ALMA MATER STUDIORUM - UNIVERSITA' DI BOLOGNA

FACOLTA' DI SCIENZE MATEMATICHE, FISICHE E NATURALI

CORSO DI LAUREA IN SCIENZE DI INTERNET

"IL PROBLEMA DELLA POSTA ELETTRONICA NON DESIDERATA (SPAM) "

Tesi di Laurea in SCIENZE DI INTERNET

Presentata da:

Domenico Torre

Relatore:

Dott. Moreno Marzolla

Sessione III

Anno Accademico 2009-2010

Ai miei genitori,che mi hanno dato questa meraviglioso opportunità di studio e di vita. Un grosso ringraziamento anche a coloro che mi hanno aiutato in questi anni di vita bolognese. Cosa aggiungere ancora.....grazie grazie grazie a tutti.

Indice

Sommario

Questo elaborato analizza uno dei servizi più utilizzati, se non il più utilizzato, dagli utenti all'interno della rete: la posta elettronica. A tutti sarà capitato di inviare o ricevere una e-mail, così come di possedere uno o più account di posta elettronica: ci sono persone, ad esempio, che posseggono un account per il lavoro, uno per l'università e uno privato per comunicare con i propri amici ecc. Proprio a causa di questa caratteristica del servizio di posta elettronica, ovvero l'ampia diffusione e il semplice utilizzo, coloro che se ne servono il più delle volte non ne sono informati sull'esatto funzionamento, così come non ne conoscono a fondo le problematiche né tanto meno come risolverle. In realtà le cose non stanno precisamente così. In questo elaborato si vuole approfondire il funzionamento della posta elettronica, quali problemi essa può presentare e come poterli risolvere o evitare.

Innanzitutto vedremo come funziona nello specifico la posta elettronica, quali sistemi utilizza, di quali protocolli fa uso.Un accenno al protocollo S/MIME, che ha come scopo principale la sicurezza delle e-mail, infatti ogni volta che riceviamo un messaggio di posta elettronica, questo non ha nessuna forma di autenticazione, ovvero non si può avere l'assoluta certezza dell'identificazione del mittente; il protocollo S/MIME cerca di risolvere il problema dell'autenticazione. Successivamente verrà trattato il problema dello Spam, attraverso l'analisi delle sue diverse tipologie, come agisce e che cosa provoca sulla posta elettronica e cosa prevede la legislazione a riguardo. Tra gli ostacoli al corretto e soddisfacente utilizzo della posta elettronica, il principale e più diffuso è certamente lo Spam. Tutti i possessori di un account di posta elettronica avranno ricevuto almeno una volta una e-mail di posta indesiderata; in questo lavoro vedremo proprio come poter evitare questo inconveniente.

Lo scopo di questa tesi è quello di comprendere quali siano gli accorgimenti, le tecniche e i sistemi per poter evitare, arginare ed eventualmente eliminare il problema della posta indesiderata o Spam.

1 Introduzione alla posta elettronica

La posta elettronica, si può considerare come uno dei servizi più consolidati e usati nelle reti. Questo è un servizio che permette lo scambio di messaggi tra utenti che fanno parte della rete. Gli indirizzi di posta elettronica sono così formati: username@hostname, dove username sarà l'interlocutore (sia mittente che destinatario), che è univoco per hostname, e hostname è un indirizzo IP o un nome DNS. Va ricordato che l'indirizzo di posta elettronica non è associato al singolo utente, ma alla casella di posta elettronica che fornisce il servizio. La nascita della posta elettronica risale al 1972, quando uno studioso americano (Ray Tomlinson), installò su ARPANET un sistema che permetteva lo scambio di messaggi fra diverse università. Tanti sono i motivi del successo della posta elettronica: velocità, facilità d'uso, accessibilità ed economicità. Infatti la posta elettronica permette in maniera semplice di inviare messaggi a più persone contemporaneamente, allegare documenti, permette lo scambio di informazioni multimediali, come ad esempio suoni, immagini, hypelink, video. Nonostante ciò, anche la posta elettronica presenta dei limiti: sicurezza nell'invio e difficoltà nell'identificare il mittente. Anche se, per questi ultimi due problemi presentati vi è una "soluzione", ovvero la ricevuta di ritorno quando si invia un messaggio, e le applicazioni di firma digitale. Lo scopo di questa tesi è di esaminare il funzionamento della posta elettronica e trattare uno dei suoi principali problemi: lo Spam e come aumentarne la sicurezza onde evitare spiacevoli inconvenienti.

1.1 Componenti di un sistema di posta elettronica : MUA-MTA

La posta elettronica si basa su un' architettura client - server, divisa in due ruoli: spedizione e ricezione. Cosi la posta viene implementata in Internet tramite la collaborazione di due componenti principali:

- MUA (Mail User Agent): Il MUA, chiamato anche *mail reader*, è il client; è un programma di gestione di posta attivo sul client. Permette agli utenti di comporre, ricevere e leggere i messaggi. Conosce la sintassi di composizione dei messaggi(RFC822 e MIME), conosce il protocollo(POP3, IMAP4), conosce il protocollo che permette la spedizione dei messaggi (SMTP), che sarà in grado di consegnarli al MTA utile per la trasmissione.
- MTA (Mail Trasfer Agent) : MTA, chiamato anche *mail server*, rappresenta un ponte fra due MUA: è l'interfaccia con la rete che si occupa della ricezione e trasmissione dei messaggi. MTA può essere o un server SMTP per la spedizione e la ricezione di messaggi verso e da altri server Smtp, un POP3 per la spedizione dei messaggi al client, e IMAP4 che permette la gestione dei messaggi sul server dal client.

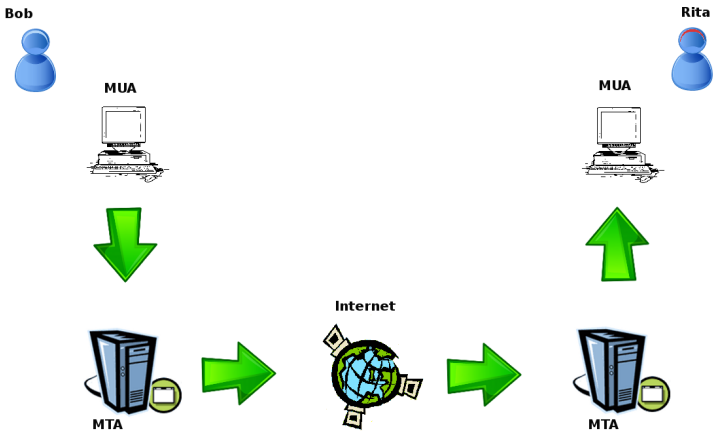

Figura 1.1: Flusso di un messaggio e-mail su internet

1.2 Funzionalità

La spedizione e la ricezione dei messaggi di posta è divisa in varie fasi, che seguono un determinata sequenza di operazioni, Infatti il mittente, finito di comporre il messaggio, il suo user agent invia il suo messaggio di posta al server dove viene accodato nei messaggi di posta in uscita. Successivamente l' e-mail viene spedita al server di posta del destinatario, che si occuperà di collocarla nella rispettiva mailbox. Qualora il server di posta del destinatario non dovesse essere attivo, il server di posta del mittente collocherà il messaggio in una coda, e riproverà a spedirlo ogni trenta minuti circa. Nel caso in cui il messaggio non dovesse essere inviato dopo un po' di giorni, allora verrà cancellato, e sarà data notifica da parte del server dell'avvenuta cancellazione del messaggio al mittente. Per poter leggere i messaggi ricevuti nella propria mailbox, il mittente dovrà autenticarsi con username e password. In ogni caso i due servizi svolgono il rispettivo lavoro su porte diverse: 25 per il primo e 110 per il secondo.

1.3 Lo standard RFC822 prima e dopo MIME (RFC 1341)

Come tutti i messaggi di posta tradizionali, anche quelli di posta elettronica contengono informazioni come indirizzo del destinatario, indirizzo del mittente e la data. Tutte queste informazioni nella posta cartacea sono visibili esternamente, mentre nella posta elettronica vengono incluse nel messaggio tramite alcune righe di intestazione, che precedono il corpo del testo lasciando una riga vuota. Solitamente un messaggio di posta elettronica si compone di due sezioni: la sezione dell'intestazione e quella del corpo della mail. Altri tipi di sezioni potrebbero essere presenti nel caso in cui la mail dovesse contenere un allegato. Le intestazioni contengono informazioni utili, che vengono interpretate dal client di posta per stabilire da chi è stata inviata la mail, a chi è diretta, se ci sono più destinatari o nel caso in cui vi sono degli allegati. Alcune di queste informazioni servono al client di posta per identificare il formato del corpo, il formato dell'allegato o per il tipo di codifica di eventuali allegati. Rispetto alla posta cartacea, tutte queste intestazioni sono visibili in seguito quando leggiamo la nostra mail. Di seguito al testo della mail troviamo la sezione degli eventuali allegati, i quali sono inglobati nella e-mail e vengono codificati con la tabella ASCII

7 bit: con questi ultimi troviamo, nel caso ci siano altre informazioni, il nome del file e suo il tipo. Originariamente il protocollo per la rappresentazione dei documenti della posta elettronica era definito dal documento RFC 822 [1], risalente al 1982. All'interno di questo documento veniva specificato il formato per i messaggi di posta elettronica e ci si limitava a messaggi esclusivamente di tipo testo ASCII, senza nessun altra possibilità di messaggi di altro tipo, come ad esempio immagini. Nel giugno del 1992 venne presentato un nuovo documento che rappresentava il nuovo formato della posta elettronica RFC 1341 [2] in cui è descritto lo standard MIME. Con questo nuovo formato si vogliono superare alcune delle limitazione della RFC 822. Infatti grazie allo standard MIME viene definito sia il formato dei messaggi testuali (ASCII), sia dei messaggi di posta multimediale, cioè che contengono immagini, video, suoni ecc. Una delle principali limitazioni del RFC 822 è che il contenuto dei messaggi è limitato a caratteri di 7 bit, quindi un testo non composto interamente in codice ASCII, deve prima essere convertito per poi essere inviato in rete, cosi per poter risolvere questo problema è stato indicato RFC 1341.

1.3.1 RFC 822

RFC 822 è uno standard internet. Un messaggio di posta elettronica secondo questo standard è formato da una busta e un contenuto. La busta, detta *header* contiene l'intestazione del messaggio, cioè tutte le informazioni come il mittente, il destinatario e la data di spedizione. Ecco un esempio di quello che contiene l' header:

- To: lista destinatari
- From: mittente
- CC: lista di destinatari per conoscenza
- BCC: lista nascosta di destinatari per conoscenza
- Date: data di spedizione
- Reply - to: indirizzo diverso dal mittente
- Subject: titolo del messaggio

Il contenuto, detto *body*, è separato dall'*header* da una linea vuota: rappresenta il testo del messaggio ed eventuali file, dove il messaggio è formato usando solo caratteri ASCII.

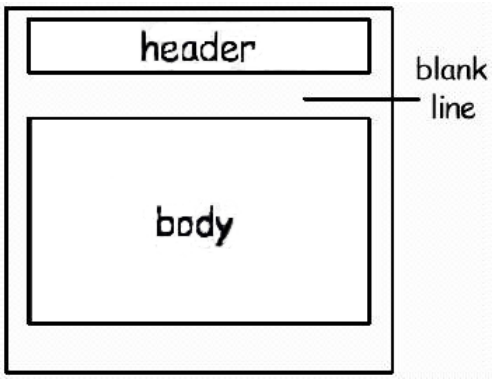

Figura 1.2: Modello RFC 822

1.3.2 RFC 1341 o MIME

Quando nella rete internet abbiamo due utenti che si scambiano un messaggio di posta elettronica tra loro, dove uno invia e l'altro riceve, il primo che invia il file dovrà specificare il tipo di file tramite lo standard MIME (Multipurpose Internet Mail Extensions), in modo che il secondo, che riceverà i dati, saprà come trattarli. Tramite lo standard MIME è possibile inserire, in una qualsiasi e-mail, oltre al testo, anche file audio e video, oppure file immagini. In questo modo il software che gestisce la posta non si preoccupa del contenuto della e-mail, ma sarà l'utilizzatore finale a gestire l'opportuna decodifica, in base alle specifiche inserite nel messaggio stesso. In MIME vengono introdotte delle estensioni che permettono di risolvere i problemi dello standard precedente (RFC 822) ad esempio inserimento di messaggi multimediali e definire il formato di messaggi testuali e non, senza introdurre delle incompatibilità con i documenti scritti con l' RFC 822. Al l' interno di un documento MIME vi è un'intestazione che contiene i seguenti campi:

- *MIME Version* : serve a identificare la versione dello standard MIME che viene usato nel messaggio.

- *Content-Trasfer-Encoding* : Specifica un metodo di codifica dei dati accessorio a quello principale; permette ai dati di transitare attraverso tutti i meccanismi di trasporto della posta elettronica, poiché potrebbero avere dei limiti nei set di caratteri ammessi. Molto importante sarà applicare questa codifica ai dati prima che questi vengano trasmessi in rete. Infatti i documenti per essere trasportati nella rete hanno bisogno di un ulteriore codifica: questa avviene nel *Content-Trasfer-Encoding* dove viene specificato che relazione c'è fra i dati nel loro formato originale ed il formato con cui vengono trasmessi. Questo campo permette una trasmissione di dati nella rete senza alcun problema, anche se i dati dovessero passare attraverso una rete simile allo standard RFC 822 e non alle sue seguenti versioni. Un esempio è un sistema di posta adattabile con il protocollo SMTP, dove è necessario che il documento venga codificato in caratteri ASCII a 7 bit, non oltre le mille linee. Questo campo viene usato per specificare come possono essere trattati i dati per essere trasportati in rete. I valori del *Content-Trasfer-Encoding* sono: 7bit, 8bit, binary, quoted-printable, base64.

7bit, 8bit, binary indicano che nessuna operazione di codifica è stata utilizzata sul contenuto del messaggio, e, contemporaneamente forniscono delle indicazioni sui tipi di dati contenuti nel messaggio stesso, rendendo possibile il tipo di codifica che potrebbe servire per trasmettere il messaggio in determinati sistemi di trasmissione.

Il valore 7 bit, sta a significare il caso in cui i dati possono essere rappresentati in gruppi di sette bit, dove ognuno rappresenta un carattere ASCII. Tale valore è assunto di default nel caso in cui non dovesse esser specificato il campo.

Il valore 8 bit, indica che potrebbero esserci dei caratteri non appartenenti al set ASCII. Quindi, frammentando il messaggio in linee di 8 bit ciascuna e associando ad ogni linea un carattere ASCII, si potrebbero ottenere delle sequenze di caratteri in apparenza senza significato.

Binary, indica che il contenuto del messaggio è in formato binario, ad esempio se ci dovesse essere un' immagine, un file audio o video.

Quoted-Printable, indica che un' operazione di codifica è già stata utilizzata sui dati, così da trasformare il messaggio in una serie di caratteri ASCII; nel caso in cui il messaggio è già formato da ca-

ratteri ASCII, resta inalterato. Questo tipo di codifica ha lo scopo di mettere i dati in un formato che difficilmente sarà alterato da parte dei vari sistemi che il messaggio attraverserà prima di arrivare a destinazione.

Base 64, indica che sui dati è stata applicata una codifica a base 64. Con questa codifica il messaggio viene trasformato in una sequenza di caratteri, che fa parte di un sottogruppo del set caratteri ASCII, e principalmente viene applicata ai caratteri che vanno da le lettere maiuscole "A" a " Z", a quelle minuscole "a" a "z", e dai numeri da "0" a "9" e al carattere "\".

Così facendo, ogni carattere codificato può essere formato con sei bit. L'operazione di codifica permette che il messaggio venga suddiviso in gruppi di 24 bit, dove ogni gruppo viene diviso in quattro gruppi di sei bit, ai quali è associato il proprio carattere ASCII corrispondente e appartenente al sottogruppo specificato.

X-Token, indica un modello di codifica esterno (*token* è il nome dato alla codifica), non standard. Molto importante sarà rendere noto a chi riceve il messaggio di questa codifica: in tal modo il destinatario saprà come ricostruire il messaggio in maniera corretta.

- *Content-Type*: indica il tipo e il sottotipo di dati contenuti nel messaggio, in modo che il software che riceve il messaggio possa capire il tipo dei dati ricevuti. La forma di questo campo è cosi composta: *Content-Type: tipo/sottotipo; [parametro]*, dove il "tipo" specifica la forma generale dei dati, e "sottotipo" specifica il tipo particolare di dati trasmessi. Il campo "parametro" è opzionale. Facciamo un esempio di come possa essere veramente: Content-Type: text/plain; charset = us-ascii: in questo caso "Text" è un tipo che sta a rappresentare informazioni in formato testuale, scritte secondo un determinato tipo di linguaggio, mentre "Plain" è un sottotipo di "Text" che indica un testo non formattato. Un altro esempio di sottotipo di "Text" potrebbe essere "Richtest" che indica un testo con una semplice formattazione. Invece il parametro, usato per i messaggi Text è Charset, che indica i tipi di caratteri si stanno utilizzando. Nel caso in cui non dovesse esser specificato il parametro, verrà assunto di default il set di caratteri US-ASCII.

- *Content-ID*: campo opzionale che serve ad identificare in maniera univoca il messaggio

- *Content-Description*: campo opzionale che serve a descrivere il contenuto testuale del messaggio

Tutti i tipi di dati MIME, utilizzati dal campo Content-Type, devono essere registrati presso IANA (Internet Assigned Numbers Authority). Infatti tutti i nuovi tipi di dati non ancora riconosciuti dalla IANA, dovranno essere contrassegnati da una "x", ad esempio multipart/x-mixed.

1.4 I protocolli

La posta elettronica utilizza il protocollo SMTP per l'invio di messaggi, anche l'agente dell'utente lato mittente, il quale, non potendo comunicare direttamente con il MTA del destinatario, si affida al protocollo SMTP per trasferire i suoi messaggi al suo MTA. Il MUA del destinatario non può utilizzare il protocollo SMTP per il recupero dei messaggi di posta, poiché SMTP è un protocollo push, che, tradotto letteralmente significa "spingi". Per quando riguarda la ricezione di messaggi, di posta elettronica vengono utilizzati altri due protocolli, chiamati POP3 e IMAP.

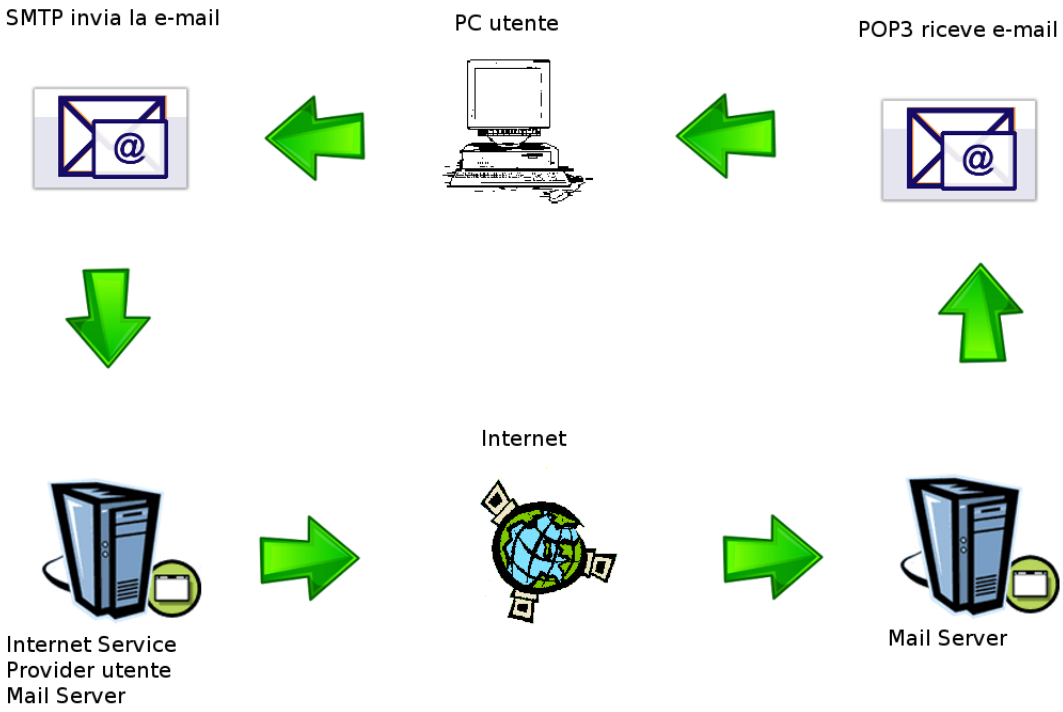

Figura 1.3 : Funzionamento posta elettronica

1.4.1 SMTP

Il protocollo SMTP [3] rappresenta il nucleo, ed è anche il protocollo principale dello strato di applicazione della posta elettronica così definito da RFC 821, che è stata sostituita dalla RFC 2821 nell' agosto del 1982. Vengono considerati anche altri standard: la RFC 822 che descrive la sintassi degli headers della mail e il formato del messaggio, la RFC 1049, che definisce le strutture dati per interpretare correttamente il contenuto delle mail e la RFC 974, che si occupa del routing delle mail tramite DNS, che lavora anche in stretta collaborazione con l'estensione MIME che definisce intestazioni extra per il formato dei messaggi, descritto nelle RFC che vanno dalla 2045 alla 2049. SMTP supporta solo il formato ASCII a sette bit per il corpo e le intestazioni dei messaggi: questo non era un problema negli anni '80, ma adesso, con lo sviluppo raggiunto nella capacità e velocità di trasmissione e la possibilità di inviare messaggi multimediali con file video, audio o immagini questa restrizione diventa un limite. Lo standard SMTP prevede alcune ingegnosità come quella di non obbligare il trasferimento di un messaggio più di una volta, qualora questo abbia più destinatari, o quella di verificare prima l'esistenza del luogo di destinazione, evitando così di lasciar trasferire invano i dati sul server per poi scoprire che il messaggio non è giunto a destinazione, a causa di un errore nella digitazione dell'indirizzo o per la chiusura di quell'indirizzo di posta elettronica. Il compito principale di SMTP riguarda il trasferimento delle e-mail tra i server di posta. Per fare questo si appoggia al servizio trasferimento affidabile dei dati del protocollo di trasporto TCP per trasferire la posta dal server di posta mittente a quello del destinatario. Così come accade per buona parte dei protocolli a livello di applicazione, SMTP presenta un lato client, in esecuzione sul server di posta del mittente, e un lato server, in esecuzione su quello del destinatario. Entrambi i lati girano su tutti i server di posta. Così quando un server invia posta a un altro, agisce da client SMTP, quando invece la riceve, funziona come server SMTP.

SMTP può essere schematizzato cosi :

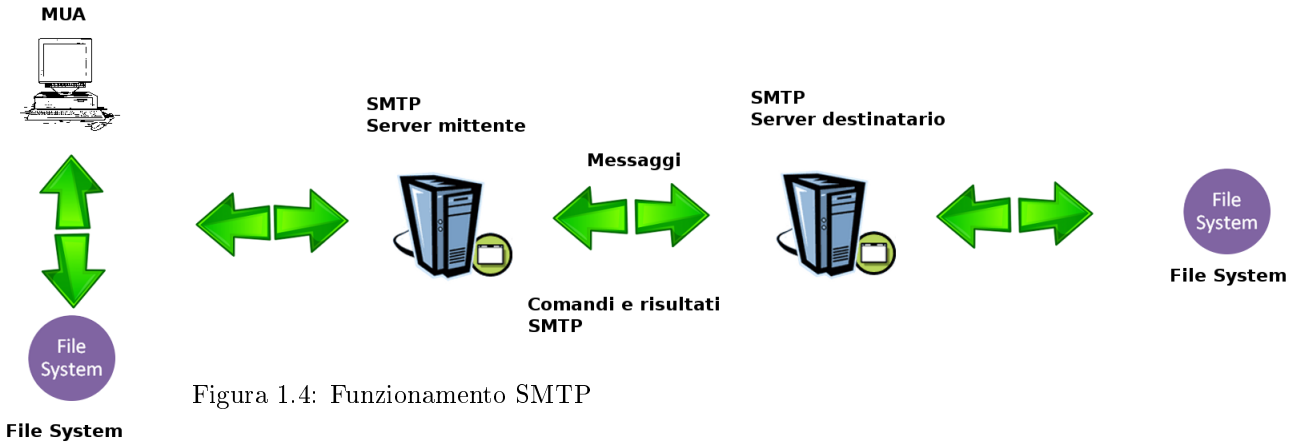

Figura 1.4: Funzionamento SMTP

SMTP utilizza quindi il protocollo affidabile TCP sulla porta 25 e di solito effettua una trasmissione diretta fra server di posta mittente e destinatario non appoggiandosi a nessun server intermedio. La spedizione è articolata nelle seguenti fasi:

- handshaking: apertura della connessione TCP, fase di presentazione necessaria per sincronizzare i server e prepararsi per il trasferimento del messaggio. In questa fase il client presenta l'indirizzo e-mail del mittente e del destinatario.
- trasferimento del messaggio che sarà suddiviso a sua volta in:

1. inizio invio del messaggio di posta
2. elenco dei destinatari
3. invio del messaggio di posta

- chiusura: chiusura della connessione TCP. Essendo una connessione persistente la chiusura della connessione avviene solo quando non ci sono più messaggi da spedire. Per ogni messaggio vengono ripetute le fasi di handshaking e trasferimento.

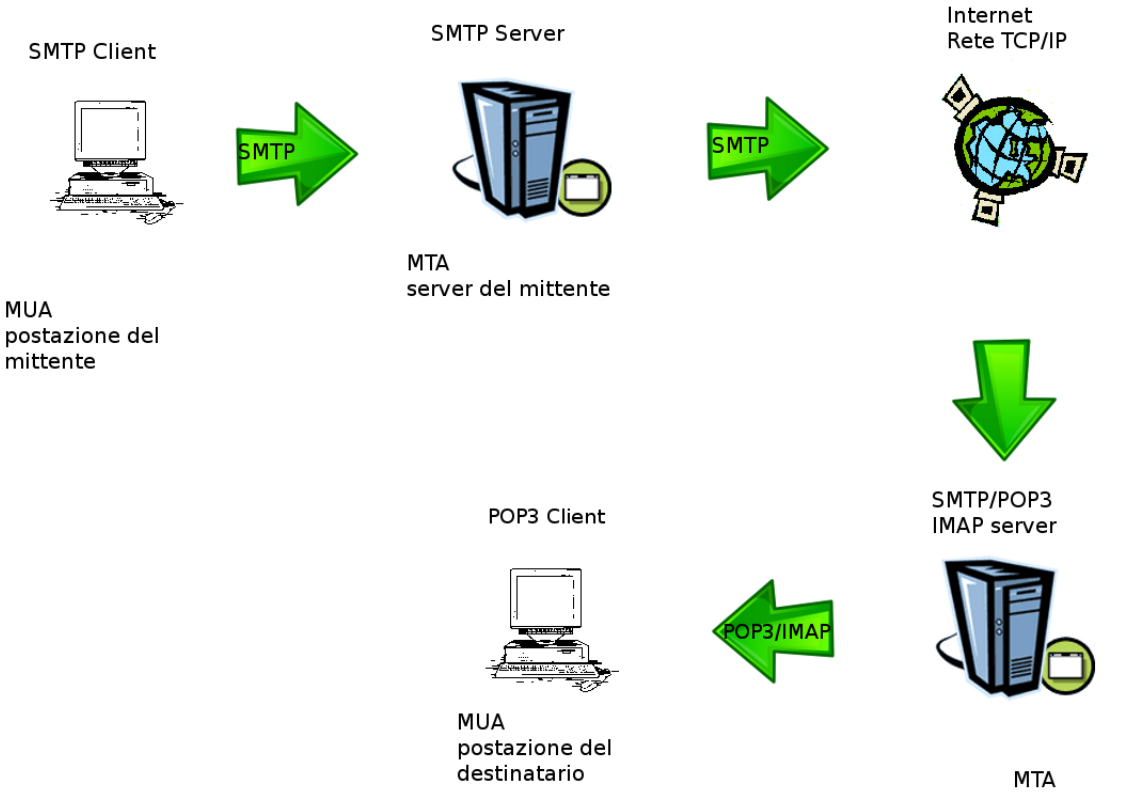

Figura 1.5: Operazioni base del protocollo SMTP e POP 3

Vediamo ora nel dettaglio le operazioni del protocollo SMTP:

1. Il mittente invoca il proprio MUA, fornisce l'indirizzo di posta del destinatario, compone il messaggio e dà istruzione all'agente utente di inviarlo.
2. Il MUA del mittente invia il messaggio al suo server di posta, dov'è collocato in una coda di messaggi.
3. Il lato client di SMTP, che gira sul lato server di posta del mittente, vede il messaggio nella coda dei messaggi ed apre una connessione TCP verso un server SMTP, che gira sul server di posta del destinatario.
4. Dopo un handshake SMTP, il client SMTP invia il messaggio del mittente sulla connessione TCP.
5. Presso il server di posta del destinatario, il lato server di SMTP riceve il messaggio. Il server di posta del destinatario quindi lo posiziona nella casella di posta
6. Il destinatario, quando lo ritiene opportuno, invoca il proprio MUA per leggere il messaggio.

Importante è osservare che di solito SMTP non usa server intermedi. In particolare, se il server di posta del destinatario è spento, il messaggio rimane nel server di posta del mittente e attende un nuovo tentativo. Ora vediamo i principali comandi del protocollo SMTP.

HELO: identifica il client SMTP al server SMTP, ovvero il computer che si sta rivolgendo al server.

MAIL: richiede al server ricevente l'accettazione di un messaggio di email per l'inoltro.

MAIL FROM: <indirizzo mittente>: Indica la casella di posta del mittente, ovvero l'indirizzo di ritorno al quale riportare eventuali errori.

RCPT TO: <indirizzo destinatario> : indica la casella di posta del destinatario (RCPT "Recipient"). Si può indicare più di un destinatario ripetendo il comando RCPT TO.

DATA: indica al server che quanto digitato successivamente saranno i dati del messaggio di posta, si segnala la fine del messaggio inviando una riga contenente solo un punto.

RSET: annulla i comandi (Reset) precedentemente inviati nella sessione SMTP corrente.

VRFY <stringa>: chiede al server se la stringa di testo immessa rappresenta un nome utente presente ed in tal caso visualizza l'intero indirizzo.

HELP: visualizza i comandi disponibili sul server.

NOOP: non esegue nessuna operazione restituisce solo un messaggio 250 (Ok) se il server risponde.

QUIT: termina la sessione SMTP corrente.

Ecco un esempio di sessione SMTP:

```
S: 220 www.example.com ESMTP Postfix

C: HELO mydomain.com

S: 250 Hello mydomain.com, pleased to meet you

C: MAIL FROM: <sender@mydomain.com>

S: 250 sender@mydomain.com ... Sender ok

C: RCPT TO: <friend@example.com>

S: 250 friend@example.com ... Recipient Ok

C: DATA

S: 354 End data with "." on a line by itself

C: Subject: messaggio di prova

C: From: sender@mydomain.com

C: To: friend@example.com

C:

C: Ciao, C: questa è una prova.

C: .

S: 250 Ok: queued as 12345

C: QUIT

S: 221 Bye
```

Figura 1.6: Sessione Protocollo SMTP

1.4.2 POP 3

La maggior parte dei sistemi di posta elettronica sfrutta il protocollo POP [4] (Post Office Protocol RFC 1725) per il trasferimento dei messaggi tra la propria mailbox e il client di posta. Esistono tre versioni diverse del protocollo POP, ma l'unica di fatto utilizzata è la versione 3 (POP3): fornisce un magazzino messaggi che trattiene le e-mail ricevute fino a che l'utente effettua il log-on e le scarica. POP3 è un sistema molto semplice con poca selettività. Tutti i messaggi in arrivo e gli attachment devono essere scaricati insieme. POP3 accetta messaggi formattati ed inviati tramite protocollo messaggi SMTP. POP3 entra in azione quando l'agente utente (il client) apre una connessione TCP verso il server di posta (il server). Quindi per eseguire il processo di scaricamento delle mail viene instaurata una connessione TCP sulla porta 110. Come per SMTP, una sessione POP3 consiste in uno scambio di comandi e informazioni tra il client (MUA del mittente) e il server (MTA del mittente). La sessione passa attraverso tre fasi:

1. "Autorizzazione" : l'agente utente invia nome e password per autenticare l'utente.
2. "Transazione" : l'agente utente recupera i messaggi: anche durante questa fase, può marcare i messaggi per la cancellazione, rimuovere i segni di cancellazione e ottenere statistiche sulla posta.
3. "Aggiornamento" ha luogo dopo che il client ha inviato il comando quit, che conclude la sessione POP 3. In questo istante, il server di posta cancella i messaggi che sono stati marcati per la rimozione.

Durante le diverse fasi, il server conserva le informazioni di stato, come i messaggi contrassegnati per l'eliminazione, mentre durante le sessioni POP3 non le mantiene, e ciò rende molto semplice l'implementazione di un server POP3. E' possibile che su un server POP3 sia stabilito un tempo di inattività: una volta trascorso, si viene automaticamente disconnessi senza passare nella fase di aggiornamento. Questo significa che la connessione TCP è terminata e che gli eventuali comandi di cancellazione impartiti al server non saranno presi in considerazione. Vi sono due possibili alternative

dopo che i messaggi sono stati prelevati. La prima soluzione, detta "download and delete" (scarica e cancella): dopo la copia i messaggi vengono rimossi dal server. La seconda soluzione detta "download and keep" (scarica e conserva): dopo la copia i messaggi restano nel server. Per la prima soluzione sono previsti i comandi LIST, RETR e DELE. Una questione da non sottovalutare riguarda il fatto che, usando la prima tecnica, se l'utente dovesse usare macchine diverse per scaricare la posta, i messaggi verranno divisi su di esse. Invece usando la seconda tecnica i messaggi rimangono sul server, e quindi l'utente può leggere i messaggi su macchine diverse senza che questi vengono cancellati dopo ogni accesso. Vediamo ora i principali comandi utilizzati dal protocollo POP 3:

USER <nomeutente>: identifica l'utente che si connette al server;

PASS <password>: invia in chiaro la password dell'utente che si sta autenticando;

STAT: restituisce il numero di messaggi presenti e lo spazio da essi occupato;

LIST <numero messaggio>: senza parametri, indica la dimensione di ogni messaggio, altrimenti solo quella del messaggio indicato;

RETR <numero messaggio>: visualizza il messaggio indicato;

TOP <numero messaggio>: visualizza un numero predefinito di linee dalla testa del messaggio;

DELE <numero messaggio>: cancella dal server il messaggio indicato;

NOOP: non esegue nessuna operazione, restituisce solo un messaggio +OK se il server risponde;

RSET: cancella le operazioni di cancellazione DELE in precedenza inviate al server;

QUIT: termina la sessione POP3 corrente e si disconnette dal server.

Ecco un esempio di sessione del protocollo POP 3:

S:+OK <22593.1129980067@example.com>

C:USER pippo

S:+OK

C:PASS pluto

S:+OK

C:LIST

S:+OK 1 817 2 124 .

C:RETR 1

S:+OK Return-Path: <pippo@example.org>

Delivered-To: pippo@example.org Date: Sat, 22 Oct 2
13:24:54 +0200

From: Mario Rossi <mario@rossi.org> Subject: xxxx
Content-Type: text/plain;

charset=ISO-8859-1 testo messaggio .

C:DELE 1

S:+OK

C:QUIT

S:+OK

Figura 1.7: Sessione protocollo POP 3

Vediamo ora nel dettaglio le operazioni del protocollo POP 3:

1. Il client POP3 si connette al server POP3 costantemente in ascolto sulla porta 110.
2. Una volta connesso, il server invia un messaggio di saluto al client normalmente indicando il nome e la versione del software server.
3. Fase di autenticazione dove il client invia al server POP3 i comandi USER (nomeutente) e PASS (password). Una volta che il client è stato autenticato si possono eseguire le operazioni per gestire la posta, come leggere, cancellare un messaggio ecc.

4. E' possibile inviare i comandi al pop server il quale risponde con +OK in caso di comando eseguito correttamente, contrario -ERR nel caso in cui non riesca ad interpretare il comando

5. La sessione viene chiusa con il comando QUIT.

1.4.3 IMAP

Il protocollo IMAP [5] (Internet Messaging Access Protocol RFC 1064) è un protocollo di accesso alla posta, che permette di compiere le elaborazioni direttamente sul server remoto centralizzato. Così facendo si possono avere a disposizione tutti i messaggi indipendentemente dalla macchina con la quale ci si sta connettendo. POP3, al contrario, scarica i messaggi localmente e non permette di creare cartelle remote: questo è uno dei motivi per cui è stato introdotto il protocollo IMAP come sostituzione di POP3. Viene definito nella RFC 2060, e la versione a cui è giunto è IMAP4. E' più complesso di POP3 ma nello stesso tempo offre maggiori funzionalità. Il server IMAP, all'arrivo di ogni messaggio, lo associa alla cartella INBOX del destinatario, che poi lo amministrerà secondo le proprie esigenze. Anche IMAP permette agli utenti di realizzare cartelle e spostare i messaggi come POP3: l'ulteriore funzionalità è quella di crearle sul server remoto, quindi anche la ricerca viene effettuata su server remoto. IMAP resta in possesso delle informazioni tra una sessione e l'altra, a differenza di POP3. Altra caratteristica molto importante e da non sottovalutare è la possibilità di scaricare solo alcune parti del messaggio; questo si rivela fondamentale soprattutto quando la connessione è a bassa larghezza di banda. Oltre a quanto fatto dal POP3, il protocollo IMAP consente inoltre di:

- Rinominare le caselle di posta elettronica:
- Cancellare i singoli messaggi senza essere costretto a recuperarli;
- Leggere le intestazioni dei messaggi senza doverli prelevare interamente;
- Prelevare solamente porzioni di messaggi;

- Supportare della modalità di lavoro offline per i client e successiva sincronizzazione quando possibile;

Una sessione IMAP consiste di una serie di comandi inviati dal client al server, il quale resta in ascolto sulla porta TCP 143. Ogni comando ha un identificatore, cioè un prefisso alfanumerico (es. A0001) chiamato "tag" che lo precede, che ha come compito quello di far corrispondere le risposte del server alle rispettive domande del client. Poiché ha la possibilità di gestire la posta in modo centralizzato sul server, e quindi essendo in grado di accedere da diversi macchine, IMAP implementa un sistema di attributi che possono indicare lo stato corrente di un messaggio.

Vediamo i comandi del protocollo IMAP. Ad ogni messaggio vengono definiti uno o più flag, che iniziano con il carattere "\" e ne indicano lo stato:

- \Seen: il messaggio e' stato letto.
- \Answered: e' stato inviato un messaggio di risposta.
- \Flagged: sul messaggio e' stato impostato un flag urgent o Special Attention.
- \Deleted: il messaggio e' stato contrassegnato per la cancellazione.
- \Draft: il messaggio non e' completo ed e' quindi contrassegnato come bozza.
- \Recent: il messaggio e' appena giunto nella casella di posta.

Una sessione IMAP può avere 5 stati:(si veda la figura 1.8)

- Autenticato
- Non Autenticato
- Selezionato
- LogOut (disconnesso)
- Attesa di IMAP che qualcuno si colleghi

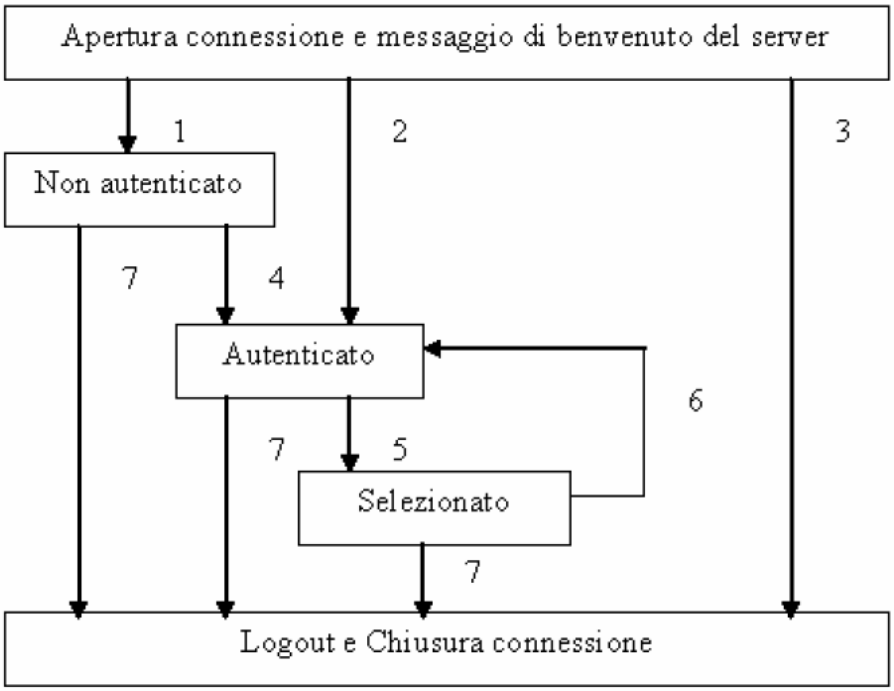

Figura 1.8 Rappresentazione scenari possibili protocollo IMAP

1. Connessione senza preventiva autenticazione
2. Connessione con preventiva autenticazione
3. Connessione respinta
4. Login riuscito
5. Comando Select eseguito con successo oppure Comando Examine
6. Comando Close command o Select fallito oppure Comando Examine
7. Logout e Chiusura Connessione o interruzione del server

Nei vari stati della sessione sono disponibile diversi comandi, che serviranno per l'elaborazione dei messaggi. Nello stato non autenticato, si trovano gli utenti che sono collegati al server ma che ancora si devono autenticare. I comandi utilizzabili in questo stato sono quindi quelli necessari per l'autenticazione:

AUTHENTICATE: indica il metodo di autenticazione da utilizzare;

LOGIN <utente> <password>: invia al server il nome utente e la password per autenticarsi;

Una volta forniti i corretti valori di login e password si passa allo stato Autenticato. E' possibile scegliere, esaminare, creare, cancellare, rinominare una mailbox, controllarne lo stato e gestire le sottoscrizioni ad altre caselle. Questi sono i comandi quando si è autenticati:

SELECT <mailbox>: seleziona la casella di posta con cui lavorare;

EXAMINE <mailbox>: seleziona la casella di posta ma con accesso in sola lettura;

CREATE <mailbox>: crea una nuova casella di posta;

DELETE <mailbox>: cancella una casella di posta;

RENAME <vecchio-nome-mailbox> <nuovo-nome-mailbox>: rinomina una casella di posta;

SUBSCRIBE <mailbox>: aggiunge la casella di posta alla lista delle caselle attive da visualizzare in un client;

UNSUBSCRIBE <mailbox>: elimina la casella di posta alla lista delle caselle attive da visualizzare in un client;

LIST <riferimento> <mailbox>: restituisce un sottoinsieme dei nomi disponibili al client;

LSUB <riferimento> <mailbox>: restituisce un sottoinsieme delle caselle sottoscritte dall'utente;

STATUS <mailbox> <stato>: indica lo stato della casella di posta;

APPEND <mailbox> <messaggio>: aggiunge un messaggio alla casella di posta selezionata.

Si passa allo stato selezionato quando si e' scelto su quale casella lavorare. I comandi per la gestione dei messaggi sono i seguenti:

CHECK: funzione di manutenzione del server IMAP;

CLOSE: riporta nello stato Autenticato e rimuove i messaggi con flag \Delete attivo;

EXPUNGE: rimane nello stato Selezionato e rimuove i messaggi con flag \Delete attivo;

SEARCH [setcaratteri] <parametri>: cerca i messaggi che soddisfano i parametri di ricerca indicati;

FETCH <messaggi> <dati>: visualizza determinati dati di un messaggio (es. soggetto, data, testo ecc);

Una volta che il client effettua una richiesta di chiusura della connessione, si passa nello stato Disconnesso, per terminare il collegamento.

Alcuni comandi non sono associati ad un particolare stato del server e quindi sono sempre disponibili:

CAPABILITY: indica le funzioni supportate dal server;

NOOP: non esegue nulla, risponde in modo positivo se il server è attivo. Utilizzato per evitare la disconnessione causa timeout;

LOGOUT: indica al server che il client vuole terminare la connessione.

1.5 La sicurezza nella posta elettronica con il protocollo S/MIME

Il protocollo SMTP, che gestisce la posta elettronica, è sicuramente uno dei servizi più usati dagli utenti di Internet. Questo servizio viene comunemente considerato sicuro perché quando un utente accede alla casella di posta elettronica, e utilizza il proprio nome utente e password, scrive una e-mail, considera questi dati sicuri e non utilizzabili da nessun altro.

In realtà il protocollo SMTP presenta seri problemi di sicurezza, ad esempio non vi è nessuna forma di autenticazione da parte dell'utente nella fase di invio della e-mail. Altro problema riguarda la riservatezza di una e-mail, poiché questa quando viene spedita al mittente è formata da tanti pacchetti in chiaro, che all'interno della rete potrebbero essere letti da terzi.

Per risolvere questa serie di problematiche, come riservatezza e autenticazione, è stato creato il protocollo S/MIME. La prima versione di questo protocollo apparve nel 1995, nel 1998 uscì una seconda versione, che venne sottoposta a studio dalla IEFT, per poter essere considerato standard di sicurezza nel servizio della posta elettronica. Nel 1999 si ebbe la conferma a livello mondiale per la protezione delle e-mail.

1.5.1 Il protocollo S/MIME e la firma digitale

Il protocollo S/MIME ha due modalità per risolvere il problema sicurezza della posta elettronica: la firma digitale e la crittografia della e-mail.

La firma digitale soddisfa tre esigenze che sono autenticità, cioè il destinatario riesca a verificare l'identità del mittente; non ripudio, dove il mittente non potrà rinnegare un documento da lui firmato; integrità dei dati, il destinatario non potrà modificare il documento. Il non ripudio fa riferimento all'art. 21 del decreto legislativo 82/2002, valore probatorio del documento informatico sottoscritto che rimanda anche all'art. 2702 del codice civile.[6]

La firma digitale è composta da tre algoritmi:

- algoritmo per generare la chiave G che produce due chiavi PK e SK. PK (public key) che è la chiave pubblica per la verifica della firma, invece SK (secret key) chiave privata che è in possesso del firmatario, che la utilizza per firmare la e-mail.
- algoritmo di firma S, questo prende in input il messaggio di posta elettronica M, e la chiave privata SK e ne produce una firma F.
- algoritmo per la verifica della firma V, questo prende in input il messaggio di posta elettronica M e la chiave pubblica PK e la firma F, e ne verifica la validità.

Vediamo come viene applicata la firma digitale a una e-mail. Fondamentali gli algoritmi per porre la firma digitale e per verificarla, questi sono: l'algoritmo di firma e l'algoritmo di verifica. Il primo, crea la firma digitale che dipenderà dall'argomento della e-mail a cui deve essere inserita la firma, e dalla chiave che possiede il mittente. Il secondo, algoritmo di verifica, viene utilizzato dal destinatario per decretare l'autenticità della firma.

Il mittente, per porre la firma digitale utilizzerà delle informazioni (password), e grazie all'algoritmo di Hash, ottiene una stringa funzione del documento; la stringa sarà a sua volta cifrata con la chiave privata del mittente. Dalla cifratura della stringa si ottiene la firma digitale, che vine inserita nella e-mail.

Il destinatario utilizza l'algoritmo di Hash per calcolare la stringa hash, poi decritta la firma digitale con la chiave pubblica del mittente, così da avere la stringa hash calcolata dal mittente.In questo

modo potranno mettere a confronto le due stringhe hash calcolate, per decretare l'identità del mittente e l'integrità dei dati della e-mail.

Piccolo accenno sull'algoritmo di Hash, questo quando viene utilizzato restituisce delle stringhe di lettere e numeri; viene utilizzato per la firma digitale poiché la stringa di output è univoca per la e-mail a cui viene applicato l'algoritmo; inoltre non è invertibile, quindi non sarà possibile partendo dalla stringa di output ricostruire la e-mail o il documento originale.

Applicazione della firma digitale ad una e-mail dal lato mittente.

1. Acquisizione e-mail.
2. Recupero da parte del mittente delle informazioni univoche (ad esempio password, chiave privata) utili per identificarlo.
3. Applicazione della la firma digitale fornita dalle informazioni univoche del mittente.
4. Inserimento della firma digitale nella e-mail.
5. Invio e-mail.

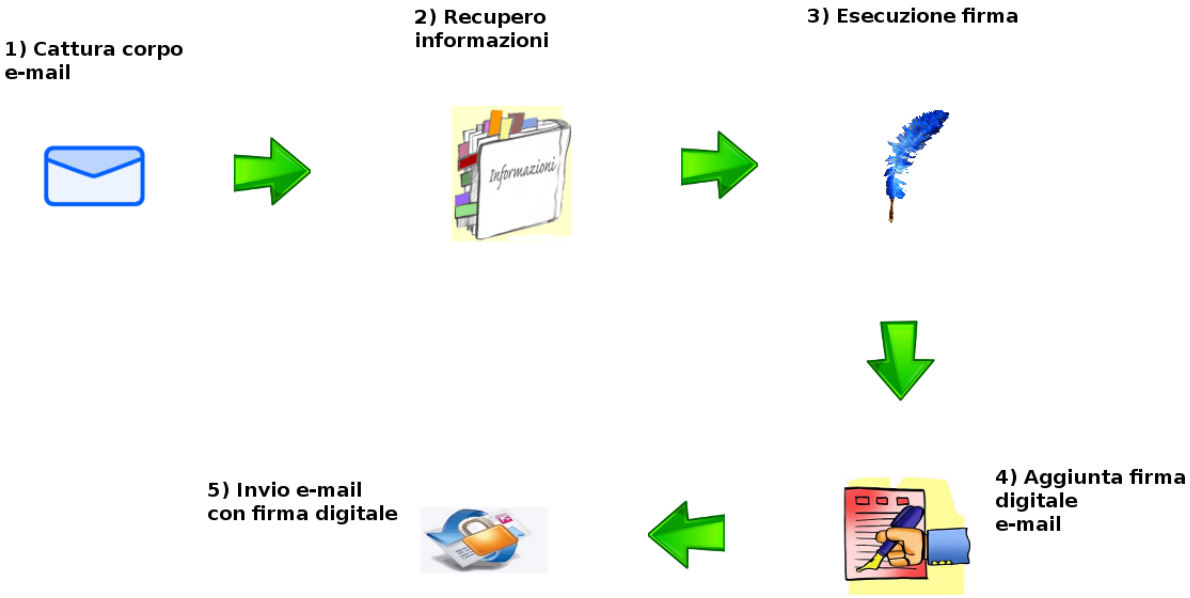

Figura 1.9 Applicazione della firma digitale ad una e-mail dal lato mittente.

Verifica firma digitale da parte del destinatario

1. Ricezione e-mail.
2. Recupero della firma digitale presente nella e-mail.
3. Recupero corpo e-mail.
4. Recupero informazioni univoche (chiave pubblica) del mittente.
5. Creazione firma da parte del destinatario.
6. Confronto delle due firme (mittente e destinatario).
7. Se le due firme digitali sono corrispondenti la e-mail sarà valida.

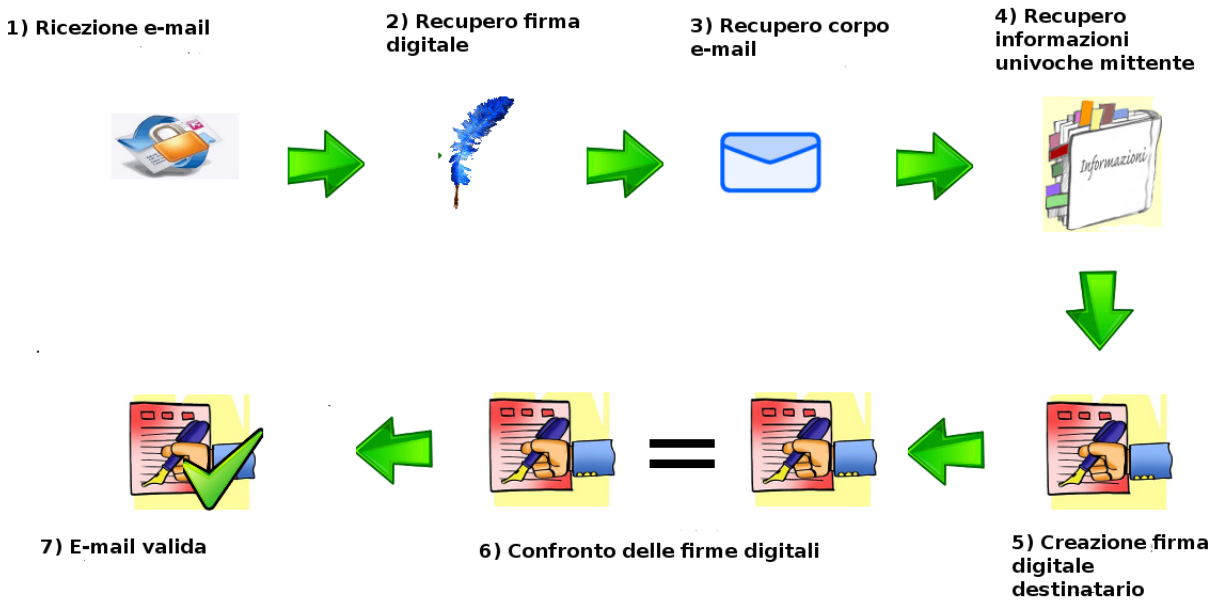

Figura 1.10 Verifica firma digitale da parte del destinatario.

1.5.2 Il protocollo S/MIME e la crittografia

Altra importante soluzione proposta dal protocollo S/MIME è la crittografia. Questa tecnica non è presente nel protocollo SMTP, infatti quando una e-mail viene spedita, prima di raggiungere il destinatario, corre il rischio che possa essere letta da terzi.

Con la crittografia viene meno questo problema, poiché una e-mail viene modificata, così da rendere vano il tentativo di lettura di utenti diversi dal destinatario della e-mail. Caratteristica fondamentale della crittografia sono: la riservatezza della e-mail, e l'integrità dei dati. Riservatezza poiché la e-mail crittografata sarà leggibile solo al destinatario; integrità dei dati poiché una e-mail crittografata, non sarà modificabile.

La crittografia, detta anche crittografia a chiave pubblica-privata, associa ai due utenti coinvolti una coppia di chiavi.

La chiave privata personale dell'utente che lo utilizza per decodificare la e-mail cifrata. La chiave pubblica, serve a cifrare la e-mail del destinatario che possiede la relativa chiave privata.

Infatti la crittografia che utilizza il sistema chiave pubblica-privata, ha una caratteristica molto importante; una e-mail cifrata potrà essere decifrata usando solo la chiave privata corrispondente. é come se avessimo una cassaforte con due chiavi distinte, una per aprirla e l'altra per chiuderla.

La coppia di chiavi viene generata da un algoritmo (RSA) a partire da numeri casuali. Fondamentale è che essendo un algoritmo asimmetrico, non permette a chi possiede la chiave pubblica di risalire alla chiave privata. Infatti un esempio riassuntivo potrà così essere descritto: se un utente X vuole spedire una e-mail crittografata, a un utente Y, l'utente X dovrà avere la chiave pubblica di Y per cifrare la e-mail; mentre l'utente Y per decodificare il messaggio utilizzerà la propria chiave privata.

Riassumiamo come viene effettuata la crittografia dal mittente:

1. Recupero corpo e-mail.
2. Recupero informazioni univoche destinatario (ad esempio chiave pubblica).
3. Esecuzione crittografia sulle informazioni fornite dal destinatario.
4. Crittografia della e-mail originale.
5. Invio e-mail

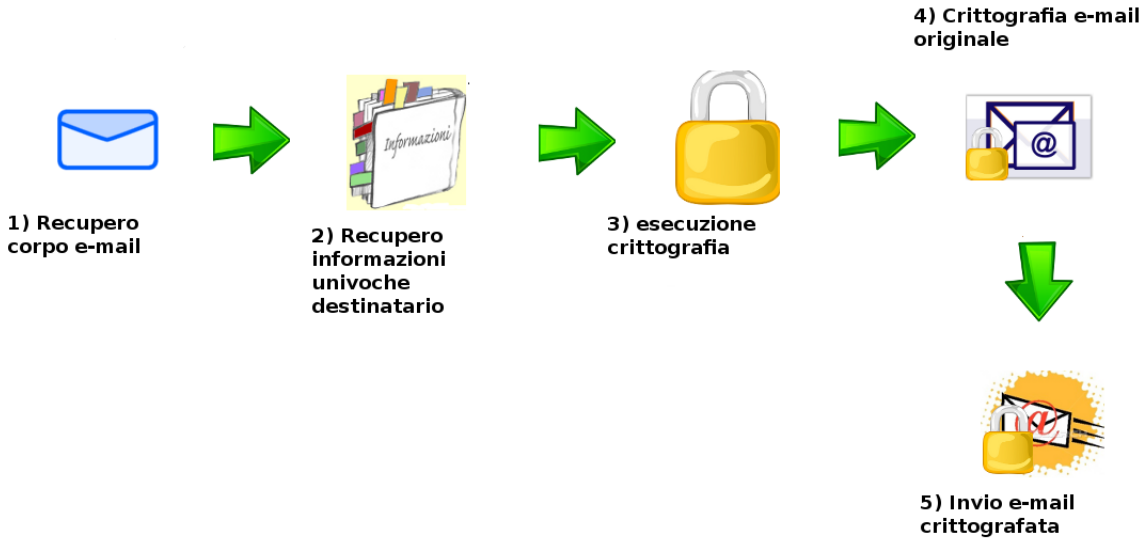

Figura 1.11 Applicazione crittografia alla e-mail

Riassumiamo come viene effettuata la de crittografia dal destinatario:

1. Ricezione e-mail.
2. Recupero e-mail crittografata
3. Recupero informazioni univoche del destinatario (ad esempio chiave privata)
4. Esecuzione de crittografia della e-mail sulla base delle informazioni univoche
5. Recupero e-mail con testo originale non crittografato.

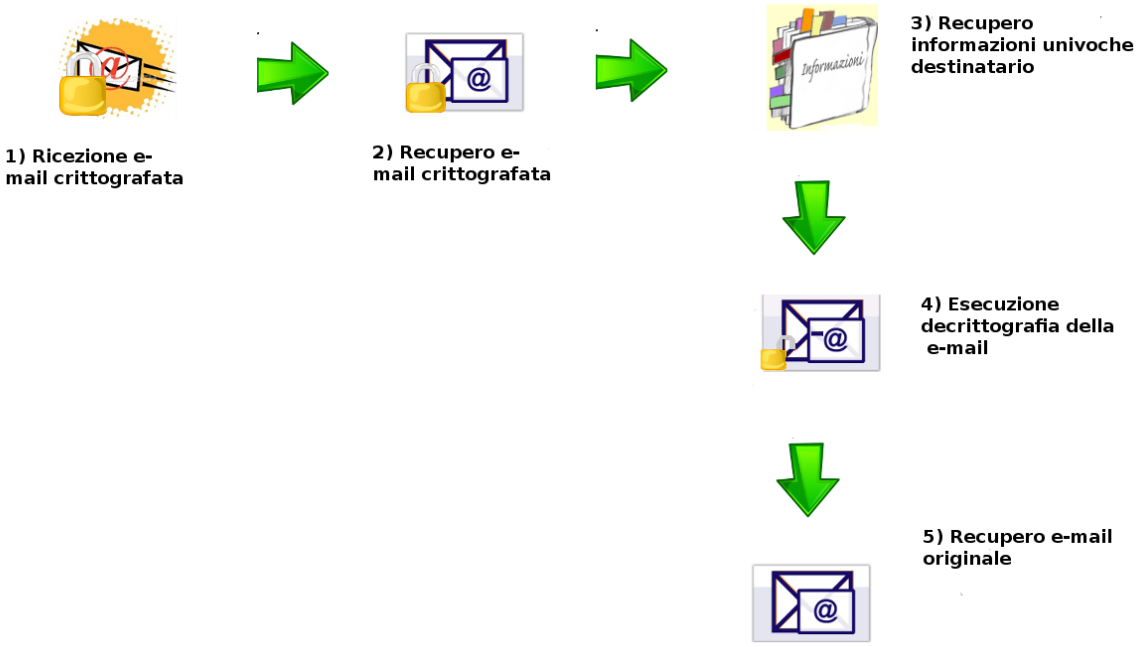

Figura 1.12 Applicazione de crittografia alla e-mail.

2 Il problema della posta indesiderata (Spam)

2.1 Cos'è lo Spam

Da quando Internet è diventato di uso abituale, ha iniziato a circolare una grande quantità di informazioni: questo ha fatto sì che molte persone se ne siano servite abusivamente. Comunemente viene detto "Spam" o "Posta indesiderata" l'invio di posta elettronica (solitamente messaggi pubblicitari) a una grande quantità di destinatari senza che questi acconsentano al servizio. Il nome Spam ha origine da uno sketch comico nel quale la cameriera di un locale proponeva continuamente ai clienti un menù tutto a base di questo "Spam", che altro non era che la pubblicità ossessiva di una carne in scatola. Lo scopo principale dello Spam infatti è fare pubblicità in maniera massiccia e ripetitiva e a basso costo. Gli indirizzi dei destinatari vengono raccolti dagli Spammer su Internet con diverse modalità: grazie a dei software nascosti nelle pagine web o installati dagli utenti che recuperano indirizzi da forum o dai vari siti Internet.

2.2 Tipi di posta indesiderata e i suoi effetti

Fra le tipologie di Spam più conosciute vi sono:

- Posta non sollecitata a molti recipienti (UBE, Unsolicited Bulk Email): sono chiamati "Urban legend" o "Hoax" quei messaggi che non hanno contenuto commerciale, le cosiddette "catene di Sant'Antonio", e-mail che arrivano da persone sconosciute a cui non è stato dato il permesso di inviarle. Vengono facilmente riconosciute grazie al loro contenuto, incentrato nell'attirare l'interesse, desiderio di sapere o comportare lo choc emotivo nei confronti del lettore per indurlo a realizzare le richieste presentate, di solito impostate nel divulgare l'e-mail a quante più persone possibili. In queste situazioni è sempre meglio analizzarne il contenuto, facendo delle dovute ricerche in Internet, per identificare eventuali elementi che diano verità o smentita al messaggio recapitato. Se le ricerche effettuate ci indicano la presenza di un "Hoax" o di una "Urban Legend", è bene eliminare la mail ricevuta.

- Posta commerciale non sollecitata (UCE, Unsolicited Commercial Email): come dice il nome stesso, queste e-mail pubblicizzano prodotti commerciali, inoltrate chiaramente senza il consenso del destinatari.
- Mail derivanti da iscrizioni a mailing-list: la mailing-list, traducibile in italiano come lista di distribuzione, non è altro che una lista di indirizzi di posta elettronica, dove una e-mail viene spedita automaticamente a una serie di destinatari appassionati a un particolare argomento. Può creare dei problemi di posta indesiderata poiché tutti possono iscriversi e solitamente non richiede la conferma diretta all'utente, quindi sarà facile trovarsi iscritti senza saperlo. Fortunatamente questo fenomeno di posta indesiderata sta scomparendo.

Inviare dello Spam è molto semplice e anche molto economico: basta infatti possedere un semplice collegamento a Internet e un PC; in questo modo uno Spammer riesce ad inviare messaggi di posta elettronica a basso costo ma con riscontri in termini economici molto elevati, dovuti alla semplicità con cui si inviano e-mail pubblicitarie, che ne sponsorizzano il prodotto o quel che sia.

In questo modo i costi aumentano per chi, ad esempio, usa una connessione a pagamento, a minuti o a traffico scaricato: infatti l'utente che riceve Spam si troverà delle e-mail indesiderate che andranno a sottrargli del tempo, perché dovrà eliminare lo Spam dalla propria casella di posta elettronica, liberarne dello spazio occupato da posta indesiderata..

Infatti ora molti ISP implementano dei filtri all'interno dei server di posta elettronica, cercando di diminuire lo Spam presente nel servizio offerto. All'interno di queste liste di filtri vi sono gli indirizzi IP e i domini riconosciuti come colpevoli di invio di Spam. Infatti in alcune situazioni, l'invio di posta elettronica indesiderata contemporaneamente a molti utenti potrebbe creare un grosso problema ai server che ospitano le e-mail. Questo avviene quando molti utenti accederanno all'interno della propria casella di posta elettronica, dove il server di posta sarà rallentato a causa dell'arrivo di molti messaggi di posta inviati dagli Spammer. Tutto ciò crea danni nei confronti del servizio della posta elettronica, ma anche a tutta la rete stessa: può verificarsi, una perdita di credibilità nel funzionamento

della posta elettronica da parte di utenti che la vorrebbero utilizzare, ad esempio per affari, o creare danni al commercio elettronico, poiché gli utenti potrebbero non effettuare acquisti all'interno della rete per paura di ricevere e-mail indesiderate.

Riassumendo quando detto, lo Spam comporta dei rischi e inconvenienti quali:

- la perdita di tempo provocata dalla cancellazione dei messaggi;
- il contenuto spesso violento o indecoroso dei testi o delle immagini di questi messaggi, che talvolta potrebbero urtare la sensibilità dei destinatari;
- lo spreco della banda all'interno della rete;
- l'occupazione eccessiva di spazio all'interno delle casella di posta elettronica;
- la difficoltà nel consultare la propria e-mail a causa dei molti messaggi pubblicitari ricevuti, e quindi il rischio di cancellare messaggi a noi utili al posto dello Spam.

La ricezione dello Spam comporta anche dei costi aggiuntivi nei confronti dei fornitori dei servizi internet, e di conseguenza una maggiorazione del prezzo degli abbonamenti. Questi costi aggiuntivi sono dovuti a:

- attivazione di sistemi anti-spam;
- consumo di risorse supplementari, come filtri nati-Spal all'interno dei server.

2.3 Spam "accettato", Opt-Out, Single Opt-in, Notified Opt-In, Double Opt-in

Lo Spam è diventato talmente diffuso all'interno della rete che si sono sviluppati diversi metodi di affrontare il problema da parte dei vari ordini giuridici. Infatti per le newsletter e le e-mail marketing sono stati adottati degli strumenti di prevenzione, al fine di non ricevere lo Spam, utili anche per la regolamentazione dello norme sulla privacy. Questi strumenti sono : Opt-Out [7], Single Opt-in [7], Notified Opt-In [7], Double Opt-in [7].

- Opt-Out: è il caso in cui la ricezione delle e-mail marketing considerate Spam arrivano al destinatario, fino al momento in cui quest'ultimo non decide in modo esplicito di non volere più ricevere posta indesiderata. Se il destinatario non dovesse effettuare la notifica ai possessori delle liste di indirizzi di posta elettronica, le e-mail marketing e newsletter continueranno ad essergli inviate.
- Single Opt-In è la modalità di ricezione delle e-mail marketing e newsletter che avviene in seguito all'esplicito consenso dell'utente a riceverle e quindi anche ad essere aggiunto alle liste di indirizzi. La comunicazione del consenso avviene tramite l'invio di una e-mail da parte dell'utente, dopo che quest'ultimo ha compilato un form di registrazione. Sicuramente le e-mail ricevute tramite l'Opt-Out sono maggiori rispetto al Opt-In, ma anche quest'ultimo sistema presenta dei limiti, nonostante abbia maggiore sicurezza rispetto all' Opt-Out: il principale limite è dato dal fatto che in seguito a errori nel formulare il form (ad esempio indirizzo di posta elettronica scritto in maniera errata) potrebbero essere inseriti indirizzi di posta elettronica non corrispondenti a chi a compilato il form; altro non secondario problema potrebbe essere quello di inserire indirizzi di posta elettronica di terze persone senza averne ricevuto il consenso dal possessore, poiché il Single Opt-In non prevede la verifica delle e-mail da parte del destinatario.
- Notified Opt-In: questo metodo potrebbe sembrare simile al Single Opt-In, ma fra i due vi è una sostanziale differenza.

Poiché nel Notified Opt-In, quando un utente si iscrive a delle newsletter o a delle e-mail marketing , quindi a liste di indirizzi, riceve una e-mail in cui può decidere di cancellarsi da queste liste, così da non ricevere e-mail indesiderate.

- Double Opt-In quest'ultimo modello è sicuramente il migliore per quando riguarda la sicurezza nella ricezione di e-mail marketing e newsletter. Infatti viene chiamato anche "confirmed Opt-In", poiché rispetto al Single Opt-In, al Opt-In e al Notified Opt-In è quello che non solo richiede l'iscrizione, ma un'ulteriore conferma avviene attraverso l'invio di una e-mail successiva da parte del destinatario. Quindi, così facendo si sarà sicuri che il consenso sarà del vero utente possessore dell'indirizzo di posta elettronica, poiché questi avrà effettuato l'accesso alla propria casella di posta elettronica per poter rispondere e quindi essere consapevole di ricevere delle e-mail marketing o newsletter.

2.4 Tipologie di Spam

Lo Spam all'interno della rete ha avuto una enorme crescita, infatti consultando i dati diffusi da Brightmail[8], vediamo le percentuali di posta indesiderate presenti nel web dal 2001 all'ultima pubblicazione fornita a settembre 2010. Nel 2001 la percentuale era molto bassa, infatti era ferma a 1%: solo una e-mail ogni cento era considerata Spam. Nel 2003 il fenomeno della posta indesiderata ha avuto una enorme crescita, infatti i dati iniziarono a essere preoccupanti, poiché si era sul 49% di Spam presente nelle caselle di posta elettronica. Nel 2004 le cose non sono sicuramente migliorata: si è arrivati persino al 65% di Spam presente nelle caselle di posta elettronica. Con l'ultimo rapporto della Brightmail, aggiornato a settembre 2010, lo Spam presente all'interno della rete risulta essere arrivato al 92.51%. Come si può notare la presenza di posta indesiderata è cresciuta in maniera consistente rispetto ai primi rapporti del 2001; potremmo anzi aggiungere che quasi l'intera posta elettronica presente nella rete è Spam.

Le categorie di posta indesiderata presenti nella rete possono essere così classificate: Internet 45%, prodotti 14%, salute 12%, finanza 9%, truffe 4%, raggiri 4%, tempo libero 4%, adulti 1%, politica <1%, altri 8%.

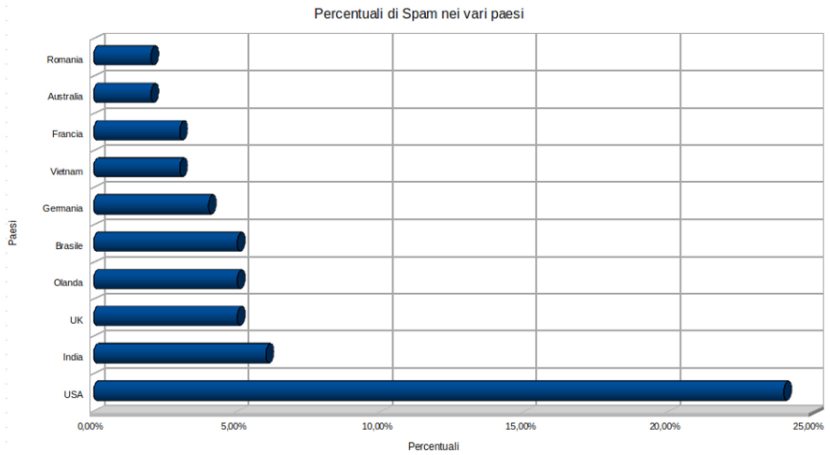

Figura 2.1 Categorie di Spam

Gli URL da cui provengono invece queste e-mail considerate Spam sono: .COM 67%, .RU13%, .ORG 6,2%, .info 4,8%.

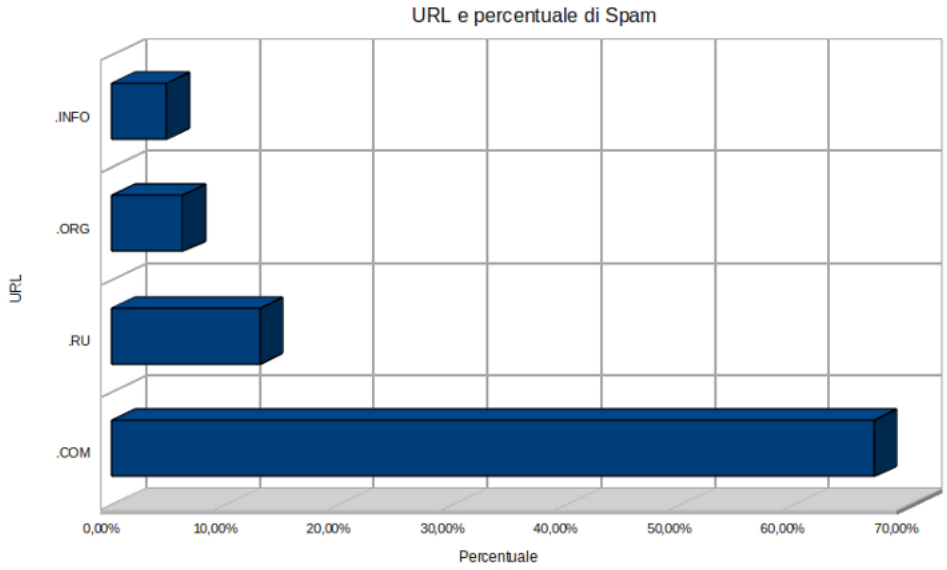

Figura 2.2 URL dello Spam

Per quanto riguarda i principali paesi da cui ha origine lo Spam, possono essere così suddivisi: USA 24%, India 6%, UK 5%, Olanda 5%, Brasile 5%, Germania 4%, Vietnam 4%, Francia 3%, Australia 2%, Romania 2%.

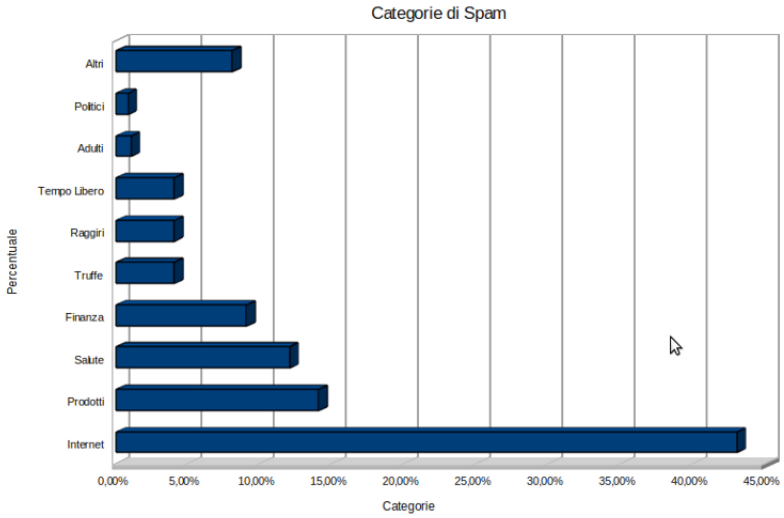

Figura 2.3 Categorie di Spam

Inoltre, secondo un rapporto di Enisa pubblicato a gennaio 2010 dal Sole 24 ore, risultano elevati anche i costi che devono affrontare i provider di posta elettronica: nei paesi più grandi i costi sopportati dai provider sono pari circa a un milione di euro l'anno, invece nei paesi più piccoli siamo sull'ordine dei diecimila euro l'anno.

3 Aspetti Giuridici

Il fenomeno dello Spam, come invio di messaggi di posta indesiderata, ha assunto un'importanza rilevante nel mondo della rete. Sicuramente fra le cause principali, che hanno portato lo Spam a divulgarsi così prepotentemente, vi è la struttura stessa di Internet. Questa, insieme all'evoluzione tecnologica, ha dato la possibilità agli Spammer di migliorare la tecniche di invio di Spam, andando ad eludere i controlli messi in atto da parte degli Internet Service Provider.

Di fronte al problema dello Spam, va detto che sia l'Italia che gli Stati Membri dell'unione Europea hanno attuato una serie di normative dirette a ridurre e prevenire il fenomeno della posta indesiderata. Queste regolamentazioni sono rivolte a tutti i soggetti, come ad esempio call center, imprese,ecc. che interagiscono ogni giorno con i dati sensibili degli utenti. Vediamo la legislazione italiana come interviene nei confronti dello Spam e di chi lo attua.

3.1 La normativa di riferimento

La legge italiana, per regolamentare e disciplinare il fenomeno dello Spam, ha messo in pratica normative a cui fare riferimento. Per quando riguarda l'invio di posta elettronica a fini commerciali, la legge italiana fa riferimento all'art. 130 Codice Privacy, rubricato come "Comunicazioni Indesiderate". Questo articolo regolamenta il fenomeno dello Spam, anche se limitato alla sola comunicazione indesiderata e non a quella non richiesta. Per regolamentare la comunicazione indesiderata, sia l'Italia che gli altri Stati della Comunità Europea hanno deciso di adottare l'Opt-In (vedere capitolo 2, paragrafo 2.3). Questo sistema avverte le imprese, le società e qualsiasi altro soggetto abbia a che fare con il trattamento dei dati personali di un utente, che si ha la possibilità di utilizzare questi dati sensibili solo dopo che l'utente stesso abbia dato esplicito consenso.

Inoltre, all'interno dell'art. 130 Codice Privacy sono presenti altri divieti, riguardanti l'invio di messaggi a scopo pubblicitario, messaggi per la vendita diretta o per indagini di mercato, per i quali l'impresa o l'attività che invia tali messaggi dovrà obbligatoriamente fornire un recapito al quale l'utente possa rivolgersi per esercitare i propri diritti.

Sempre all'interno dell'art. 130 Codice Privacy, capo I Sistemi informativi, comma 6 dice che: "In caso di reiterata violazione delle disposizioni di cui al presente articolo il Garante può, prevedendo ai sensi dell'articolo 143 comma 1, lettere b, altre sì prescrivere ai fornitori di servizio (Internet Service Provider) di comunicazione elettronica di adottare procedure di filtraggio o altre misure praticabili relativamente alle coordinate di posta elettronica da cui sono state inviate le comunicazioni" [9].

Invece, art. 143 comma 1, lettera b dice che:"Esaurita l'istruttoria preliminare, se il reclamo non è manifestatamente infondato e sussistono i presupposti per adottare un provvedimento, il Garante, anche prima della definizione del procedimento: prescrive al titolare le misure opportune o necessarie per rendere il trattamento conforme alle disposizioni vigenti"(lettera b) [10].

Con questo articolo si vuole indicare che il Garante dovrà comunicare ai fornitori di tale servizio (ISP), solitamente Internet Service Provider, che devono adottare misure di filtraggio/prevenzione utili alla protezione dell'indirizzo di posta elettronica dell'utente, nel momento in cui vengono sottratte informazioni personali, ad esempio indirizzi di posta elettronica, da parte di aziende o chi per loro (Spammer).

La legislazione prevede un altro ordinamento, il decreto legislativo del 9 Aprile 2003 n.70 [11]. Questo riferisce sul commercio elettronico, in particolare l'art. 9 di tale decreto asserisce che: "Le comunicazioni commerciali non sollecitate trasmesse da un prestatore di posta elettronica devono, in modo chiaro ed inequivocabile, essere identificate come tali fin dal momento in cui il destinatario le riceve e contenere le indicazioni che il destinatario del messaggio può opporsi al ricevimento futuro di tali comunicazioni".

Anche in questo caso risulta chiaro che quando un utente riceve delle e-mail pubblicitarie o commerciali, all'interno di tale messaggio vi devono essere tutte quelle informazioni utili per consentire al destinatario del messaggio (ad esempio l'indirizzo del mittente), di informare il mittente di non volerne più ricevere.

Il primo articolo che impone una pena a chi è fonte di Spam, è l'art. 167 Codice Privacy che asserisce: "Salvo che il fatto costituisca più grave reato, chiunque al fine di trarne per se o per altri profitto o di recare ad altri un danno, precede al trattamento di dati personali in violazione in quanto previsto dal Codice stesso, è punito, se dal fatto deriva nocumento, con la reclusione da sei a diciotto mesi o, se il fatto consiste nella comunicazione e diffusione, con la reclusione da sei a ventiquattro mesi" [12].

Inoltre l'art. 161 Codice Privacy, impone sanzioni a chi fa Spamming:"La violazione delle disposizioni dell'articolo 13 (informativa dove all'interessato dei dati, deve essere data comunicazione) è punita con la sanzione amministrativa del pagamento di una somma da tremila euro a diciottomila euro, nei casi di dati sensibili o giudiziari o di trattamenti che presentano rischi specifici all'articolo 17 (dove sono presentati i rischi dovuti al trattamento dei dati sensibili) o, comunque, di maggiore rilevanza del proprietario per uno o più interessati, da cinquemila euro a tremila euro. La somma può essere aumentata sino al triplo quando risulta inefficace delle condizioni economiche del contravventore" [13].

Come risulta da questi articoli, vi possono essere due tipi di sanzioni, applicabili a chi sia considerato fonte di Spam, a chi utilizzano i dati sensibili di un utente senza previo avviso o chi tratta tali dati senza considerare i rischi possibili e infine chi trarrà beneficio da tali dati. Nell'articolo 167 è punibile con la reclusione chi invia posta indesiderata, e ne trae beneficio economico. Invece nell'art. 161, chi è causa di Spam è punibile con sanzioni amministrative molto elevate, che vanno dai 300 euro fino a 500 euro.

In questi casi rientrano tutti gli Spammer o chi li commissiona, aziende o organizzazioni. Quindi tutti gli utilizzatori di Web Bugs, di Spyware, di Spambot sono sanzionabili, poiché recuperano e utilizzano in maniera illegittima i dati sensibili dell'utente.

Uno dei primi casi riguardanti l'attività di Spamming portata in giudizio è avvenuto a Napoli il 26 Giugno 2004: è stato infatti condannato il primo Spammer in Italia, con il risarcimento nei confronti del soggetto danneggiato. In questo processo le cause del risarcimento riguardavano lo stress subito dall'utente titolare della casella di posta elettronica, dovuto alla ricezione dello Spam.

Per quando riguarda la comunicazione via e-mail relativa all vendita di prodotti o servizi, si è attuato un sistema diverso: per la

comunicazione pubblicitaria in Italia si utilizza il sistema dell'Opt-In, per quella che commerciale invece si fa riferimento al decreto legislativo 206 del 2005, in particolare all'art. 58, comma 1 e 2 che asseriscono: "L'impiego da parte di un professionista del telefono, della posta elettronica, di sistemi automatizzati di chiamata senza l'intervento di un operatore o di fax richiede il consenso preventivo del consumatore (1). Tecniche di comunicazione a distanza diverse da quelle di cui al comma 1, qualora consentano una comunicazione individuale, possono essere impiegate dal professionista se il consumatore non si dichiara esplicitamente contrario(2)" [14].

4 Le tecniche usate dagli Spammer per il recupero di indirizzi

I costi che affronta uno Spammer per portare avanti la sua attività e per recuperare gli indirizzi di posta elettronica sono decisamente ridotti; chi in realtà subisce tali costi è il destinatario della posta elettronica indesiderata. Lo Spammer, con l'invio di e-mail spazzatura, riesce ad effettuare pubblicità in tutto il mondo a costi bassi. Diverse sono le tecniche utilizzate dagli Spammer per il recupero di indirizzi di posta elettronica. Fra questi vi sono:

- WEB Bug
- Dictionary Attack
- Spambot
- Spyware

4.1 Dictionary Attack

E' una tecnica utilizzata per indovinare gli indirizzi di posta elettronica; più precisamente lo Spammer cerca di generare indirizzi che potrebbero realmente esistere. Solitamente un indirizzo di posta elettronico è così formato: nome.cognome o username@dominio. Per la parte sinistra dell'indirizzo di posta elettronica, ovvero la parte dello user name, gli Spammer compongono stringhe in base a qualche logica, per lo più nomi di persone. Invece per la parte alla destra della "@", gli Spammer usano domini validi. Questo è dei motivi principali per cui l'indirizzo nome.cognome@dominio.it è uno dei più soggetti a questo tipo di attacco.

4.2 Spambot

Gli Spambot [15] sono particolari software usati dagli Spammer per analizzare il web e ricercare indirizzi di posta elettronica. Lo Spambot fa parte della categoria di software chiamati Web Crawler. Questo è un programma che analizza i contenuti presenti nella rete, compresi i database, in maniera totalmente automatizzata. Una volta

raccolte queste informazioni le inserisce in un indice, che fornirà al motore di ricerca a cui fa riferimento lo Spambot. Fra le principali fonti di indirizzi si trovano i sistemi di mailing-list, il web e i vari gruppi di discussione che si basano su indirizzi di posta elettronica. Alle mailing-list si accede per richiedere l'iscrizione del proprio indirizzo, così da riceverne notizie riguardanti quella mailing-list. Per evitare l'iscrizione di indirizzi non consentita è stata inserita la conferma di iscrizione, quindi tutte le mailing-list che non prevedono la verifica possono essere più facilmente vittime di attacchi da parte degli Spammer. Per quanto riguarda i newsgroups il campo più a rischio risulta essere il campo "From:". Ne esistono di molte tipologie di Spambot, possono anche essere realizzati in casa; vi sono persino aziende che li creano e li rivendono anche a prezzi molto elevati. Fra gli Spambot più importanti troviamo Web Data Extractor che, oltre ad estrarre indirizzi e-mail, estrae numeri di fax, numeri di telefono, URL, testo e tag particolari.

4.3 Spyware

Gli Spyware [16] sono software che raccolgono informazioni sugli utenti presenti sulla rete senza il loro consenso. Fra le informazioni prelevate vi sono siti visitati, password, o indirizzi e-mail. Queste saranno necessarie agli Spammer o aziende che tramite l'uso di Spyware cercano di raccogliere le informazioni dell'utente, così potranno effettuare ricerche, inviare messaggi pubblicitari (non richiesti) e quindi trarne profitto pubblicizzando i loro prodotti.

Gli Spyware, a differenza dei Web Bugs, dei Dictionary Attack e Spambot, non sono capaci di diffondersi in automatico, ma hanno bisogno che l'utente li installi. Un esempio è quello dei vari programmi presenti e distribuiti sulla rete in maniera gratuita. In realtà molti di questi programmi non sono totalmente "gratuiti", poiché nascondono uno stratagemma al loro interno per essere "ripagati". Infatti quando l'utente andrà ad installarli, gli Spyware raccoglieranno moltissime informazioni su quell'utente. Molte volte questi programmi chiedono agli utenti se vogliano tenerli come pagina principale sul proprio browser, o effettuare delle ricerche sul web con un loro motore di ricerca. Le informazioni raccolte sono, ad esempio, i gusti di navigazione dell'utente, a cosa è interessato, indirizzi e-mail, il tempo medio di navigazione e nei casi peggiori anche la password

della casella di posta. Grazie a tutte queste informazioni, gli Spammer o le aziende, si ripagheranno il software gratuito installato sul PC dell'utente, e così con queste informazioni raccolte, gli Spammer possono inviare dello e-mail indesiderate e pubblicità mirate all'interno delle casella di posta elettronica, a seconda delle abitudini di navigazione dell'utente.

4.4 Web Bugs

I Web Bugs, sono tra i metodi più subdoli per il recupero di informazioni che riguardano l'utente (ad esempio browser utilizzato, indirizzo di posta elettronica, indirizzo IP, orario di navigazione ecc.). Purtroppo il rischio di incappare in un Web Bugs non è risolvibile utilizzando un browser al posto di un altro.

I Web Bugs sono elementi grafici, immagini piccolissime spesso della dimensione di un pixel, sono anche chiamati e-mail tracking, questi sono utilizzati dagli Spammer per verificare che il destinatario abbia letto le e-mail e sopratutto che l'indirizzo di posta elettronico sia valido.

Anche all'interno della propria casella di posta elettronica si possono trovare dei Web Bugs, infatti una e-mail, che potrebbe sembrare solo testo, in realtà è anche formata da linguaggio HTML; quindi anche nelle e-mail vi potrebbero essere immagini che in realtà nascondono dei Web Bugs.

Molti client di posta elettronica, onde evitare gli inconvenienti sopra citati, mettono in allarme l'utente quando riceve delle e-mail, ad esempio bloccando le immagini presenti nelle e-mail, e solo nel caso in cui l'utente sarà sicuro della provenienza della e-mail ricevuta, allora potrà sbloccarle.

In conclusione i Web Bugs cercano di recuperare informazioni di navigazione dell'utente, andando così a violare la privacy. Importante sarà il comportamento dell'utente quando naviga sulla rete o quando andrà a leggere le e-mail ricevute; dovrà stare bene attento alla provenienza dei messaggi di posta elettronica ricevuti e quindi presenti nella propria casella di posta, evitando così di imbattersi con i Web Bugs.

5 Strumenti utili per limitare lo Spam

Vediamo in questo capitolo come cercare di eliminare lo Spam sia a livello di server di posta elettronica, con le liste di blocco (Black List), sia a livello utente con vari software disponibili (Spam Terminator, Mail Washer, Spamhilator).

5.1 Le liste di blocco (Black List)

Dai dati forniti da parte di alcune aziende (come ad esempio Brightmail), si è visto come ormai il problema della posta elettronica indesiderata a livello mondiale sia molto cresciuto, tanto da far nascere una "lotta antispam" da parte degli utilizzatori della rete Internet. Alcuni strumenti molto utili su cui andremo a soffermarci sono le liste di blocco, chiamate anche Black List [17].

Le liste di blocco sono uno strumento non disponibile a livello client di posta elettronica, ma utilizzate a livello server. Il compito principale delle liste di blocco è quello di bloccare lo Spam prima che possa arrivare all'utente finale e quindi all'interno della sua casella di posta elettronica.

Le liste di blocco sono composte da indirizzi IP, che vengono selezionati, e quindi considerati causa di Spam, quando risultano violare certi standard, che sono fondamentali all'interno della rete affinché non vi siano problemi (violazione della privacy, casella di posta elettronica piena di Spam) fra quelli presi in considerazione vi possono essere: norme di comportamento (violazione della privacy), standard di sicurezza (proxy o relay aperti). Le liste di blocco vengono conservate nel tempo, e sono soggette a cambiamenti e aggiornamenti, ad esempio aggiungendo o eliminando indirizzi IP, dichiarati fonte di Spam; inoltre sono consultabili e in alcuni casi aggiornabili da chiunque lo ritenga opportuno.

In alcuni casi però vi possono essere liste di indirizzi IP non del tutto affidabili, chiamate "Bastard Admin". Queste sono create non seguendo i citati standard per determinare se un indirizzo IP è fonte di Spam, ma seguendo criteri del tutto arbitrari. Quindi per chi volesse verificare alcune e-mail siano considerate Spam tramite queste liste di blocco, dovrà considerare che le liste rispettino certi standard che assicurino la validità di ciò che si sta consultando.

Fra questi standard e requisiti che determinano la validità delle liste di blocco vi sono: la possibilità di controllare documentazioni che giustificano il perché quell'indirizzo IP è presente nella lista, oppure perché alcuni indirizzi IP vengano prima inseriti e poi rimossi, e infine che l'autore delle liste di blocco abbia ben chiaro quali siano gli standard utilizzabili per considerare un indirizzo IP fonte di Spam.

Diverse sono le strategie per bloccare la posta indesiderata. Vediamo alcune delle liste di blocco e come sono state catalogate.

- Liste di indirizzi o blocchi lasciati aperti a disposizione di Spammer

All'interno di questa categoria di liste di blocco sono presenti gli indirizzi IP che appartengono a mail server o web server che fanno parte di provider o fornitori di hosting. La denuncia effettuata dai creatori di queste liste di blocco nei confronti dei provider riguarda il fatto che questi rifiutino di intervenire quando all'interno dei loro server vengono individuati dei loro clienti che sono considerati fonte di Spam. Inoltre in queste liste di blocco vi sono indirizzi IP che risultano appartenere a organizzazioni propense allo sviluppo di software e risorse varie, utili agli Spammer e alla divulgazione dello Spam. In sintesi, in queste liste di blocco sono presenti indirizzi IP di provider o fornitori di hosting che hanno un comportamento assolutamente irresponsabile, dovuto al fatto che questi dovrebbero essere fra i primi a controllare che i propri clienti non siano degli Spammer. Invece, purtroppo, i provider sono spesso i primi che supportano gli Spammer e la loro attività.

- Liste di relay aperti

I relay aperti per diversi anni sono stati la principale causa di Spam. Il relay aperto è una particolare configurazione del server di posta elettronica che permette sia di accettare che di trasmettere messaggi di posta a chiunque, senza che il mittente e il destinatario delle e-mail in questione appartengano al sistema locale. Gli Spammer fanno uso illegale di questi relay aperti al fine di poter inviare una e-mail di posta indesiderata a molti destinatari contemporaneamente.

Però, a causa di ciò, si può verificare l'intasamento e l'esaurimento delle risorse disponibili del server di posta. Fortunatamente i relay aperti sono diminuiti e con essi anche lo Spam. Per poterli inserire in queste liste di blocco, i relay vengono sottoposti a test e a verifiche che controllano che non siano già utilizzati dagli Spammer.

- Liste di risorse abusabili per via di falle di sicurezza.

Anche in queste liste di blocco possono esservi dei relay aperti, ma quello che viene elencato sono le falle di sicurezza presenti all'interno di pagine web. Un esempio potrebbe essere lo script Formmail, utilizzato per poter inviare e-mail da form presenti all'interno di pagine web. Questo script infatti presenta delle falle di sicurezza che permettono agli Spammer di inviare e-mail indesiderate.

- Liste con criteri particolari.

L'ultimo caso di liste di blocco di indirizzi IP sono create andando ad analizzare alcune particolarità. Ad esempio si inseriscono indirizzi IP di domini dove non risulta funzionante l'indirizzo "abuse". Questo è un indirizzo ormai inserito e messo a disposizione da quasi tutti i provider di posta elettronica per segnalare abusi dovuti alla ricezione di Spam all'interno della propria casella di posta elettronica. Gli indirizzi IP sono inseriti in queste liste di blocco anche quando la registrazione nel "Whois" dell'indirizzo IP è falso. Il "Whois" è un servizio di rete che permette, attraverso un'interrogazione a un determinato database, di determinare a quale provider Internet appartiene quell'indirizzo IP.

Sicuramente le liste di blocco sono un ottimo sistema alla prevenzione dello Spam, prima che questo possa arrivare all'interno della casella di posta elettronica dell'utente.

5.2 Le liste di blocco utilizzabili

Le liste di blocco [17] non sono sempre consultabili poiché sono sempre in continua evoluzione; ne nascono e ne vengono chiuse continuamente. Inizialmente le liste di blocco nacquero per essere utilizzate da chi le creava, ora invece sono consultabili da tutti. Fra le prima liste di blocco create troviamo RBL (Realtime Blackhole List) di MAPS [17].

MAPS è un organizzazione guidata da Paul Vixie, il quale creò le RBL, che altro non sono che liste di indirizzi IP con relay aperti; contengono anche informazioni sugli host noti per l'invio di Spam. Queste liste di blocco sono state create per dare un aiuto contro la posta indesiderata e per poter filtrare a livello server di posta lo Spam. Questi blocchi di indirizzi possono essere consultati sia gratuitamente che a pagamento; inoltrare vengono anche consultate dal singolo utente, tramite una ricerca manuale sul sito presente nella rete. Fra le liste di blocco a pagamento troviamo:

- MAPS-RBL: in questa lista di blocco sono inseriti tutti gli indirizzi IP dei network utilizzati dagli Spammer, quelli che sono fonte di Spam o che comunque forniscono dei servizi utili allo Spammer;

- MAPS-RSS (Relay Spam Stopper): in queste liste di blocco vi sono elencati tutti gli indirizzi IP dei server che hanno problemi di relay aperti.

- MAPS-NML (Non Confirmed Mailing-list): in queste liste di blocco sono elencati tutti gli indirizzi IP, per i quali è stato dimostrato che ospitano mailin-list gestite in maniera scorretta, ovvero che usano le e-mail di vari utenti senza il consenso di questi ultimi.

Vediamo ora altre tipologie di liste di indirizzi che possono essere consultate, suddivise per categorie.

Liste di relay aperti o proxy aperti.

- ORDB (http://www.ordb.org/): era un'importante lista di relay aperti presenti nella rete, venne chiusa nel 2006. Questa permetteva di consultare i test che venivano effettuati sui proxy aperti affinché questi potessero essere giudicati fonte di Spam;

- DSBL (http://dsbl.org/): anche questa lista conteneva proxy aperti, e come ORDB è stata chiusa. Su questa era possibile consultare i test che si effettuavano sui proxy aperti per verificarne se fossero fonte di Spam;

- NSABL (http://www.njabl.org/): lista di blocco molto variegata, dove sono presenti elenchi delle diverse fonti di Spam, ad esempio relay e proxy aperti, indirizzi IP e i relativi test effettuati per decretare se fossero fonte di Spam;

- CBL (http://cbl.abuseat.org/): lista di blocchi dove sono presenti indirizzi IP, nella quale però non è possibile consultare molte informazioni; le uniche rese disponibili sono gli indirizzi IP e data e ora in cui questi indirizzi sono stati individuati come fonte di Spam.

Liste di indirizzi o di blocco lasciati a disposizione degli Spammer

- SBL: si tratta di una lista di blocco creata e gestita da uno dei più importanti antispammer, Steve Linford. Per questa lista di blocco è stato creato un vasto database on-line, all'indirizzo www.spamhouse.org, consultabile da chiunque lo ritenga opportuno. Dalle molte ricerche effettuate da Steve Linford è emerso che esistono poche ma ampie centrali fonte di Spam, che sono responsabili di quasi tutta la posta indesiderata presente nella rete.

Caratteristica fondamentale è che il database, di SBL è in continuo aggiornamento e non presenta alcun falso positivo. Questi si verificano quando anche una e-mail sicura viene considerata per errore Spam da parte del proprio client di posta elettronica.

- SPEWS: è sempre una lista di blocco, ma rispetto alle altre viste finora è molto innovativa per alcuni sui comportamenti nei confronti di chi diffonde Spam. La sua particolarità sta nel fatto che non è possibile risalire a chi siano i suoi amministratori poiché questi si tengono nell'anonimato.

Una delle politiche adottate da questa lista di blocchi è evitare che uno Spammer, una volta scoperto, possa continuare a inviare posta indesiderata. Questo atteggiamento cerca di limitare al massimo lo Spam bloccando i provider che ospitano gli Spammer. Questa modalità, il più delle volte crea problemi dovuti al fatto che bloccando i provider dove risiedono gli Spammer, si vanno a bloccare anche altri siti ospitati in quel provider, che però non risultano essere fonte di Spam. D'altra parte in questo modo, SPEWS è sicuro che molti provider andranno ad eliminare immediatamente quei suoi clienti che sono considerati fonte di Spam. Infatti da quando SPEWS esiste, sono molti i provider che hanno eliminato gli Spammer presenti al loro interno. In questi anni, SPEWS ha ricevuti molti attacchi da parte degli Spammer, dovuti al fatto che la sua politica sta ottenendo degli ottimi risultati, utili al bene di tutta la rete.

Liste di indirizzi su base geografica.

- The South Korea blocking list: è sicuramente fra le varie liste di blocco viste sin ora, quella più curiosa: questa infatti cerca di bloccare l'intero flusso di e-mail provenienti dalla Corea del Sud, poiché gli Spammer si sono accorti della grande quantità di proxy insicuri, quindi di facile uso per l'invio di posta indesiderata.

Stessa scelta è stata intrapresa da altre nazioni, come ad esempio: Brasile, Nigeria e Cina, dato che questo tipo di politica adottata nella lotta allo Spam è risultata un'ottima soluzione per la sua velocità e anche per gli ottimi risultati che sta ottenendo.

5.3 Strumenti software e tools

Per cercare di contenere lo Spam l'esperienza gioca un ruolo molto importante ed è sicuramente fondamentale, ma un valido aiuto ci è fornito dai software presenti sulla rete. Diamo uno sguardo ai programmi presenti in rete, utilizzabili da un utente per poter eliminare lo Spam all'interno della propria casella e-mail.

5.3.1 Spam Assassin

Fra i software presenti all'interno della rete, utili per difendersi dallo Spam, troviamo "Spam Assassin". Questo è un programma open-source distribuito dalla Apache Software Foundation, all'indirizzo www.spamassassin.org.

Spam Assassin, per ogni singolo messaggio di posta elettronica ricevuto, ricava informazioni, (ad esempio testo ed header) che vengono elaborate, confrontate ed infine viene espresso un giudizio sulla e-mail analizzata. Spam Assassin è un e-mail filter, quindi è un filtro applicabile sul client di posta elettronica utilizzato, capace di filtrare le e-mail in arrivo, onde evitare l'arrivo di Spam nella casella di posta elettronica.

Inoltre, per approfondire questi controlli e renderli sicuri, utilizza le varie liste di blocco presenti nella rete e i vari database di Spam-Tracking (ad esempio Razor, DCC).

Analizziamo i controlli in maniera più dettagliata.

- Razor è una network collaborativa e distribuita per la rilevazione dello Spam che si basa sul contributo degli utenti. Il funzionamento di Razor è molto semplice ed è utile per Spam Assassin: considerato che un messaggio di posta indesiderato viene inviato a molte persone ed è uguale per tutti, l'utente che per primo riceve il messaggio di Spam lo inserisce nel database di Razor, così se ci dovessero essere invii seguenti saranno riconosciuti e subito bloccati. Quando Spam Assassin confronterà le e-mail che arrivano, se sono presenti già nel suo database assegnerà loro un punteggio, altrimenti le invierà al database di Razor. Stessa cosa succede anche con DCC, infatti anche questo è una network collaborativa come Razor. In sintesi Razor e DCC collaborano con Spam Assassin affinché possano sempre

essere aggiornati sulla presenza di nuovi Spam presenti nella rete.

- Black List: Spam Assassin utilizza le varie black list presenti nella rete, che sono molto utili poiché sono sempre aggiornate, quindi sono un ottimo strumento su cui verificare la possibilità che una e-mail possa essere o meno considerata Spam.

- Analisi del testo: Spam Assassin riesce ad analizzare il testo delle e-mail, poiché questo nei messaggi di posta indesiderata ha delle caratteristiche particolari che il software riesce a individuare (ad esempio presenza di caratteri particolari, blocchi di codice).

- Analisi dell'intestazione: Spam Assassin, quando analizza l'intestazione di una e-mail, è in grado di rilevare incoerenze presenti nei messaggi di posta generati automaticamente.

Caratteristica importante di Spam Assassin è la bassa percentuale (0,1%) di falsi positivi, inoltre ha la capacità di bloccare oltre il 99% di posta indesiderata.

Il metodo usato da Spam Assassin per valutare e quindi considerare che una e-mail sia Spam o meno, è un metodo euristico. Il funzionamento è il seguente: sulle e-mail vengono effettuati dei test che prendono in considerazione diversi elementi; per ogni elemento controllato si assegna un punteggio alla e-mail analizzata, che potrà essere negativo o positivo, quindi utile per poter considerare una e-mail Spam o no. Spam Assassin effettua moltissimi test del genere onde evitare che gli Spammer li possano aggirare. Ogni test ha una sua valutazione, ottenuta controllando un campione di e-mail (considerate Spam) dove verranno estratte le qualità caratteristiche di un messaggio di posta indesiderato. Oltre ad effettuare diversi test sulle e-mail, anche il database di Spam Assassin è continuamente aggiornato nel momento in cui vengano trovati nuovi messaggi considerati Spam, e ciò rende il software affidabile e sicuro. I test eseguiti prendono in considerazione tre elementi principali: le proprietà tecniche, il contenuto delle e-mail e le informazioni di fonti esterne.

Le proprietà tecniche forniscono dati e informazioni più precise, infatti Spam Assassin riesce ad individuare se una e-mail sia stata generata automaticamente, quale sia il software utilizzato per spedire la e-mail, verifica la validità di data e ora presente nella e-mail, la dimensione del testo e il formato dell'indirizzo.

Quando invece Spam Assassin analizza il testo del messaggio di posta, effettua delle verifiche sul mittente, e se vi sono vocaboli o frasi ricorrenti nelle e-mail considerate Spam.

Infine Spam Assassin analizza le informazioni provenienti da fonti esterne, ovvero i siti conosciuti come fonte di Spam. Una volta che il software ha valutato i tre elementi sopra citati, ad ogni singola e-mail analizzata vengono assegnati dei punteggi che saranno confrontati con la somma dei test effettuati: se i punteggi dovessero superare una soglia limite allora una e-mail sarà considerata Spam.

Spam Assassin aggiunge delle intestazioni ai messaggi, che sono utili per capire se una e-mail è considerata Spam. Le intestazioni aggiunte sono:

- X Spam Level: indica il livello di Spam della e-mail; vengono aggiunti degli asterischi, secondo regole interne del software, nel campo subject: se questi superano il numero di 5, allora comparirà anche la scritta Spam nel campo subject.

- X Spam Status: racchiude una serie di indicazioni sulla e-mail; ad esempio, con il campo "Hits" si vuole indicare il punteggio ottenuto dalla e-mail, dopo le varie verifiche e test effettuati; il campo "Required" indica il limite per considerare una e-mail Spam; il campo "Test" invece indica i vari test che sono stati effettuati sulla e-mail.

- X Spam Checker Version: indica semplicemente la versione di Spam Assassin attualmente in uso sul PC utilizzato.

Su Spam Assassin inoltre è possibile apportare delle modifiche ai parametri del software, ad esempio per ottenere un filtraggio diverso in base alle proprie esigenze. Riassumendo, Spam Assassin è molto efficace nell'individuare lo Spam, ha un'ottima documentazione inoltre è sempre aggiornato. Unica pecca è che non è in grado di

gestire le e-mail, cioè non è possibile usarlo anche per eliminare o archiviare la posta indesiderata, poiché Spam Assassin si limita solo ad individuare lo Spam e a marchiarlo come tale.

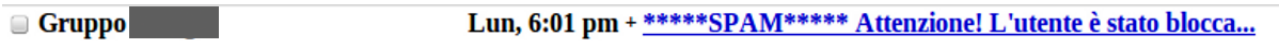

Figura 5.1 Esempio di e-mail segnalata come Spam da parte di Spam Assassin

5.3.2 Spamhilator

Spamhilator [18] è un software open source, sviluppato da programmatori tedeschi, utile agli utenti per eliminare lo Spam all'interno della propria casella di posta elettronica. Spamhilator supporta il protocollo di posta POP3, ed è possibile usarlo con tutti i clienti di posta elettronica (Eudora, Outlook, Netscape, Thunderbird). Questo software si comporta come un proxy server locale, da poter usare come filtro fra il client di posta elettronica (utilizzato dall'utente) e il server POP3 del provider Internet, infatti ha bisogno di essere accorpato come server di posta local host nei veri client di posta elettronica.

Spamhilator utilizza due modalità per poter filtrare le e-mail in arrivo nella casella di posta elettronica: il "Learning Filter" e il "Word Filter". La prima analizza le parole che solitamente non sono usate nei messaggi considerati Spam, ed ad ogni parola associa una certa probabilità; le parole analizzate sono presenti in un database utilizzato dal Learning Filter.

Nel caso in cui non basti utilizzare il Learning Filter per poter decretare che una e-mail sia Spam, si utilizzerà il Word Filter. Come la prima modalità, consulta un proprio database, ma, al contrario del Learning Filter, controlla tutte quelle parole che solitamente sono presenti nei messaggi considerati Spam.

Fondamentale di Spamhilator è che ogni utente potrà aggiungere delle parole all'interno dei database, utilizzati da Learning Filter e da Word Filter.

5.3.3 Spam Terminator

Spam Terminator [19] è un software open source, che permette di filtrare i messaggi di posta.

Questo programma può essere integrato con tutti i programmi client di posta elettronica, come Outlook, Thunderbird, Eudora e altri. Spam Terminator permette all'utente di aggiornare il database del software nel caso in cui questo riceva una e-mail Spam non presente nel database del software, inoltre offre anche la possibilità di personalizzare il programma in base alle proprie esigenze. Infatti l'utente che utilizza Spam Terminator, può segnalare al software quali messaggi non desidera accettare, ad esempio quei messaggi che arrivano da un utente non desiderato, o quei messaggi già ricevuti in passato e considerati Spam.

Spam Terminator si inserisce fra il proprio client di posta elettronica ed il server remoto. Tutto il lavoro svolto dal software è completamente automatizzato: Spam Terminator andrà immediatamente ad eliminare i messaggi rilevati come Spam, collegandosi al provider remoto ed eliminando lo Spam direttamente sul server di posta elettronica dell'utente. Per sicurezza vi è la possibilità di poter salvare messaggi eliminati in automatico dal software, poiché vi potrebbero essere alcuni errori, come l'eliminazione di una e-mail che in realtà non è Spam.

Dal punto di vista tecnico Spam Terminator funziona allo stesso modo di un server proxy POP3, infatti quando un utente vorrà leggere la propria casella di posta elettronica, utilizzando quindi il proprio client di posta, questo si collegherà a Spam Terminator, il quale, a sua volta, si connetterà sul server di posta remoto dove risiedono le e-mail dell'utente e qui inizierà a filtrare i messaggi di posta elettronica considerati Spam.

Le regole utilizzate per poter filtrare le e-mail da Spam Terminator sono le seguenti:

- nel campo "From" eliminerà quelle e-mail che non specificano questo campo; che non hanno indirizzi validi o non sono presenti; e-mail i cui destinatari specificati dall'utente siano non validi o non desiderati; tutte le e-mail provenienti da domini indesiderati dall'utente.

- nel campo "To" eliminerà quelle e-mail che non specificano questo campo o non è presente; e-mail con eccessivo numero di destinatari; e-mail ricevute da un mittente indesiderato

- nel campo "Oggetto" eliminerà quelle e-mail che contengono caratteri anomali, stringhe considerate indesiderate dall'utente; e-mail dove il campo oggetto sia vuoto.

- nel testo delle e-mail, controllerà se al suo interno vi possono essere stringhe considerate indesiderate o che solitamente sono presenti nelle e-mail di Spam.

5.3.4 MailWasher

Mail Washer [20] è un software utile agli utenti per evitare di ricevere dello Spam all'interno della propria casella di posta elettronica. Di questo software esiste anche una versione free, ma, a differenza di quella a pagamento, è utilizzabile per un solo account di posta elettronica.

Mail Washer, per evitare l'arrivo di e-mail indesiderate all'interno della propria casella di posta elettronica, si collega all' Internet Service Provider dove risiede la mailbox contenente le e-mail dell'utente. Il software controllerà automaticamente le e-mail considerate Spam e quelle che contengono dei file infetti da virus. Questa è la principale differenza rispetto agli altri software analizzati, come Spam Terminator, Spamhilato e Spam Assassin. Infatti Mail Washer riesce ad individuare i virus presenti negli allegati delle e-mail,

poiché solitamente gli allegati che contengono virus sono di grandi dimensioni. Perciò Mail Washer, oltre ad essere in grado di eliminare lo Spam, è anche un ottimo antivirus, utile a prevenire che il PC di un utente sia infettato dai virus presenti nelle e-mail che riceve.

Conclusioni

Per quanto riguarda il problema dello Spam, dato che questo fenomeno, come abbiamo visto, ha un diffusione vastissima nella rete, ho trovato molto interessante l'analisi delle metodologie utilizzate dagli Spammer per diffondere in maniera illegittima materiale di vario tipo, dal prodotto commerciale a materiale di dubbio gusto (ad esempio materiale pornografico). Abbiamo visto come alcuni accorgimenti semplici ma essenziali, come ad esempio non aprire messaggi di posta provenienti da indirizzi sconosciuti o scegliere un nome account particolare in modo da essere difficilmente indovinato dagli Spammer, così da limitare la proliferazione dello Spam. Questo per quanto riguarda ciò che un utente della rete può fare per arginare il problema in maniera semplice e rapida. Ma sarebbe certamente utile conoscere tutti gli altri metodi, decisamente più efficaci, per evitare l'intrusione di Spam nelle nostre caselle di posta elettronica; ad esempio software antiSpam. Come abbiamo visto, è stato fatto molto da quando l'uso della posta elettronica si è diffuso in maniera capillare, affinando e migliorando, grazie anche al contributo degli stessi utenti, le tecniche di prevenzione ed eliminazione dello Spam. Bisogna anche aggiungere, però, che anche gli stessi Spammer affinano e migliorano le tecniche di diffusione del loro materiale, e per fronteggiarli al meglio può risultare utile un'azione congiunta tra i sistemi informatici di prevenzione, che devono essere in continuo aggiornamento, l'attenzione da parte degli utenti nell'utilizzo della posta elettronica, una legislazione adeguata, attenta e aggiornata che preveda non solo sanzioni per chi agisce in maniera illegittima, ma anche tutela per chi invece cerca continue soluzioni al problema dello Spam.

Bibliografia

- [1] http://www.rfc-editor.org/rfc/rfc822.txt
- [2] http://www.rfc-editor.org/rfc/rfc1341.txt
- [3] http://www.rfc-editor.org/rfc/rfc821.txt
- [4] http://www.rfc-editor.org/rfc/rfc1725.txt+ù
- [5] http://www.rfc-editor.org/rfc/rfc1064.txt
- [6] http://gazzette.comune.jesi.an.it/2005/112/12.htm
- [7] http://www.blogmarketing.it/opt-in-e-opt-out-un-po-di-chiarezza/
- [8] State of Spam & Phishing Settembre 2010
- [9] http://www.privacy.it/codiceprivacy.html#art117
- [10] http://www.privacy.it/codiceprivacy.html#art117
- [11] http://www.parlamento.it/parlam/leggi/deleghe/03070dl.html
- [12] http://www.camera.it/parlam/leggi/deleghe/03196dl2.htm#161
- [13] http://www.camera.it/parlam/leggi/deleghe/03196dl2.htm
- [14] http://www.camera.it/parlam/leggi/deleghe/testi/05206dl.htm
- [15] http://it.wikipedia.org/wiki/Spambot
- [16] http://www.spyware.it/spyware.htm
- [17] http://www.collinelli.net/antispam/as0150.htm
- [18] http://www.ilsoftware.it/querydl.asp?ID=703
- [19] www.spamterminator.it
- [20] http://www.ilsoftware.it/articoli.asp?ID=1842

Lista sigle presenti nella tesi:

- DNS (Domain Name Server)
- ARPANET (Advanced Research Projects Agency Network)
- MUA (Mail User Agent)
- MTA (Mail Transfer Agent)
- MIME (Multipurpose Internet Mail Extentions)
- IMAP (Internet Message Access Protocol)
- IANA (Internet Assigned Numbers Authority)
- TCP (transmission Control Protocol)
- IP(Internet Protocol)
- UBE (Unsolicited Bulk Email)
- UCE (Unsolicited Commercial Email)
- UROL(Uniform Resource Locator)
- HTML (Hyper Text Markup Language)
- MAPS (Mail Abuse Protection System)
- RBL (Realtime Blackhole List)
- RSS (Relay Spam Stopper)
- NML (Non-Confirmed Mail List)
- ORDB
- DSBL (Distribuited Sender Boycom List)
- NSABL (Not Just Another Bogos List)
- CBL (Composite Blocking List)
- SBL(Spamhouse Block List)
- SPEWS (Spam Prevention Early Blocking System)
- S/MIME (Secure/Multipurpose Internet Mail Extensions)
- SMTP (Simple Mail Transfer Protocol)