). L'aumento del carico sul mailbox server e' compreso tra il 30% ed il 50%, a seconda dei periodi. Negli ultimi mesi dell'anno l'incidenza di questo fattore si e' ridotta, principalmente in conseguenza del fatto che sono state chiuse le mailing list locali a mittenti non iscritti, cosa che ha ridotto notevolmente lo spam indirizzato a destinatari molteplici.

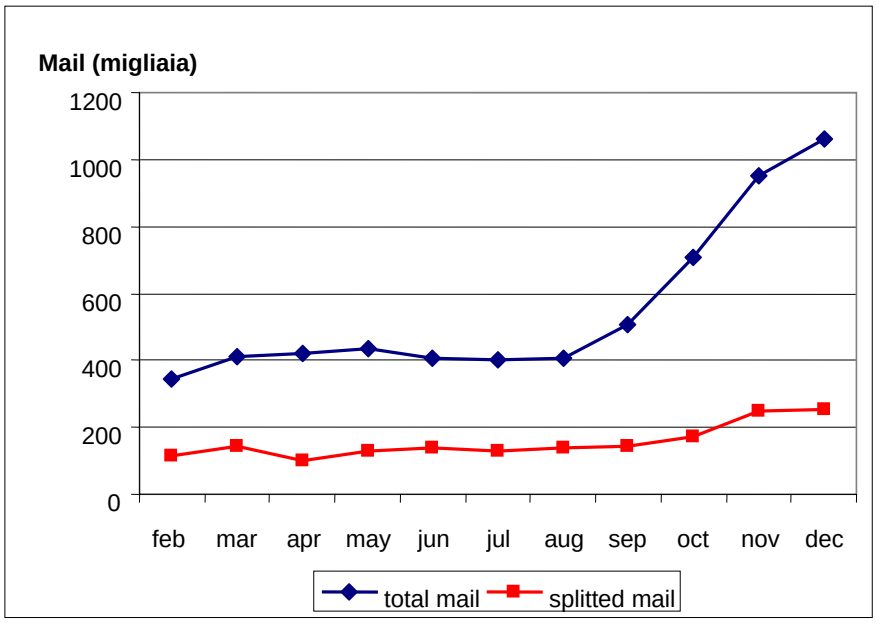

Fig. 9.1: Incidenza dello splitting sul numero di messaggi totali

L'aumento del numero delle mail non incide sulla capacita' del sistema nel trattare i messaggi in arrivo: la valutazione dei periodi in cui i mail relay rifiutano nuove connessioni per eccesso di carico mostra come in un anno l'aggregato di questi tempi sia contenuto entro le sette ore complessive, suddivise in sporadici periodi di durata inferiore al minuto.

La figura 9.2 rappresenta il grafico del numero di messaggi giornalieri che i due relay passano al mailbox server, ed il numero di tali messaggi identificati come SPAM da Sophos. Si puo' vedere come la percentuale di SPAM sia dell'ordine dell'80% del totale.

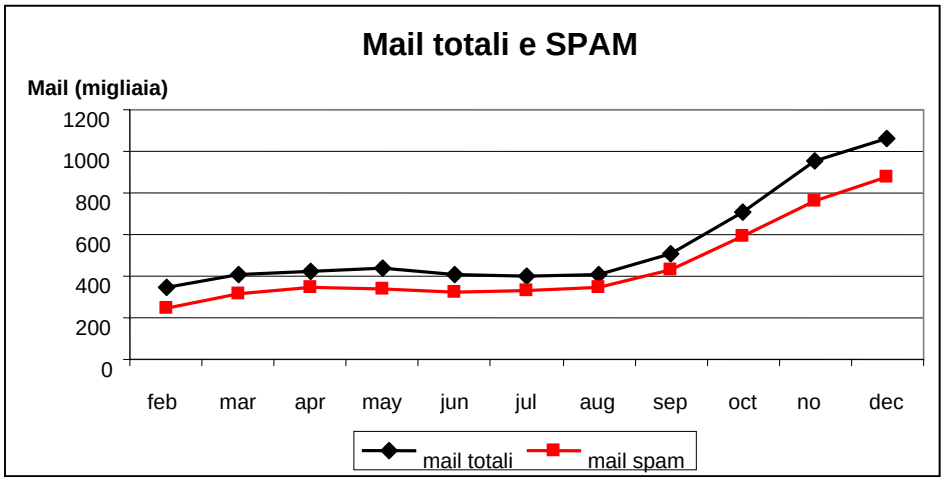

Fig 9.2: Mail totali e SPAM ricevuti ed inoltrati dai mail relay

Si nota anche come nel corso degli ultimi mesi si sia verificato un notevole incremento dello SPAM, che ha quasi triplicato il traffico di mail complessivo gestito dal sistema di posta, portandolo ad un valore di circa 35000 messaggi al giorno.

E' interessante osservare in fig. 9.3 i traffici separati per mail relay: si nota come il mail relay secondario tratti circa il 30% dei messaggi complessivi, contrariamente a quanto ci si sarebbe aspettato (nessun carico in conseguenza della configurazione a priorita' inferiore dell'MX record relativo). Si nota anche come il traffico gestito dal relay secondario sia per piu' del 90% costituito da SPAM identificato, mentre la percentuale scende sotto l'80% per il relay primario. Una possibile spiegazione e' che alcuni spammers possano utilizzare come relay destinatario il server associato all'MX record a priorita' inferiore.

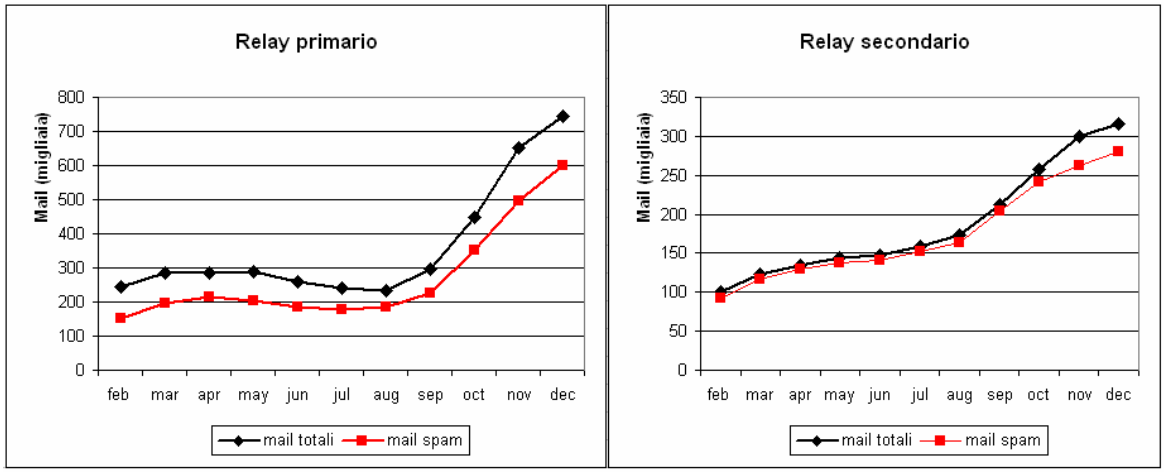

Fig 9.3: mail totali e SPAM divise per relay server

A sostegno di questa tesi si puo' anche citare il fatto che la classifica dei relay esterni che maggiormente inviano spam ai server locali e' guidata con amplissimo margine dal relay esterno alla LAN per il dominio locale (vedi §4.5.1).

Le statistiche riportate sopra sono state valutate utilizzando i log file di sistema, attraverso appositi applicativi sviluppati localmente; Sophos PureMessage dispone di un sistema di statistica integrato abbastanza ricco, ma e' stato configurato solo recentemente e non dispone ancora di statistica sufficiente.

Il filtro antivirus ha manifestato un comportamento assolutamente soddisfacente, al punto che, unitamente alla opera di educazione dell'utenza nella gestione degli allegati di dubbia provenienza, non si sia manifestata alcuna infezione virale conseguente a messaggi di posta elettronica negli ultimi 12 mesi.

10 CONSIDERAZIONI CONCLUSIVE

La ristrutturazione del servizio di posta elettronica ha richiesto due mesi di progettazione, spesi nella raccolta dei requisiti da soddisfare in termini di risorse hardware,

tipo e qualita' dei servizi da implementare, valutazione delle possibili soluzioni e delle strategie di migrazione. La parte realizzativa, suddivisa in diverse fasi (migrazione del sistema di mail relay, migrazione del mailbox server, installazione e configurazione iniziale dei filtri, installazione e configurazione iniziale dei servizi accessori) e' stata operata in due mesi, seguiti da un periodo di sei mesi dedicato alla messa a punto delle configurazioni ed allo sviluppo di applicativi locali (Webspam, modifiche a Mailman, pachetti di analisi statistica) per ottenere esattamente le funzionalita' richieste. Infine e' stato fatto un dettagliato lavoro di documentazione e supporto puntuale per consentire agli utenti di poter fare un utilizzo corretto e proficuo dei servizi resi disponibili.

Il dimensionamento delle risorse si e' rivelato idoneo all'aumento di carico verificatosi nel seguente anno e mezzo di attivita', ed ha margini che garantiscono la piena funzionalita' entro il tempo previsto di altri due anni, oltre il quale si rendera' probabilmente necessaria la sostituzione dell'hardware.

Il mantenimento del sistema ha comportato un carico di lavoro importante durante la messa a punto, valutabile in 0.5 fte, ma relativamente leggero dopo (0.2 fte, essenzialmente dovuti alla soluzione di problemi specifici degli utenti piu' che del sistema in se').

Per quanto riguarda il tipo e la qualita' dei servizi forniti agli utenti, l'unico metro disponibile e' il feed-back fornito dagli utenti stessi: l'uptime dei sistemi e' soddisfacente, in quanto le mancanze di servizio sono sporadiche, quasi sempre legate a motivi indipendenti dai server coinvolti e per la maggior parte pianificati per motivi di manutenzione di sala macchine, quadri elettrici, UPS. L'utenza oggi non manifesta mancanza di funzionalita': i soli problemi lamentati sono dovuti allo sporadico calo di efficienza dei filtri antispam, generalmente conseguenti a configurazioni non correttamente calibrate.

L'esperienza maturata nel corso di 18 mesi in produzione ha evidenziato i seguenti punti sui quali si porta' lavorare nel futuro prossimo:

- mantenimento e miglioramento dell'efficienza dei filtri antispam, con l'eventuale introduzione di tecniche quali l'auto apprendimento (filtri bayesiani) o soluzioni di trusting dei relay mittenti che potrebbero coinvolgere configurazioni globali dell'INFN o della comunita' HEP;
- realizzazione di soluzioni di ridondanza per il mailbox service, attualmente non supportate dai software disponibili su piattaforma linux; in particolare si vogliono indagare soluzioni di clustering;
- realizzazione di soluzioni di alta affidabilita' dei sistemi, utilizzando tecniche di virtualizzazione dei server, che garantirebbero la possibilita' di operare un piu' frequente upgrade dei pacchetti software utilizzati con un minimo impatto sull'utenza;
- miglioramento della soluzione di backup delle mail, che attualmente comporta tempi piuttosto lunghi (fino ad una giornata intera) per il recupero di mail in virtu' dell'aumento della dimensione dei dati (oggi circa 160 GB);

- realizzazione di sistemi di allarmistica che possano utilizzare messaggistica di tipo SMS per la notifica di malfunzionamenti;
- eliminazione dei residui meccanismi di autenticazione non integrati nel sistema centrale (vedi Mailman);

11 SCHEMA DEL SISTEMA

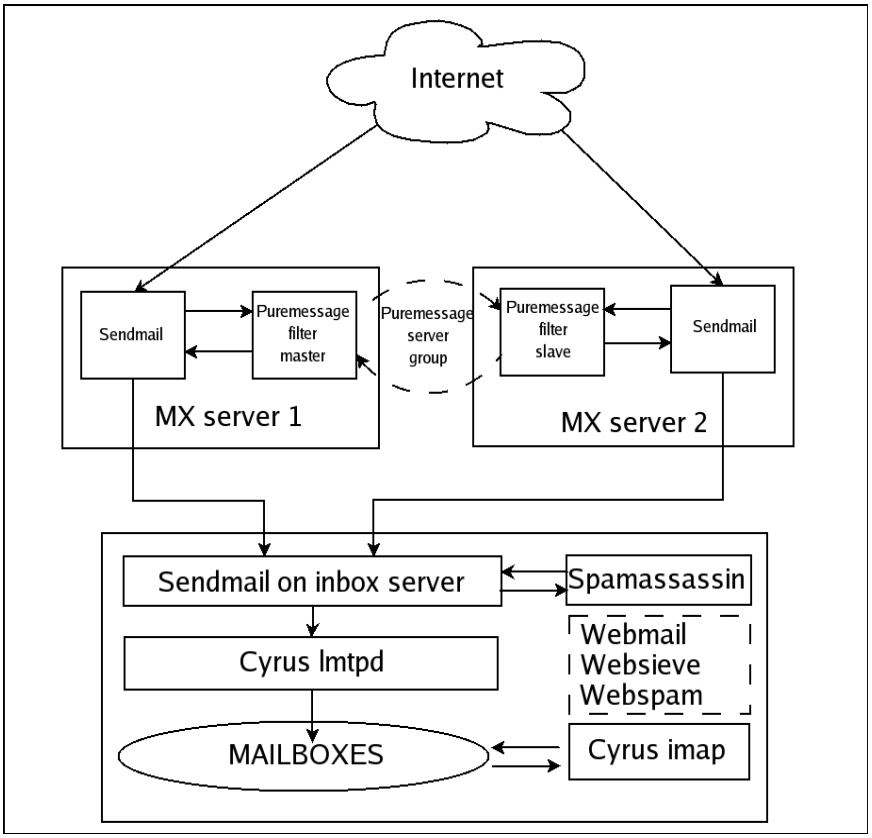

Fig 11.1

La figura 11.1 rappresenta lo schema finale del sistema descritto.

.

APPENDICE A – ESEMPIO DI SCRIPT SIEVE

```
require "fileinto";


### Sposta lo SPAM in una folder predefinita
if header :contains "Subject" "[SPAM]" {
    fileinto "INBOX.Spam";
### Sposta il messaggio indirizzato ad una mailing list
} elsif header :contains ["to", "Cc"] "mailinglist-name" {
    fileinto "INBOX.MailingList-Name";
}
```

APPENDICE B – FILE DI CONFIGURAZIONE

File di configurazione di sendmail sul relay server:

```
 divert(-1)
#
# file di configurazione relay.mc si sendmail
#
divert(0)
include(`../m4/cf.m4')
VERSIONID(`@(#)relay.mc 8.12 (Berkeley) 9/7/99')
OSTYPE(linux)dnl
define(confDOMAIN, `ge.infn.it')
MASQUERADE_AS(confDOMAIN)dnl
define(confUSERDB_SPEC, `/etc/mail/userdb.db')
define(`ALIAS_FILE', `/etc/mail/aliases,/etc/mail/mailman_aliases')
define(`confBIND_OPTS',`-DNSRCH -DEFNAMES')
define(`LOCAL_SHELL_PATH', `/usr/sbin/smrsh')
MASQUERADE_DOMAIN(confDOMAIN)dnl
DOMAIN(confDOMAIN)dnl
FEATURE(`nocanonify')
FEATURE(`always_add_domain')
FEATURE(use_cw_file)dnl
FEATURE(allmasquerade)
FEATURE(masquerade_envelope)
FEATURE(masquerade_entire_domain)
FEATURE(relay_entire_domain)dnl
FEATURE(access_db)
EXPOSED_USER(postmaster)

#
```

```
# Configurazione domini locali
#
Cw ge.infn.it
Cw genova.infn.it
Cw dns2.ge.infn.it

FEATURE(smrsh)

#
# Definizione dello userdb per riscrivere
# gli indirizzi dei destinatari
#
Kuserdb btree -o confUSERDB_SPEC

LOCAL_CONFIG

#
# Configurazione dei certificati e metodo di autenticazione
#
define(`CERT_DIR', `/usr/share/ssl/private')dnl
define(`CACERT_DIR', `/usr/share/ssl/private/CA')dnl
define(`confCACERT_PATH', `CACERT_DIR')dnl
define(`confCACERT', `CACERT_DIR/INFNCA.pem')dnl
define(`confSERVER_CERT', `CERT_DIR/smtp-cert.pem')dnl
define(`confSERVER_KEY', `CERT_DIR/smtp-key.pem')dnl
define(`_CERT_REGEX_SUBJECT_',`-s1
     (/C=IT/O=INFN/L=.+)(/CN=.+)(/Email=.+)')dnl
TRUST_AUTH_MECH(`GSSAPI DIGEST-MD5 CRAM-MD5 LOGIN')dnl
define(`confTRUST_AUTH_MECH',`GSSAPI DIGEST-MD5 CRAM-MD5 LOGIN')dnl
define(`confAUTH_MECHANISMS',`GSSAPI DIGEST-MD5 CRAM-MD5 LOGIN')dnl
define(`confDONT_BLAME_SENDMAIL',`GroupReadableSASLDBFile')
define(`confAUTH_OPTIONS',`A p y')dnl

#
# Configurazione dei demoni (porta su cui ascoltare, nome demone)
#
DAEMON_OPTIONS(`Port=smtp, Name=MTA')dnl
DAEMON_OPTIONS(`Port=465, Name=MTA-SSL, M=s')dnl
define(`confMAX_DAEMON_CHILDREN', `50')dnl

#
# Definizione del filtro antispam
#
INPUT_MAIL_FILTER(`Sophos',`S=inet:3366@localhost,F=T,T=C:5m;E:8m;R:4m;S:
     2m')
define(`confMILTER_LOG_LEVEL', 9)
```

```
#
# Splitting delle e-mail inviate a piu' destinatari
#
QUEUE_GROUP(`mqueue', `F=f, r=1, R=16')

############################################################
LOCAL_RULE_0
Rbb + $+ < @ $=w . >    $#cyrusbb $: $1

LOCAL_RULE_1
############################################################
### Local Ruleset 1, rewrite sender header & envelope ##
############################################################
#Thanks to Bjart Kvarme <bjart.kvarme@usit.uio.no>
S1
R$-           $1 < @ $j . >           user =>user@localhost
R$- < @ $=w . > $* $: $1 < @ $2 . > $3 ?? $1    user@localhost ?
R$+ ?? $+     $: $1 ?? $(userdb $2 : mailname $: @ $)
R$+ ?? @      $@ $1                   Not found
R$+ ?? $+     $>3 $2                  Found, rewrite


MAILER(local)
MAILER(smtp)
```

File di configurazione di cyrus imapd.conf:

```
#
# Percorsi di configurazione e mailbox parition
#
configdirectory: /var/imap
partition-default: /var/spool/imap

admins: cyrus root

#
# Demone di autenticazione
#
sasl_pwcheck_method: saslauthd

allowanonymouslogin: no
#
# Certificati per imaps
#
tls_cert_file: /var/imap/mbox-sum.pem
tls_key_file: /var/imap/mbox-sum.pem
```

File di configurazione di cyrus cyrus.conf:

```
START {
  # Esegue all’avvio il controllo dei database
  recover cmd="ctl_cyrusdb -r"
}

# I socket unix sono creati in /var/imap/socket
SERVICES {
  # definizione dei servizi da attivare
  imaps       cmd="imapd -s" listen="imaps" prefork=0
  sieve       cmd="timsieved" listen="sieve" prefork=0

  # LMTP deve essere sempre presente
  lmtpunix    cmd="lmtpd" listen="/var/imap/socket/lmtp" prefork=0
  prefork=1
}

#Job da eseguire a periodi determinati
EVENTS {

  # scrive su disco I db caricati in memoria (obbligatorio)
  checkpoint  cmd="ctl_cyrusdb -c" period=30

  # Cancella alle 04:00 di ogni giorno i duplicati (se configurato)
  delprune    cmd="ctl_deliver -E 3" at=0400

  # Cancella le sessioni tls se e’ attivo il caching
  tlsprune    cmd="tls_prune" at=0400
}
```

File di configurazione di spamassassin local.cf:

```
#
# Riscrittura del subjecr e definizione della soglia
#
required_hits 3
rewrite_subject 1
subject_tag [[SPAM]]

bayes_auto_learn 0

######################
#attivo il timeout su check mx
check_mx_delay 3
######################
```

```
####################
# whitelist

whitelist_from *-bounces@ge.infn.it

#
# Configurazione di DCC
#
use_dcc 1
dcc_dccifd /usr/libexec/dcc/dccifd
dcc_home /etc/dcc
dcc_options -Rw whiteclnt

#
# Disattivazione del filtro bayesiano
#
use_bayes      0

#
# attivo i check rbl con timeout 5 sec.
#
rbl_timeout 5


#
# attivazione di razor e pyzor
#
use_razor2 1
use_pyzor  1
razor_timeout 2
razor_config /etc/razor/razor-agent.conf

#
# riscrittura di una regola
#
score SORTED_RECIPS 2.0

#
# Non tagga i messaggi provenienti dal webmail
#
header LOCAL_WEBMAIL_HEADER X-WebMail-Server =~ /INFN-GE/
score LOCAL_WEBMAIL_HEADER -60

#
# non tagga gli utenti autenticati
#
```

```
header            LOCAL_SMTPMAIL_HEADER2            Received            =~
/authenticated\sbits=0\).+by\sdns2\.ge\.infn\.it/
score LOCAL_SMTPMAIL_HEADER2 -50

#
# Assegna un punteggio di 40 ai messaggi riconosciuti da Sophos
#

header LOCAL_SOPHOS_SPAMFILTER exists:X-Spam-Sophos-Genova
score LOCAL_SOPHOS_SPAMFILTER 40
```

APPENDICE C – CARATTERISTICHE DELLE MACCHINE

Inbox server:

- Modello: Dell 2650
- CPU: doppio processore Intel Xeon 2.4 GHz
- Memoria RAM: 2GB
- Storage: 5 dischi SCSI in configurazione RAID5 per un totale di 420 GB
- Doppio alimentatore ridondato
- Doppia scheda Broadcom Netextreme BC5703 Gigabit Ethernet

Relay primario:

- Modello: Dell 1750
- CPU: doppio processore Intel Xeon 3 GHz
- Memoria RAM: 2GB
- Storage: 2 dischi SCSI in mirror per un totale di 70 GB
- Doppio alimentatore ridondato
- Doppia scheda Broadcom Netextreme BC5704 Gigabit Ethernet

Relay secondario:

- Modello: Dell 1750
- CPU: doppio processore Intel Xeon 3 GHz
- Memoria RAM: 2GB
- Storage: 2 dischi SCSI in mirror per un totale di 70 GB
- Doppio alimentatore ridondato
- Doppia scheda Broadcom Netextreme BC5704 Gigabit Ethernet

RIFERIMENTI

(1) http://www.ietf.org/rfc.html
(2) http://www.sendmail.org/
(3) http://www.namesys.com/X0reiserfs.html
(4) http://asg.web.cmu.edu/cyrus/ag.html
(5) http://asg.web.cmu.edu/cyrus/config.html
(6) http://prdownloads.sourceforge.net/cyrus-utils
(7) http://www.irbs.net/internet/info-cyrus/0210/0024.html
(8) http://spamassassin.apache.org/
(9) http://openrbl.org/
(10) http://razor.sourceforge.net/
(11) http://pyzor.sourceforge.net/
(12) http://www.rhyolite.com/anti-spam/dcc/
(13) http://www.sophos.com
(14) http://www.ge.infn.it/~corosu/webspam.tar.gz
(15) http://www.horde.org
(16) http://www.gnu.org/software/mailman/