UNIVERSITÀ DEGLI STUDI DI PADOVA

FACOLTÀ DI SCIENZE STATISTICHE

Corso di Laurea in STATISTICA E GESTIONE DELLE IMPRESE

TESI DI LAUREA

ANALISI DEL TRAFFICO SMTP DI UN ISP

Relatore: Ch.mo Prof. Graziano Deambrosis

Laureando: Roberto Matteotti

Matricola: 546288-GEI

ANNO ACCADEMICO 2005-2006

"perché sono le sfumature a dare vita ai colori
e a farci tornare in mente le cose più pure
dei giorni migliori"

("I giorni migliori", Tiromancino)

.. *a mamma*,
per avermi *fortemente* consigliato
di compiere questo passo ..

.. *a papà*,
per la disponibilità
sempre manifestata ..

.. *e a me*,
pronto per
un'altra avventura :-)

PREMESSA

Ho scelto di far passare sotto la lente d'ingrandimento la dinamica del server che regola il traffico dei messaggi di posta elettronica, in entrata ed in uscita, gestiti dal provider internet presso cui lavoro.

Analizzerò composizione e dimensione di tale flusso, per capire se e in che misura la mole di messaggi di posta elettronica potrebbe rendere precaria la stabilità dello stesso server.

Se i virus restano pericolosi per la salute del computer, e tenendo in considerazione gli altri attacchi che minacciano la salvaguardia dei dati personali, ricevere ogni giorno decine, se non centinaia messaggi di posta indesiderata infastidisce non poco l'utente, intento nel ripulire la propria casella.

Focalizzerò quindi l'attenzione sui messaggi considerati spam, vero bubbone di questi ultimi anni, poiché in prima persona sono impiegato nel contrastare le insidie che minano la tranquillità ed il lavoro dei nostri clienti.

Auguro a tutti un'incuriosita lettura!

Roberto Matteotti

La società Deerfield mi autorizza a pubblicare i nomi dei marchi registrati e le immagini derivanti dai propri software.

SOMMARIO

SPAM, EPIDEMIA GLOBALE

"Il 70% di tutti i messaggi e-mail in internet è costituito da spamming"

[Fonte: "Guide indispensabili per la sicurezza informatica", Watchguard, 2005]

Le prime cinque categorie di spam classificate da gennaio a giugno 2005:

1. cure/pillole: 41,4 %
2. mutui: 11,1 %
3. contenuti per adulti: 9,5 %
4. frodi azionarie: 8,5 %
5. prodotti: 8,3 %

Altro: 21,2 %

[Fonte: Sophos, da "ICT Security Newsletter" del 6/9/5]

L'utilizzo del termine spam per identificare la moltitudine di posta elettronica indesiderata risale ad uno sketch del "Monty Python's Flying Circus", dove in un locale ogni pietanza del menù era a base di un tipo di carne in scatola (chiamato appunto "spam"). Il continuo invocare tale piatto ricordò e quindi suggerì il nome da attribuire a questo fenomeno informatico.

Pur restando una delle principali insidie alla sicurezza ed affidabilità del computer, la minaccia dei virus sembra essere oscurata, almeno come emergenza, dall'epidemia dello spam. Diffusione apparsa peraltro di recente ed inizialmente sottovalutata, considerando che i messaggi pubblicitari da cancellare erano in media meno di dieci!

A chi giova generare spam e come ci riesce?

L'operazione "anomala" di marketing punta in genere sul fatto che, spedendo un'offerta pubblicitaria a centinaia di migliaia di indirizzi pescati da internet (ogni navigatore lascia traccia della sua rotta), il committente potrà averne un ritorno economico magari da un decimo di essi. Tale indirizzario "improprio" verrebbe poi venduto ad altre figure senza scrupoli.

Cosa può fare l'utente?

Attualmente esistono dei filtri che marchiano i messaggi indesiderati da quelli che possiamo definire "normali", e che per fortuna evitano all'utente una considerevole perdita di tempo (e quindi di denaro).

Oltre a non rispondere a tali messaggi, neppure per farsi rimuovere da queste liste, il malcapitato dovrebbe evitare di inserire il proprio indirizzo e-mail nei siti internet, né tanto meno nei newsgroup (gruppi di discussione).

Cosa può fare il provider?

Il lettore può iniziare a comprendere quanto arduo e critico sia il compito di chi configura ed amministra caselle di posta elettronica, dovendo intercettare

non più solo i messaggi contenenti virus, ma anche interpretare ciò che potrebbe risultare indesiderato.

Fermo restando che viene sempre salvaguardata la privacy del cliente (i messaggi non vengono analizzati manualmente né da terze persone, ma in modo automatico secondo certi canoni e regole internazionali), il provider dà libertà all'utente di decidere se archiviare ciò che viene considerato spam in quarantena.

E' fortemente sconsigliato rigettare o cestinare a priori la posta indesiderata, perché può capitare che nell'"immondizia" finisca "erroneamente" qualche messaggio sano (il cosiddetto "falso positivo"). Virgoletto "erroneamente" perché la presenza di "falsi positivi" (così come di "falsi negativi") non è dovuta alla distrazione del server di posta, quanto piuttosto alla limitata severità di alcune regole adottate o dalla dubbia professionalità con cui il mittente formatta i messaggi, soprattutto quelli automatici (costruzione impropria di newsletter, conoscenza superficiale del linguaggio html, ecc.).

L'alfabetizzazione e la presa di coscienza degli utenti, unite ad un maggiore impegno da parte dei provider, riusciranno a ridimensionare, se non a debellare, questa antipatica (più che dannosa) forma di epidemia.

parte I

TRIVENET, PROFILO DI UN ISP

Trivenet S.p.A. nasce nel 1996 come internet provider, offrendo servizi di connettività dial-up (analogica/isdn), registrazione e mantenimento nomi a dominio e caselle di posta elettronica.

Nel 2002 diventa operatore di telecomunicazioni per il wireless local loop (WLL). Tecnologia che, grazie a celle di ripetizione installate nel territorio, consente di raggiungere direttamente l'utente finale senza passare per il cavo telefonico, ma con un terminale radio installato presso il cliente trasportando dati ad alta velocità e fonia.

Dal 2003 lavoro in Trivenet in qualità di sistemista (gestione pc e server aziendali, sia interni che pubblici), fornendo anche assistenza tecnica ai clienti nella risoluzione di problemi legati alla posta elettronica e allo spazio web da noi gestito.

ADAM

“Adam chi?!?”

Attribuire un nome alle proprie “creature” gratifica il sistemista che configura ed installa i server: mentre c’è chi pesca tra nomi di pesci o di stelle, io ho scelto di assegnare al mailserver questo nome, sfogliando la lista degli “angeli” di una serie animata giapponese.

Adam è un elemento chiave di tutta la serie, così come il server di posta di un internet provider riveste un ruolo importante, fondamentale.

Alcune caratteristiche di Adam:

- modello: HP Proliant DL380
- sistema operativo: Microsoft Windows Server 2003
- processore: Intel Xeon
- memoria ram: 2 gb

Il software che configura e gestisce le migliaia di caselle di posta elettronica si chiama VisNetic, prodotto dalla statunitense Deerfield. Grazie ad opportuni filtri ed opzioni è in grado di individuare messaggi contenenti virus ed interpretare le comunicazioni non desiderate, riconoscendole come spam.

Situazione all’11 di agosto 2005:

- domini gestiti: 353
- caselle configurate: 2218

TECNOLOGIA ANTISPAM

La tecnologia antispam multilivello di Adam si avvale dei seguenti filtri:

White & Black List

Anticipa il motore antispam. Se il destinatario garantisce per il mittente inserendolo in una lista bianca, viene evitato il controllo antispam.

Bayesian Filtering

Statisticamente col metodo di Bayes calcola la probabilità che ogni messaggio d'ingresso sia in realtà spam, investigandone il contenuto e riferendosi ad una banca dati di riferimento. Se almeno al 90% (valore personalizzabile) quel messaggio è ritenuto spam, allora esso viene marchiato come tale.

Heuristic Analysis

Il successivo livello si basa sulla tecnologia *SpamAssassin*, che euristicamente esamina l'intestazione (header) e il testo di ogni messaggio utilizzando centinaia di regole e controlli. Ad ogni messaggio viene dato un punteggio basato sulla probabilità che il messaggio sia spam.

HTML Filtering

Verifica come sono formattate le mail, soprattutto quelle automaticamente generate. Spesso lo spam non ha la formattazione corretta.

Content Filters

Controlla il contenuto e, in base a determinate regole definite dall'utente o a livello di server, compie determinate azioni.

Self Learning Mode

Il sistema "impara da solo" dopo che l'utente gli ha insegnato come comportarsi quando arrivano "falsi negativi" (= spam considerato invece come posta "legittima") o falsi positivi (messaggi "sani" ritenuti erroneamente spam).

SURBL (Spam URI Realtime Blocklist)

Diversamente dalle RBL che controllano solo l'indirizzo e-mail del mittente o l'indirizzo ip di provenienza, le SURBL marchiano i messaggi che contengono al loro interno i domini ritenuti "portatori" di spam.

SPF (Sender Policy Framework)

Verifica l'autenticità del mittente facendo una ricerca sul dns, riuscendo a capire se il server mittente è autorizzato a spedire email per conto di quel mittente.

L'ANALISI

Intendo dividere l'oggetto di questo studio in due tronconi:

- monitoraggio in tempo reale di alcuni contatori del mailserver (*periodo: 14 giugno – 14 luglio 2005*)

- sintesi di alcune serie storiche, che diano un'idea dei "numeri" che amministra un internet provider
  (*periodo: 1 marzo – 31 agosto 2005*)

È importante far presente che se il protocollo che trasmette i messaggi di posta (smtp) non dovesse essere attivo e funzionante ciò provocherebbe l'interruzione di ogni comunicazione e l'impossibilità di consegnare e recapitare i messaggi ai destinatari. Fondamentale quindi garantirne la continuità.

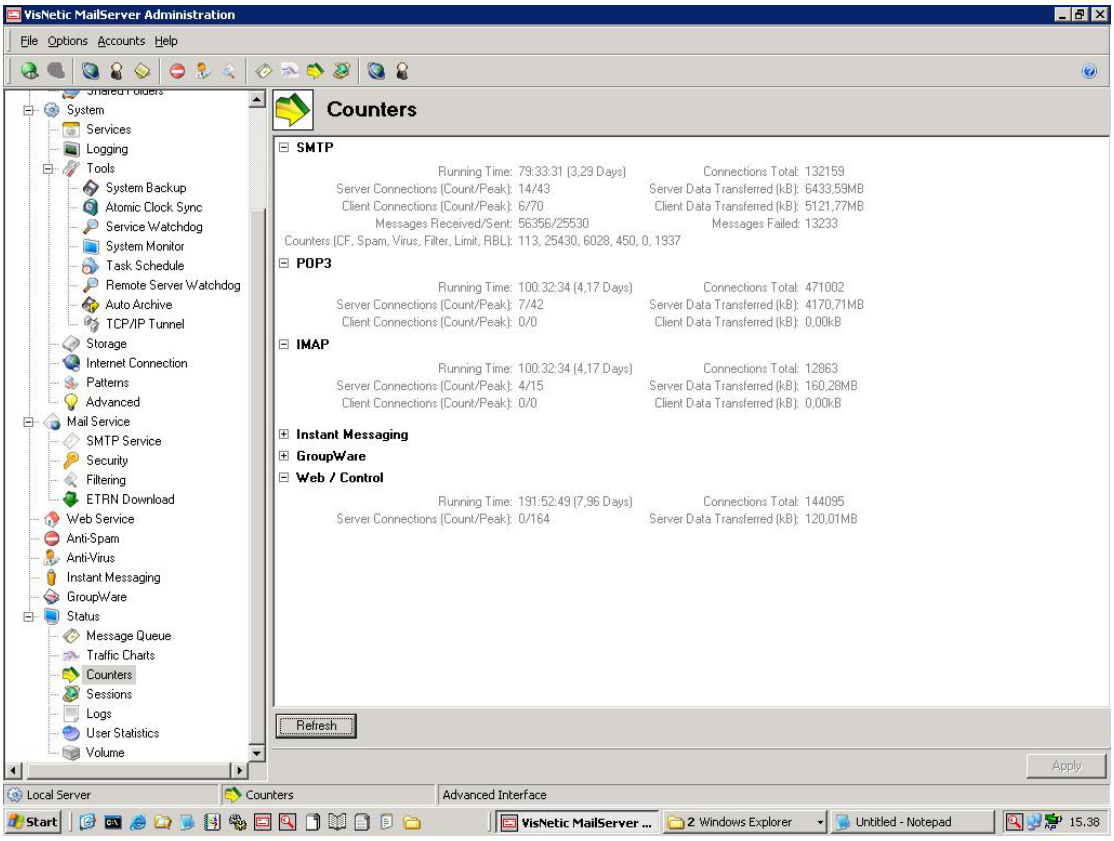
Dalle rilevazioni del primo periodo constaterò invece che il servizio smtp si arresterà più volte, per poi ripartire dopo alcuni minuti.

Le considerazioni prodotte dalla seconda parte dell'analisi permetteranno di individuare soluzioni che potrebbero evitare l'eventuale congestione del server.

parte II

RILEVAZIONI ISTANTANEE

L'interfaccia di configurazione e gestione del mailserver appare come in figura 1.



[Figura 1: interfaccia di configurazione del mailserver]

Alla sinistra il menù dei comandi. Nella parte di destra appare invece il monitoraggio dei contatori, che in tempo reale aggiornano l'incremento dei messaggi ricevuti, spediti e delle connessioni smtp/pop3/imap stabilite dai clienti, nonché la dimensione dei dati trasferiti.

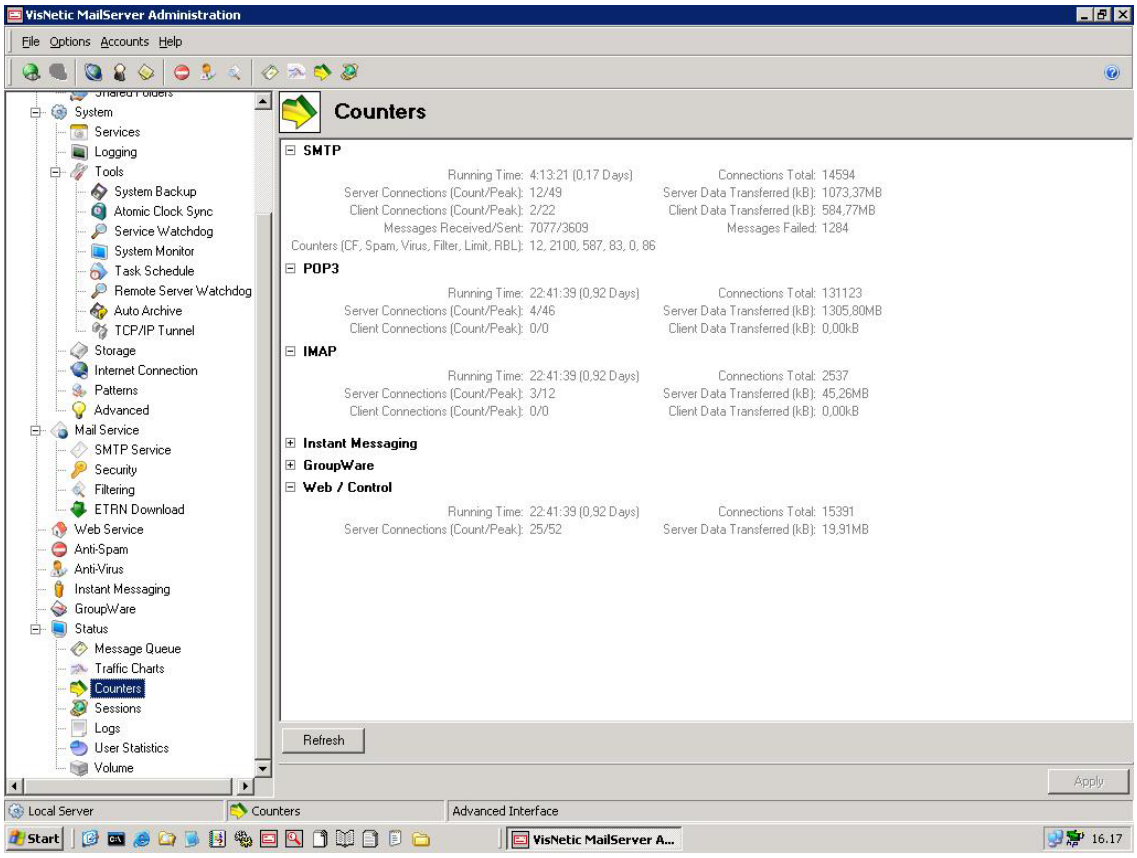
Al riavvio di ogni servizio i contatori si azzerano: se si osservasse un valore basso nel conteggio delle ore trascorse si dedurrebbe quindi che il servizio è stato riavviato da poco e sarebbe auspicabile indagarne la motivazione.

Scattando delle istantanee durante i mesi di rilevazione mi sono accorto che troppo spesso si azzeravano i contatori: già una volta al giorno sarebbe di per sè un indicatore negativo per un provider, che come già ricordato deve garantire la continuità del servizio.

Il periodo di osservazione dei rilevatori di Adam è iniziato il 14 giugno ed è terminato il 14 luglio; due volte al giorno, durante la settimana lavorativa, ho estrapolato il valore dei contatori che mi appresto a descrivere.

CONTATORI

Riportando l'istantanea scattata il 21 giugno 2005 alle ore 16.17 descrivo sinteticamente il significato dei contatori più importanti.



[Figura 2: istantanea del 21 giugno 2005]

Running Time

Il periodo intercorso dall'ultimo riavvio del servizio, espresso in ore e giorni.

Count/Peak

Il valore di un contatore in tempo reale e il suo picco.

Server Connections

Nell'architettura client/server quando Adam figura come server. Quando ad esempio l'utente, interno o esterno al mailserver si collega ad Adam tramite il servizio smtp per spedire un messaggio.

Client Connections

Nell'architettura client/server quando Adam figura come client. Quando ad esempio un messaggio di posta viene indirizzato da un utente interno verso un utente esterno, il mailserver stabilisce una connessione con il server del destinatario e spedisce il messaggio, figurando quindi come client.

Messages Received/Sent

Il numero di messaggi ricevuti e spediti dal mailserver nel periodo di tempo indicato dal contatore "Running Time".

Messages Failed

La consegna di un messaggio può fallire per diverse ragioni, principalmente dovute a destinatari sconosciuti o alle restrizioni di sicurezza del mailserver.

Spam/Virus

Il numero di messaggi identificati dal mailserver come spam e come contenenti virus.

RBL

Quanti messaggi provengono da indirizzi ip e/o domini inseriti in qualche "lista nera" ("Realtime Blocklist").

Server Data Transferred

Quantità di dati, espressi in kb, trasferiti durante le "Server Connections".

Client Data Transferred

Quantità di dati, espressi in kb, trasferiti durante le "Client Connections".

Prenderò in considerazione le rilevazioni del servizio smtp, in quanto è tramite questo protocollo che viene consentito il flusso in entrata e in uscita dei messaggi di posta. All'ingresso essi vengono scandagliati dagli opportuni filtri di sicurezza (antispam/antivirus/rbl, ecc.) e solo poi destinati alle caselle di posta dei "condòmini" del mailserver.

Riporto un estratto delle rilevazioni relative al periodo oggetto d'esame.

Data	***Mar 14/6***	***Mar 14/6***	***Mer 15/6***	***Gio 16/6***	***Lun 20/6***	***Lun 20/6***
Ora	***15.39***	***17.54***	***10.18***	***9.26***	***10.00***	***14.09***
Running Time	79:33:31	81:49:46	98:13:13	15:46:07	21:55:40	26:04:43
Days	3,29	3,38	4,08	0,63	0,88	1,08
Server Connections						
Count	14	11	10	20	13	11
Peak	43	43	139	106	23	29
Client Connections						
Count	6	3	2	4	6	4
Peak	70	70	70	40	33	33
Messages						
Received	56356	60607	72053	10528	11477	18567
Sent	25530	28006	32897	4262	4367	8126
Failed	13233	13856	16736	2613	2956	4357
Counters						
CF	113	120	145	15	37	45
Spam	25430	26432	31379	4949	6555	8594
Virus	6028	6361	7643	1063	975	1743
Filter	450	484	742	272	0	56
Limit	0	0	0	0	0	0
RBL	1937	1984	2376	351	445	521
Data Transferred						
Server D. T. (MB)	6433,59	7183,31	8443,12	942,73	595,04	1546,86
Client D. T. (MB)	5121,77	5606,95	6558,69	656,49	336,27	927,41

[Tabella 1: rilevazioni delle istantanee dal 14 al 20 giugno 2005]

Semplicemente scorrendo i valori di questa tabella si può osservare come tra il 15 e 16 giugno il servizio smtp si è fermato ed è poi ripartito, azzerando difatti tutti i contatori.

Ricordo che quando l'smtp non è attivo i messaggi non possono essere presi in consegna (e quindi opportunamente incasellati), né spediti verso l'esterno. Ecco perché il disservizio anche di pochi minuti penalizza il provider.

Fa inoltre un certo effetto prendere coscienza che in poco più di 3 giorni oltre 25.000 (venticinquemila!) messaggi sono stati considerati come spam e che sono stati trasferiti qualcosa come 6 (sei!) gigabyte di dati!

STATISTICHE

Dai valori osservati elaboro alcune statistiche, di cui riporto una sintesi:

Data	***Mar 14/6***	***Mar 14/6***	***Mer 15/6***	***Gio 16/6***	***Lun 20/6***	***Lun 20/6***
Ore	***78,96***	***81,12***	***97,92***	***15,12***	***21,12***	***25,92***
Server Conn. (Peak / Ore)	0,544579534	0,530079	1,419526	7,010582	1,089015	1,118827
Client Conn. (Peak / Ore)	0,886524823	0,862919	0,714869	2,645503	1,5625	1,273148
Msg Received / Ore	713,7284701	747,1277	735,8354	696,2963	543,4186	716,3194
Msg Sent / Ore	323,3282675	345,2416	335,9579	281,8783	206,7708	313,5031
Msg Failed / Ore	167,5911854	170,8087	170,915	172,8175	139,9621	168,0941
CF / Ore	1,431104357	1,47929	1,480801	0,992063	1,751894	1,736111
CF / Msg Received (%)	0,20	0,20	0,20	0,14	0,32	0,24
Spam / Ore	322,0618034	325,8383	320,4555	327,3148	310,3693	331,5586
Spam / Msg Received (%)	45,12	43,61	43,55	47,01	57,11	46,29
Virus / Ore	76,34245187	78,41469	78,05351	70,30423	46,16477	67,24537
Virus / Msg (%)	10,70	10,50	10,61	10,10	8,50	9,39
Filter / Ore	5,699088146	5,966469	7,577614	17,98942	0	2,160494
Filter / Msg Received (%)	0,80	0,80	1,03	2,58	0,00	0,30
Limit / Ore	0	0	0	0	0	0
Limit / Msg Received (%)	0,00	0,00	0,00	0,00	0,00	0,00
RBL / Ore	24,53140831	24,45759	24,26471	23,21429	21,07008	20,10031
RBL / Msg Received (%)	3,44	3,27	3,30	3,33	3,88	2,81
Server D. T. (MB) / Ore	81,47910334	88,55165	86,22467	62,34987	28,17424	59,67824
Client D. T. (MB) / Ore	64,86537487	69,11921	66,98009	43,41865	15,92188	35,77971

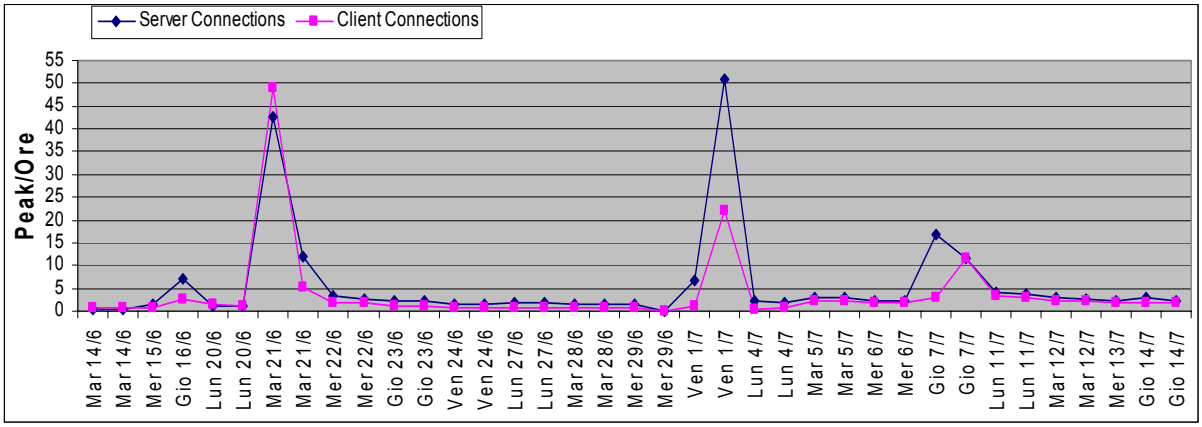
% su Msg Received

	Mar 14/6	***Mar 14/6***	***Mer 15/6***	***Gio 16/6***	***Lun 20/6***	***Lun 20/6***
CF	0,20	0,20	0,20	0,14	0,32	0,24
Spam	45,12	43,61	43,55	47,01	57,11	46,29
Virus	10,70	10,50	10,61	10,10	8,50	9,39
Filter	0,80	0,80	1,03	2,58	0,00	0,30
Limit	0,00	0,00	0,00	0,00	0,00	0,00
RBL	3,44	3,27	3,30	3,33	3,88	2,81
Msg Sani	39,74	41,62	41,31	36,84	30,19	40,98

[Tabella 2: statistiche delle istantanee dal 14 al 20 giugno 2005]

GRAFICI

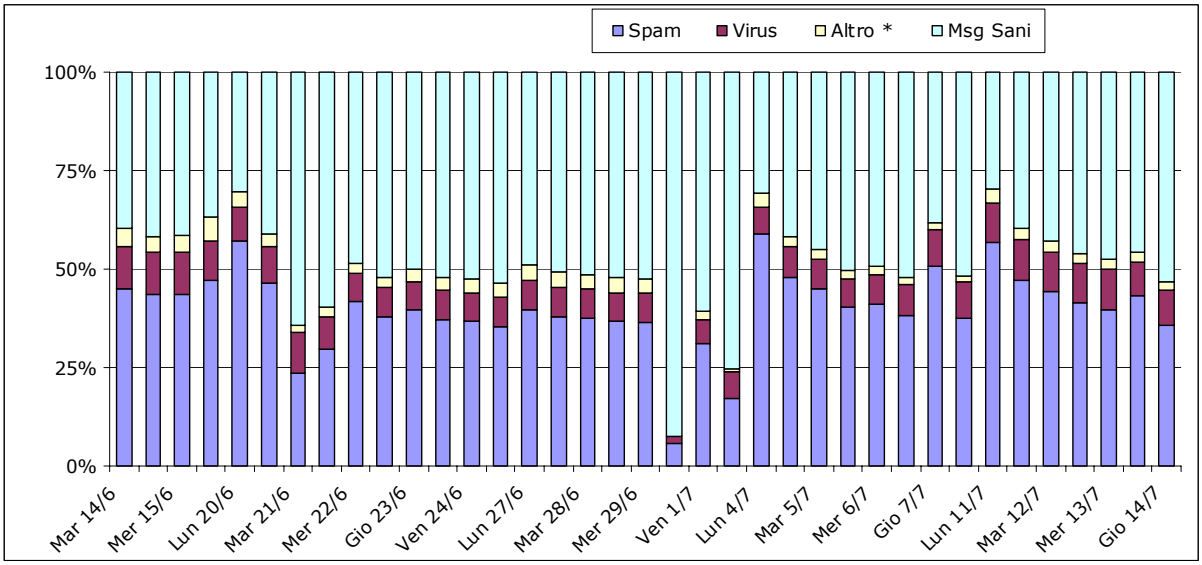
Dalle statistiche elaborate costruisco dei grafici che aiutano a suggerire alcune considerazioni. Pubblico solo alcuni dei più significativi.



[Grafico 1: i picchi delle connessioni client/server rapportati alla durata del servizio smtp ("running time")]

Scelgo questo tipo di grafico perché sono convinto che l'andamento delle linee rispecchi meglio i picchi causanti o successivi al riavvio del servizio smtp.

Per dare un'idea di quanto sia l'incidenza dei messaggi di spam o contenenti virus rispetto ai messaggi legittimi preferisco utilizzare un istogramma piuttosto che una torta. Sintetizzare i valori con una media, magari geometrica, non permetterebbe di intuire l'andamento del rapporto tra i messaggi "maligni" da quelli "benigni".



[Grafico 2: istogramma in pila 100% che confronta il contributo in percentuale di ogni mail al totale in più categorie: messaggio ritenuto spam, contenente virus, *inserito in qualche "black list"]

Grazie a questo grafico denotiamo come raramente i messaggi di spam abbiano costituito la metà dei messaggi ricevuti, anche se in media ne siamo molto vicini!. Siamo peraltro lontani dal quel 70% citato in apertura.

Ribadisco che comunque stiamo osservando un periodo molto limitato: la finestra temporale di un mese non aiuta a puntellare le nostre convinzioni, ma sicuramente offre le sembianze dell'architettura del traffico smtp.

Non tragga inganno la modesta percentuale di virus: nascono ogni giorni nuove infezioni e ne basta una per dire addio al nostro calcolatore e ai preziosi dati in esso contenuti!

CONSIDERAZIONI

Un'analisi statistica approfondita richiederebbe una standardizzazione dei valori rispetto al numero di domini o di caselle di posta elettronica, riuscendo a dare delle indicazioni indipendenti dal numero di mesi rilevati e di caselle configurate.

Non avrebbe senso altrimenti confrontare un mese di rilevazione rispetto ad un altro, venendo magari alla conclusione che il fenomeno spam è diminuito, solo perché alcune caselle di posta elettronica infestate di spam sono state trasferite presso un altro provider (e quindi cancellate da Adam), piuttosto che considerare un aumento di virus intercettati in seguito ad un incremento di clienti gestiti.

Riprendendo l'ultimo grafico è interessante notare come la percentuale di messaggi infetti risulti essere assai bassa. In realtà possiamo contare anche 7.643 messaggi contenenti virus in su quattro giorni (rilevazione del 15 giugno).

Di gran lunga inferiore comunque al numero di messaggi marchiati come spam: ben 31.379!

Conseguenza o causa?

I picchi derivano dal riavvio del servizio smtp o sono proprio essi la causa di tale interruzione/ripartenza?

Entrambe le ipotesi paiono verosimili: picchi elevati identificano una congestione del traffico rappresentati dai server di posta esterni che, riaperto il

valico, sgomitano per interagire con Adam (*conseguenza*), oppure ricordano la condizione di un blackout per sovraccarico di apparecchiature elettriche (*causa*), nel nostro caso connessioni smtp.

Se non avessi "fotografato" quotidianamente le istantanee probabilmente non mi sarei accorto di queste microinterruzioni del servizio smtp e quindi della presenza di eventuali problemi o criticità.

Interruzioni che in questo periodo (14 giugno – 14 luglio) non sono state poi molte, ma che si sono ripetute con più frequenza successivamente e che citerò nelle considerazioni finali.

Sottolineando che una buona analisi dovrebbe basarsi su un periodo di rilevazione ben più longevo, è altrettanto vero che non posso scattare ogni giorno un paio di istantanee!

Auspico quindi che Deerfield preveda in futuro uno sviluppo ed un radicamento delle statistiche insite in VisNetic, certo che altri isp suggeriranno le mie stesse raccomandazioni.

parte III

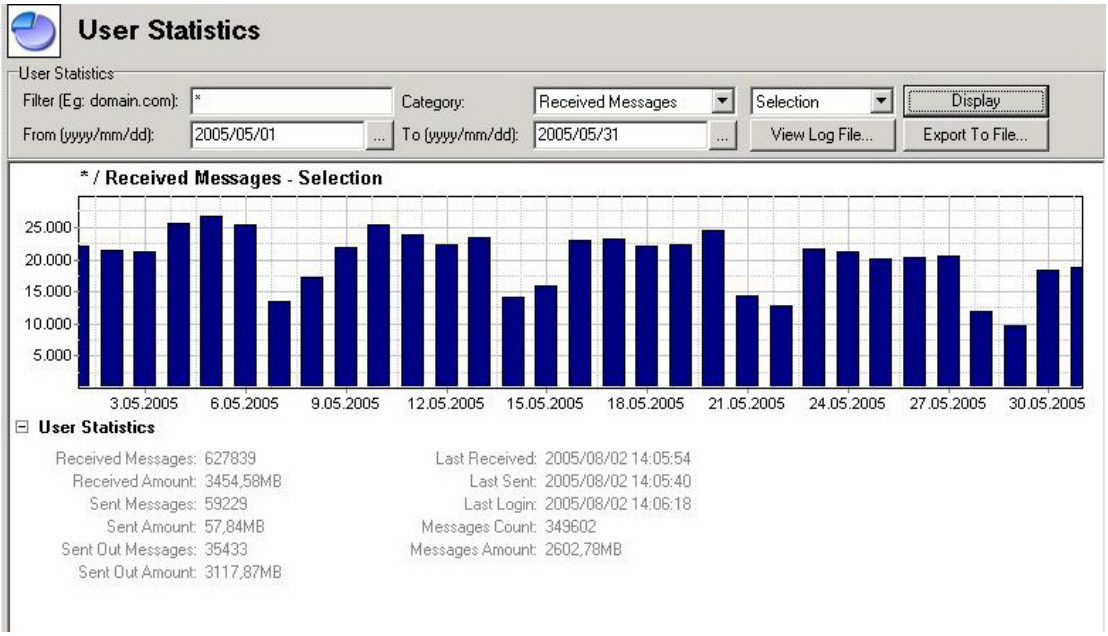
CHE NUMERI!

Nella seconda parte dell'analisi del traffico smtp di Adam riporto le statistiche dei messaggi ricevuti e spediti, non derivate da rilevazioni istantanee bensì da serie temporali riprodotte dal quadro "User Statistics", altra interessante utility di VisNetic.

Tali rappresentazioni aiuteranno il lettore a rendersi conto delle dimensioni della "popolazione" di messaggi che affollano un internet provider, ma supporteranno anche la direzione aziendale di Trivenet nel decidere se implementare un secondo server o se corazzare maggiormente quello esistente.

Il periodo considerato inizia il *primo marzo* e termina il *trentuno agosto*: prima di allora erano ancora in corso le operazioni di migrazione dal vecchio server di posta e il filtro antispam non era ancora a pieno regime.

Anche se non si osservano particolari anomalie pubblico in figura 3 il risultato dell'estrazione dai file di log dei messaggi ricevuti, giorno per giorno, nel mese di maggio 2005.



[Figura 3: messaggi ricevuti nel mese di maggio 2005]

Confrontando anche gli altri mesi registro durante il fine settimana di ogni mese, ma non è una sorpresa, una diminuzione dei messaggi ricevuti, ricordando comunque che virus e spam non vanno mai in vacanza!

Più interessante rispetto all'istogramma giornaliero si rivelerà essere il conteggio su base mensile dei messaggi gestiti da Adam.

CONTATORI

Descrivo brevemente i contatori presi in considerazione.

Received Messages

Messaggi consegnati nella casella dell'utente.

Sent Messages

Messaggio spediti, compresi quelli tra utenti del mailserver (posta "interna").

Sent Out Messages

Messaggi spediti dal mailserver verso l'esterno.

STATISTICHE

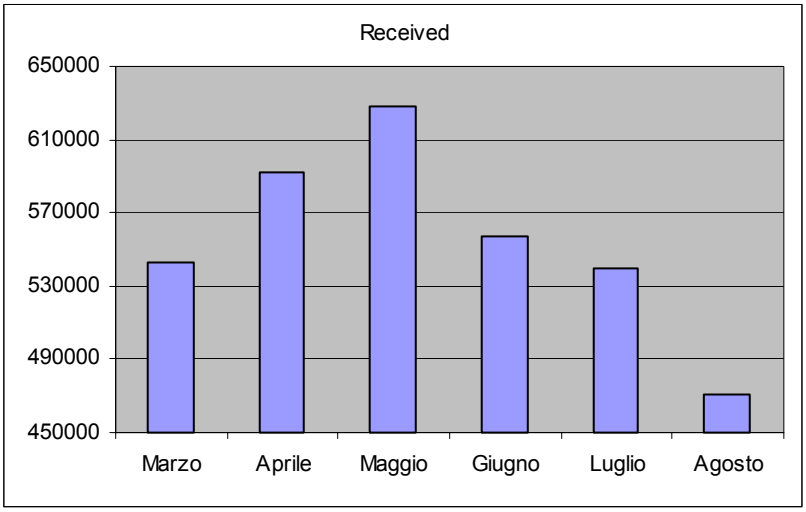
Riporto le categorie di messaggi ed il loro ammontare, mese per mese.

	Received	*Sent*	*Sent out*
Marzo	543.241	45.179	25.645
Aprile	591.972	46.406	26.522
Maggio	627.839	59.299	35.433
Giugno	557.394	58.367	38.284
Luglio	539.299	61.001	39.200
Agosto	471.269	47.481	24.410

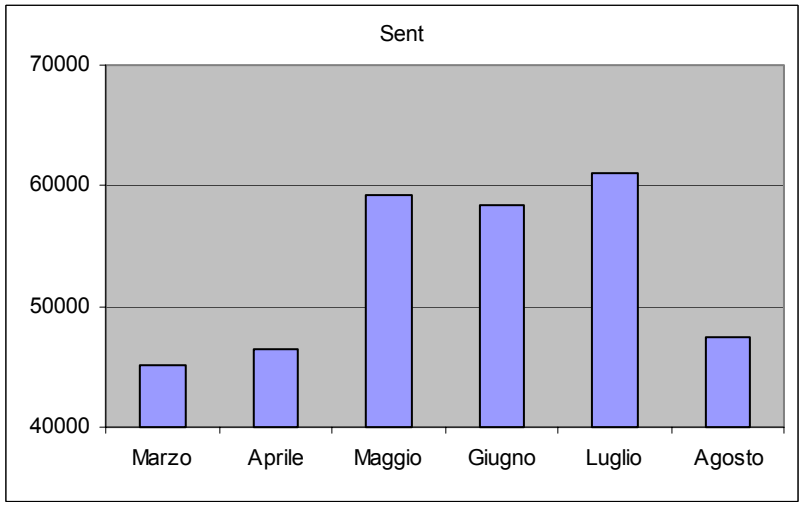
[Tabella 3: conteggio dei messaggi ricevuti e spediti, marzo-agosto 2005]

GRAFICI

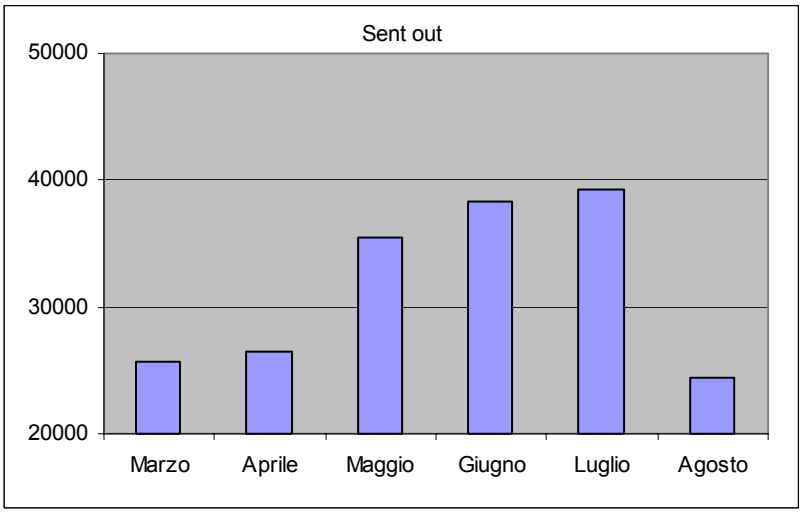
Per agevolare la lettura e la comprensione dei valori trascritti in precedenza inserisco i seguenti istogrammi:



[Grafico 3: messaggi ricevuti, marzo-agosto 2005]



[Grafico 4: messaggi spediti (compresa la posta "interna"), marzo-agosto 2005]



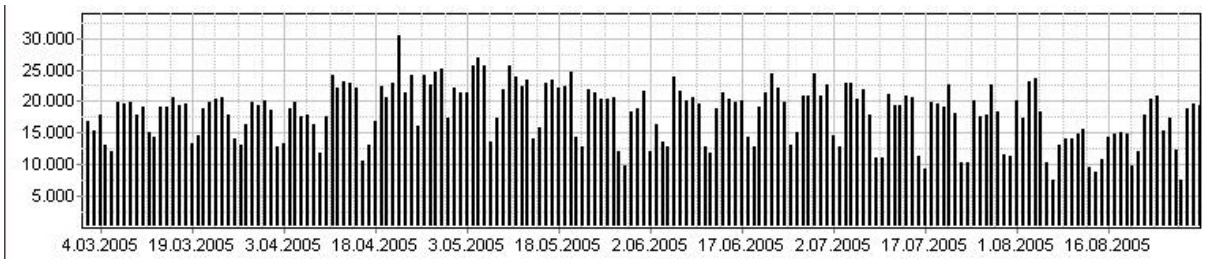
[Grafico 5: messaggi spediti verso l'esterno, marzo-agosto 2005]

Considerando che molti clienti spediscono la loro posta attraverso il nostro mailserver (relay di posta), pur non avendo domini o caselle configurate in Adam (abilitiamo infatti il relay a chi utilizza nostre connessioni internet), mi sarei aspettato che fossero di più i messaggi spediti rispetto a quelli ricevuti.

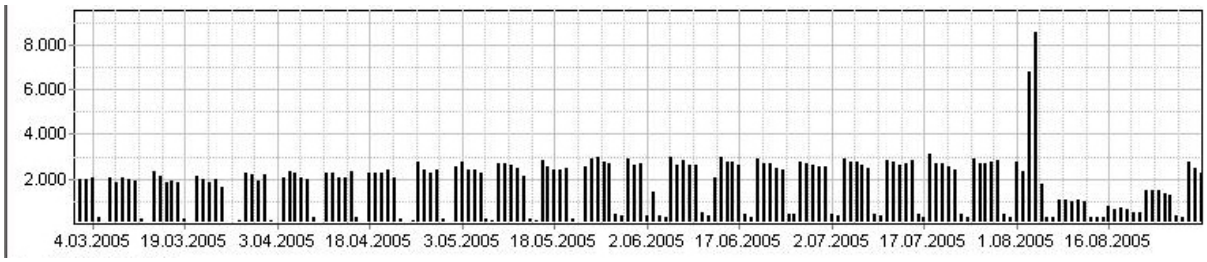
In realtà la tendenza è l'opposta di quella prevista: sono maggiormente numerosi i messaggi in entrata, anche dieci volte di più rispetto a quelli in uscita! Dimensioni gonfiate sicuramente dai messaggi di spam (in misura minima dai virus) che quotidianamente aggravano ogni mailserver mondiale.

Basterebbe questa considerazione a far riflettere su quanto imponente sia la diffusione di spam, tanto importante il suo annientamento: la sfida informatica e sociale del secolo!

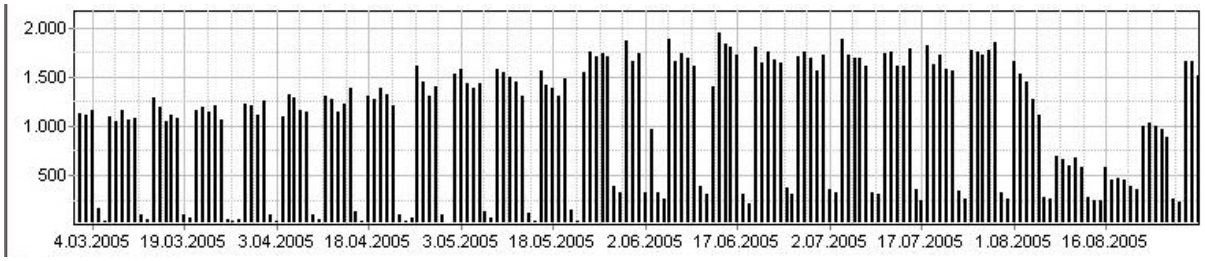
Riuscendo a raggruppare in un unico grafico tutti i messaggi passati per il mailserver, giorno per giorno in quei sei mesi, ci si rende maggiormente conto del trend in entrata ed uscita che affronta quotidianamente Adam.



[Grafico 6: messaggi ricevuti giorno per giorno, marzo-agosto 2005]



[Grafico 7: msg spediti (posta "interna" compresa) giorno per giorno, mar-ago 2005]



[Grafico 8: messaggi spediti verso l'esterno giorno per giorno, marzo-agosto 2005]

parte IV

CONCLUSIONI

Punto fermo che accompagna queste considerazioni finali è la consapevolezza documentata che *lo spam aggrava il carico di consegna del mailserver di un isp*.

Grazie alle "istantanee" fotografate si è visto come *i messaggi ritenuti spam siano quasi la metà dei messaggi ricevuti*. Che siano un numero considerevole lo dimostra anche la seconda parte dell'analisi, raffrontando il numero di messaggi ricevuti rispetto alla quantità della posta uscente.

Anche dimezzandoli, i messaggi in entrata risulterebbero di gran lunga superiori rispetto a quelli in uscita: ciò è probabilmente dovuto ai vari accodamenti di posta che Trivenet gestisce rispetto a singole caselle. Da non trascurare il fatto che comunque oggigiorno molti clienti sono autonomi nella spedizione della posta, disponendo di un server smtp in casa.

Sarebbe interessante oltreché importante provare ad inserire un soggetto "*gateway*" tra internet ed il mailserver, con l'unico compito di filtrare a monte tutta la posta: si eviterebbe così di appesantire Adam di ciò che contiene virus, alleggerendolo soprattutto dal marchiare i messaggi di spam.

Prima di affrontare questo studio prevedevo di suggerire l'implementazione di un secondo mailserver, destinato unicamente a spedire posta. Col senno di poi devo ricredermi, indicando che sarebbe più opportuno

concepire un server parallelo a quello esistente, capace solo di accogliere i messaggi in entrata.

Non riesco ancora a motivare la presenza delle *brevi ma saltuarie interruzioni del servizio smtp*. Interruzioni che non dovrebbero mai accadere proprio perché, come ripetuto più volte, un internet provider dovrebbe sempre garantire la continuità del servizio.

Ritengo lontana l'ipotesi di un attacco informatico verso la porta 25 del protocollo smtp. Suggerirò comunque alla direzione aziendale di installare in fase di test un *firewall software*, che permetta di tracciare ogni richiesta di traffico verso Adam, limitando il passaggio solo a chi è intenzionato a dare in consegna un messaggio di posta.

Oltre ad attendere gli *ultimi aggiornamenti da Deerfield*, che potrebbero coprire eventuali bachi a livello software oppure offrire migliorie al prodotto, potrei dotare Adam di una *seconda cpu* , considerando che le famigerate interruzioni erano spesso preannunciate da un preoccupante carico di lavoro della cpu esistente.

Per restringere il numero di "falsi negativi" (la cui intestazione riporto come esempio in Appendice) dovrei provare a *rendere più severo il filtro antispam*, rischiando però di incrementare la quantità di "falsi positivi" dovuti soprattutto alla negligenza di alcuni mittenti nell'osservanza di alcune fondamentali regole internazionali.

parte V

APPENDICE

Intestazione (header) di un FALSO NEGATIVO

X-Spam-Flag: **NO**
X-Spam-Status: No, hits = 4,35 required = 5,00
tests = HTML_MESSAGE, HTML_TAG_EXISTS_TBODY, BAYES_60, NO_RDNS2, J_CHICKENPOX_14, J_CHICKENPOX_34, J_CHICKENPOX_62, J_CHICKENPOX_73
X-Spam-Level: ****
X-Spam-Score: **4,35**
X-Spam-Checker-Version: SpamAssassin 3.0 (1.3)

Intestazione (header) di un FALSO POSITIVO

X-Spam-Flag: **YES**
X-Spam-Status: Yes, hits = 7,06 required = 5,00
tests = HTML_MESSAGE, HTML_TAG_EXISTS_TBODY, BAYES_99, NO_RDNS2, LONGWORD, MEGALONGWORD, J_CHICKENPOX_35
X-Spam-Level: *******
X-Spam-Score: **7,06**
X-Spam-Checker-Version: SpamAssassin 3.0 (1.3)
X-Spam-Reason: HTML

GLOSSARIO

Dns

Sistema di identificazione: risolve il nome di ogni terminale collegato ad internet in indirizzo ip. Ad esempio: tbox.trivenet.it = 212.103.192.15.

Falso negativo

Messaggio classificabile come spam, ma non catalogato come tale: succede ad esempio quando non si applicano tutte le opzioni del filtro antispam.

Falso positivo

Messaggio "sano", catalogato invece come spam per diversi motivi: ad esempio la formattazione del messaggio di una newsletter non rispettante determinati requisiti viene considerata spam.

Internet Service Provider (ISP)

Fornitore di servizi internet, tra i quali: connettività, domini, caselle di posta elettronica.

Pop3

Il pop3 è la versione più recente del protocollo pop, che consente all'utente di scaricare la posta sul proprio pc.

Smtp

“Simple Mail Transfer Protocol”: protocollo standard che regola il trasferimento dei messaggi e-mail da un server all'altro.

Spam

Così viene chiamata la cosiddetta posta “spazzatura”, o non desiderata.
Il termine trae origine da un vecchio sketch del “Monty Python's Flying Circus”, ambientato in un locale dove ogni pietanza del menù era a base di spam (un tipo di carne in scatola). Il continuo invocare tale piatto ricordò e quindi suggerì il nome da attribuire a questo fenomeno informatico. [Fonte: Wikipedia.org]

Caratteristiche dello spam:

- il messaggio è inviato ad un vasto numero di persone pubblicizzando un prodotto oppure un servizio;
- il destinatario del messaggio non ha richiesto e non vuole ricevere pubblicità dal mittente;
- il mittente dello spam spesso tenta di nascondere o di oscurare la propria identità;
- il mittente continua ad inviare e-mail indesiderate anche quando ha ricevuto la richiesta dei destinatari di bloccare l’invio.

[Fonte: “Spam: un problema per le aziende”, Windows&Net Magazine, Nov 2004]

Spamming

Mandare spam.

BIBLIO/SITOGRAFIA

Guide VisNetic, http://deerfield.com/support/visnetic-mailserver/documentation

"Introduzione alla statistica", Benito V. Frosini, La Nuova Italia Scientifica, 1988, Roma.

"Protezione per la vostra azienda – guide indispensabili per la sicurezza informatica. Guida uno: comprendere i problemi", http://www.watchguard.com

"Spam: un problema per le aziende", Windows&.NET Magazine, Nov. 2004.

"ICT Security Newsletter" del 6/9/2005, http://www.nstecna.com

Wikipedia, enciclopedia libera creata dagli internauti, http://www.wikipedia.org

LINK

http://www.deerfield.com/products/visnetic-mailserver

Il sito ufficiale di VisNetic Mailserver.

http://www.trivenet.it

Il sito istituzionale di Trivenet S.p.A.

http://www.evangelion.mangaitalia.it

La serie animata giapponese “Neon Genesis Evangelion”.

http://www.bboobb.it

Il mio sito personale.

Indirizzo e-mail: bboobb@bboobb.it