



MINISTERO DELLO SVILUPPO ECONOMICO  
DIREZIONE GENERALE PER LA LOTTA ALLA CONTRAFFAZIONE  
UFFICIO ITALIANO BREVETTI E MARCHI

DOMANDA NUMERO	102007901528789
Data Deposito	04/06/2007
Data Pubblicazione	04/12/2008

Sezione	Classe	Sottoclasse	Gruppo	Sottogruppo
H	04	M		

Titolo

METODO PER RILEVARE UN SINGOLO FLUSSO DATI ALL'INTERNO DI UN FLUSSO AGGREGATO DI DATI A PACCHETTI E PER IDENTIFICARE L'APPLICAZIONE GENERATRICE DEL SINGOLO FLUSSO DATI.

**Titolo: "Metodo per rilevare un singolo flusso dati all'interno di un flusso aggregato di dati a pacchetti e per identificare l'applicazione generatrice del singolo flusso dati"**

5

**DESCRIZIONE**

La presente invenzione riguarda un metodo per rilevare un singolo flusso dati all'interno di un flusso aggregato di dati a pacchetti e per identificare l'applicazione generatrice del singolo flusso dati"

10 Allo stato dell'arte, è noto il problema di rilevare un singolo flusso dati in un flusso dati a pacchetti e di identificare l'applicazione generatrice del flusso stesso, ad esempio identificare un singolo flusso voce e la applicazione che lo ha generato all'interno di un flusso o  
15 traffico aggregato su rete IP.

In particolare, tale problema è noto con riferimento alla telefonia VoIP in cui viene stabilita una comunicazione vocale su rete IP tra due utenze utilizzando protocolli non noti e cifrati. Un tipico esempio di software che genera  
20 flusso dati vocale su rete IP è Skype.

I protocolli e gli algoritmi che permettono a Skype, così come alla maggior parte dei programmi vocali, di generare flusso dati vocale su rete IP, sono sconosciuti e spesso criptati e si basano su cifratura del contenuto.

25 Per tale motivo risulta molto difficile rilevare la

Ing. Davide BONVICINI  
N. Iscriz. ALBO 1212 B  
(in proprio e per gli altri)

presenza di un singolo flusso dati generato da una particolare applicazione, quale ad esempio Skype, in un flusso dati aggregato comprendente flussi generati da applicazioni di vario tipo, sia vocali, sia di trasporto  
5 dati, sia di comunicazioni video, ecc..

Da quanto sopra esposto emerge l'esigenza di poter rilevare la presenza di un singolo flusso dati in un flusso aggregato di dati a pacchetti e di identificare l'applicazione generatrice del singolo flusso dati in  
10 assenza della conoscenza dei protocolli e degli algoritmi utilizzati dall'applicazione stessa per generare il singolo flusso dati e per includere tale singolo flusso dati nel flusso aggregato di dati a pacchetti.

In vista dello stato della tecnica descritto, scopo  
15 della presente invenzione è quello di realizzare un metodo per rilevare un singolo flusso dati in un flusso aggregato di dati a pacchetti ed identificare l'applicazione generatrice del singolo flusso dati in grado di superare gli inconvenienti presenti nella tecnica nota.

20 In accordo con la presente invenzione, tale scopo viene raggiunto da un metodo per rilevare un singolo flusso dati in un flusso aggregato di dati a pacchetti ed identificare l'applicazione generatrice del singolo flusso dati in accordo con la rivendicazione 1.

25 Grazie alla presente invenzione è possibile ottenere

un metodo per rilevare un singolo flusso dati in un flusso aggregato di dati a pacchetti ed identificare l'applicazione generatrice del singolo flusso dati su rete IP utilizzando una tecnica semplice.

5           Ulteriori caratteristiche ed i vantaggi del metodo per rilevare un singolo flusso dati in un flusso aggregato di dati a pacchetti ed identificare l'applicazione generatrice del singolo flusso dati secondo la presente invenzione risulteranno dalla descrizione di seguito riportata di un  
10 esempio preferito di realizzazione, data a titolo indicativo e non limitativo, con riferimento alle annesse figure, in cui:

- la figura 1 mostra uno schema di principio esplicativo del metodo per rilevare un singolo flusso dati  
15 in un flusso aggregato di dati a pacchetti ed identificare l'applicazione generatrice del singolo flusso dati in accordo con la presente invenzione,

- la figura 2 mostra distribuzioni di deviazioni di frequenza elaborate per blocchi di bit deterministici,  
20 casuali e misti.

Nel proseguo della presente descrizione si farà uso di funzioni statistiche di misura della deviazione di frequenza, in particolare della funzione chi-quadro di Pearson. Qui di seguito viene illustrato la funzione  
25 statistica chi-quadro di Pearson.

**Ing. Davide BONVICINI**  
N. Iscriz. ALBO 1212 B  
(in proprio e per gli altri)

La funzione chi-quadro di Pearson permette di verificare se il comportamento di un oggetto, osservato per un numero finito di volte, segue un comportamento atteso.

5        Ciò è realizzato calcolando la deviazione dei valori misurati dell'oggetto rispetto alla distribuzione attesa dei valori dell'oggetto.

Si assuma ad esempio di osservare un oggetto per un numero di volte  $N_{TOT}$  e che l'oggetto sotto osservazione  
10 possa assumere  $N$  possibili uscite o valori per ciascuna osservazione.

Se la distribuzione attesa dei valori è tale che il valore  $i$ , con  $i=0, \dots, N-1$ , ricorre con una probabilità  $p_i$ , allora il numero atteso di eventi o frequenza di  $i$  è dato  
15 dalla relazione  $E_i = N_{TOT} p_i$ . Chiamato  $O_i$  il numero di eventi o frequenza di  $i$  realmente osservati durante l'osservazione, si ha che il valore

$$\chi^2 = \sum_{i=0}^{N-1} \frac{(O_i - E_i)^2}{E_i}$$

20        rappresenta una misura della deviazione del comportamento osservato rispetto al comportamento atteso, ovvero della frequenza osservata rispetto alla frequenza attesa.

Se l'oggetto osservato si comporta così come atteso, allora il valore di  $\chi^2$  è distribuito secondo una

distribuzione chi-quadro con  $N-1$  gradi di libertà.

La funzione chi-quadro può essere impiegata anche per una singola osservazione. In particolare, si ipotizza che il valore dell'oggetto osservato sia distribuito con  
5 probabilità  $p_i$ .

Nella fattispecie di un flusso aggregato di dati a pacchetti, il flusso dati a pacchetti è generato da una specifica applicazione generatrice ed è suddiviso in messaggi, ciascun messaggio comprendendo una pluralità di  
10 blocchi  $g$ .

Ciascun blocco  $g$  della pluralità di blocchi ha  $n$  bit per identificare  $2^n$  valori di blocco  $i$ , ad esempio con  $i=0, 1, 2, \dots, 2^n-1$ .

Con riferimento alle annesse figure, il metodo per  
15 rilevare un singolo flusso dati in un flusso aggregato di dati a pacchetti ed identificare l'applicazione generatrice del singolo flusso dati comprende le fasi di:

- a) fornire, per ciascun valore di blocco  $i$ , un valore di frequenza atteso  $E_i$ ,
- 20 b) misurare, per un predefinito numero  $G$  di blocchi  $g$  della pluralità di blocchi, ovvero per  $Gg$  bits, i valori di frequenza  $O_i^g$  con cui ciascun blocco  $g$  assume ciascun valore di blocco  $i$  così da ottenere una pluralità di valori di frequenza misurati  $O_i^g$ ,
- 25 c) elaborare, per ciascun blocco  $g$ , i valori di

frequenza misurati  $O_i^g$  ed i valori di frequenza attesi  $E_i$ ,  
per generare un valore di deviazione di frequenza  $\chi_g^2$   
rappresentativo della deviazione della pluralità di valori  
di frequenza misurati  $O_i^g$  rispetto ai valori di frequenza  
5 attesi  $E_i$ ,

d) elaborare i valori di deviazione di frequenza  $\chi_g^2$   
generati per ciascun blocco  $g$  con almeno un valore soglia di  
deviazione di frequenza  $\chi_{th}$  per rilevare la presenza di un  
singolo flusso dati in detto flusso aggregato di dati a  
10 pacchetti ed identificare l'applicazione generatrice del  
singolo flusso dati.

Il singolo flusso dati può essere sia flusso vocale  
che flusso peer-to-peer (P2P).

In particolare, come sarà descritto in dettaglio nel  
15 seguito, la fase d) consente di determinare la sorgente  
generatrice del singolo flusso dati, ovvero l'applicazione  
utilizzata per generare il singolo flusso dati rilevato.

In accordo con una forma di realizzazione, la fase d)  
comprende le fasi di:

20 d1) elaborare i valori di deviazione di frequenza  $\chi_g^2$   
generati per ciascun blocco  $g$  per generare almeno un valore  
di deviazione di frequenza di riferimento  $\chi_{ref}$  per detto  
predefinito numero di blocchi  $G$ , e

d2) confrontare tali valori di deviazione di frequenza di  
25 riferimento generati  $\chi_{ref}$  con il valore soglia di deviazione

di frequenza  $\chi_{th}$  per determinare la sorgente generatrice del singolo flusso dati.

Secondo una forma di realizzazione, la fase c) comprende la fase di applicare la pluralità di valori di frequenza misurati  $O_i^g$  ed i valori di frequenza attesi  $E_i$ , ad una funzione di misura statistica della deviazione di frequenza.

In particolare, la funzione di misura statistica della deviazione di frequenza può essere scelta tra una delle  
10 funzione entropia, media, varianza, chi quadro e simili.

Nella fattispecie, scelta la funzione chi-quadro, espressa dalla seguente formula

$$\chi_g^2 = \sum_{i=0}^{2^n-1} \frac{(O_i^g - E_i)^2}{E_i}$$

in cui

15  $\chi_g^2$  corrisponde al valore di deviazione di frequenza  $\chi_g^2$ ,

$O_i^g$  corrisponde alla pluralità di valori di frequenza misurati  $O_i^g$ , e

$E_i$  corrisponde ai valori di frequenza attesi  $E_i$ .

20 I valori di frequenza attesi  $E_i$  possono essere ottenuti in funzione dell'applicazione che si vuole identificare oppure, in assenza di tale informazione a priori, possono essere distribuiti in modo uniforme.

Con riferimento alle figure allegate, nel seguito si

Ing. Davide BONVICINI  
N. Iscriz. ALBO 1212 B  
(in proprio e per gli altri)

descrive l'applicazione del metodo secondo l'invenzione per rilevare un singolo flusso dati generato da un'applicazione Voce su IP Skype all'interno di un flusso aggregato di dati a pacchetti ed identificare tale applicazione generatrice  
5 del singolo flusso dati.

Poiché Skype è un programma chiuso e proprietario che utilizza algoritmi di cifratura, non è possibile identificare un flusso dati generato da Skype con tecniche tradizionali di analisi dei contenuti dei pacchetti.

10 Tuttavia, vi è un'importante differenza sui messaggi immessi in rete in funzione del sottostante protocollo di trasporto utilizzato.

Ad esempio, il protocollo TCP implementa un protocollo di trasmissione orientato alla connessione e quindi  
15 garantisce che vengano ricevuti tutti i segmenti di dati nella stessa sequenza con cui vengono immessi nella rete, eventualmente con un ritardo.

Il servizio senza connessione di un collegamento offerto dal protocollo UDP invece non garantisce la consegna  
20 di tutti i dati e con la stessa sequenza con cui sono stati immessi.

Di conseguenza, un codificatore Skype non può criptare tutto il messaggio ma deve consentire al ricevitore Skype di estrarre dall'intestazione del livello di applicazione  
25 alcune informazioni aggiuntive per rilevare e gestire

Ing. Davide BONVICINI  
N. Iscriz. ALBO 1212 B  
(in proprio e per gli altri)

eventuali messaggi persi o consegnati fuori sequenza al ricevitore.

Tale informazione non può essere protetta mediante cifratura ma può essere solo offuscata in modo tale da  
5 essere agevolmente identificata in ricezione. Tale porzione del messaggio è chiamata Inizio del Messaggio o SoM (Start of Message).

Ad esempio, quando un messaggio è trasportato su protocollo TCP, l'intero contenuto del messaggio Skype è  
10 criptato e quindi i byte del messaggio casualmente assumono valori casuali. Al contrario, in caso di trasporto su UDP, solo una porzione del messaggio è distribuita casualmente mentre altre porzioni presentano proprietà statistiche tipiche di dati deterministici, ad esempio il SoM.

15 Il metodo sopra descritto permette di distinguere pertanto il singolo flusso di dati generato da applicazioni Skype rispetto a flussi di dati generati da altre applicazioni di generazione flusso di dati o voce su IP, in quanto tali applicazioni utilizzando differenti formati di  
20 intestazione, risultando in differenti distribuzioni dei byte dei messaggi.

Occorre quindi verificare se i valori di deviazione di frequenza  $\chi_g^2$  sono tali da soddisfare l'ipotesi attesa. Quale ipotesi attesa, si utilizzano le caratteristiche di  
25 contenuto del messaggio riassunte nella seguente tabella nel

caso di messaggi End-to-End (E2E) su UDP, End-to-Out (E2O) su UDP e End-to-End o End-to-Out su TCP, in cui End-to-End rappresenta traffico generato tra due host terminali, ciascuno dei quali utilizza un client Skype, mentre End-to-Out rappresenta traffico generato tra un host terminale ed un tradizionale terminale PSTN.

Modalità Skype	Inizio Messaggio (SoM)			Payload
	Posizione Byte 1-2	3	4	
E2E su UDP	Casuale	Misto	Casuale	Casuale
E2O su UDP	Deterministico	Deterministico	Deterministico	Casuale
E2E-E2O su TCP	Casuale	Casuale	Casuale	Casuale

**Tabella**

Ad esempio, il flusso E2E su UDP presenta i bytes 1,2 e 4 cifrati ovvero casuali mentre il byte 3 contiene alcuni bit casuali e alcuni bit costanti (misto nella tabella), così come i byte di inizio messaggio del flusso E2O su UDP assumo valori deterministici.

Al fine di determinare se un blocco ha distribuzione casuale, deterministica o mista, si considera quale distribuzione attesa, la distribuzione di bit uniformemente distribuiti. In questo caso il valore di frequenza atteso  $E_i$  è pari a  $N_{TOT}/2^n$  per tutti i valori di blocco  $i$ , in cui  $N_{TOT}$  è il numero di messaggi analizzati appartenenti al flusso.

Si confrontano perciò i valori di deviazione di

Ing. Davide BONVICINI  
N. Iscriz. ALBO 1212 B  
(in proprio e per gli altri)

frequenza generati  $\chi_g^2$  con una o più soglie derivate dalla distribuzione chi-quadro con  $2^n-1$  gradi di liberta. Tali soglie vengono indicate con  $\chi_{Rnd}^2$ ,  $\chi_{Mix}^2$  e  $\chi_{Det}^2$  rispettivamente per blocchi casuali, misti e deterministici.

5 I valori  $G$  del predefinito numero di blocchi e  $n$  numero di bit possono essere fissati ad esempio a  $n=4$  bit and  $G=16$ . In questo caso, si ha che la distribuzione chi-quadro di riferimento ha  $2^n-1=15$  gradi di liberta e  $E_i=N_{TOT}/16$  per tutti i valori di blocco  $i=0, \dots, 15$ .

10 Il confronto tra i valori di deviazione di frequenza generati  $\chi_g^2$  ed i valori di deviazione di frequenza di riferimento  $\chi_{Rnd}^2$ ,  $\chi_{Mix}^2$  e  $\chi_{Det}^2$  è realizzato ad esempio come segue:

- E2E su UDP

$$15 \quad \max_{g \in G'} \chi_g^2 < \chi_{Rnd}^2 \wedge \min_{g \in \{5,6\}} \chi_g^2 > \chi_{Mix}^2$$

dove,

$G' = \{g | 1 \leq g \leq G, g \neq 5, 6\}$  sono i blocchi  $g$  corrispondenti

alla parte casuale del messaggio E2E,

$\max_{g \in G'} \chi_g^2$  è un primo valore di deviazione di frequenza di

20 riferimento generato,

$\min_{g \in \{5,6\}} \chi_g^2$  è un secondo valore di deviazione di frequenza

di riferimento generato, e

$\chi_{Rnd}^2$  e  $\chi_{Mix}^2$  due valori di soglia di deviazione di

frequenza.

25 In sostanza, ci si aspetta che i blocchi  $g$  con

distribuzione casuale abbiano distribuzione uniforme  
 cosicché i valori di deviazione di frequenza generati  $\chi_g^2$   
 devono essere relativamente piccoli e quindi minori del  
 valore di soglia di deviazione di frequenza  $\chi_{Rnd}^2$ , e che i  
 5 blocchi g con distribuzione mista che contengono alcuni  
 blocchi deterministici abbiano valori di deviazione di  
 frequenza generati  $\chi_g^2$  elevati e quindi maggiori del valore  
 di soglia di deviazione di frequenza  $\chi_{Mix}^2$ .

- E2O su UDP

$$10 \quad \min_{g=1,\dots,8} \chi_g^2 > \chi_{Det}^2 \wedge \max_{g=9,\dots,16} \chi_g^2 < \chi_{Rnd}^2$$

In questo caso, ci si aspetta che l'inizio del  
 messaggio SoM, ovvero i primi 4 byte cioè g=8 blocchi di n=4  
 bits siano deterministici e che la restante porzione sia  
 casuale, essendo tutto il messaggio cifrato.

15 - E2E-E2O su TCP

$$\max_{g=1,\dots,16} \chi_g^2 < \chi_{Rnd}^2$$

In questo caso, ci si aspetta che tutti i blocchi di  
 bit abbiano distribuzione casuale.

Vantaggiosamente, il numero di messaggi appartenenti  
 20 al flusso  $N_{TOT}$  è grande. Ad esempio, il numero  $N_{TOT}$  è tale  
 per cui il valore di frequenza atteso  $E_i \geq 5$  per tutti i  
 valori di blocco i. Nell'esempio qui riportato, ciò equivale  
 a dire che  $\frac{N_{TOT}}{2^n} \geq 5$ , ovvero che  $N_{TOT} \geq 80$  con n=4 bit.

Giova altresì rilevare che la differenza tra i valori  
 25 di deviazione di frequenza generati  $\chi_g^2$  per un blocco g

deterministico o casuale aumenta in funzione del valore del numero di messaggi appartenenti al flusso  $N_{TOT}$ .

Per un blocco deterministico  $g$  si ha che

$$\begin{aligned} \chi_g^2 &= \sum_{i=0}^{2^n-1} \frac{(O_i^g - E_i)^2}{E_i} = \\ &= \frac{(N_{TOT} - E)^2 + (2^n - 1)E^2}{E} = \\ &= N_{TOT}(2^n - 1). \end{aligned}$$

Pertanto  $\chi_g^2$  cresce sostanzialmente linearmente con  $N_{TOT}$  cosicché maggiore è la lunghezza del flusso, maggiore è  $N_{TOT}$  e maggiore è l'attendibilità che il blocco  $g$  sia deterministico, ovvero superi il valore di soglia di riferimento  $\chi_{Det}^2$ .

Nel caso di un blocco  $g$  misto, se un bit è fisso e gli altri hanno distribuzione casuale,  $O_i = 0$  per metà dei possibili valori di blocco  $i$ , e  $O_i > 0$  per i restanti valori di blocco  $i$ . Poiché i possibili valori di  $i$  sono  $2^n$ , il valore di deviazione di frequenza generato  $\chi_g^2$  è

$$\begin{aligned} \chi_g^2 &= \sum_{i=0}^{2^n-1} \frac{(O_i^g - E_i)^2}{E_i} = \\ &= \sum_{i=1}^{2^{n-1}} \frac{(O_i^g - E)^2}{E} + 2^{n-1} \frac{E^2}{E} = \\ &= 2\chi_{2^{n-1}-1}^2 + N_{TOT}. \end{aligned}$$

In cui  $\chi_{2^{n-1}-1}^2$  è la funzione chi-quadro con  $2^{n-1}-1$  gradi di libertà. In altre parole,  $\chi_{2^{n-1}-1}^2$  è un valore che può essere ricavato da un'osservazione di bit casuali con  $2^{n-1}$

possibili valori di bit, anziché  $2^n$  possibili valori.

Ciò significa che nel caso di un blocco  $g$  con un bit deterministico,  $\chi_g^2$  aumenta ancora linearmente con  $N_{TOT}$ .

In figura 2 sono riportati i valori deviazione di  
5 frequenza generati  $\chi_g^2$  per blocchi di bit misti, casuali e deterministici su flussi identificati quali flussi Skype. Si può rilevare come  $\chi_g^2$  cresca linearmente con  $N_{TOT}$  sia per blocchi deterministici che per blocchi misti, laddove per blocchi completamente deterministici ha una deviazione di  
10 frequenza maggiore rispetto ai blocchi misti. In figura 2 sono riportati anche i valori  $\chi_g^2$  assunti da blocchi casuali che non dipendono da  $N_{TOT}$ . Dalla figura 2 si evince blocchi misti, deterministici e casuali possono essere distinti tra loro in funzione dei valori di deviazione di frequenza  
15 generati  $\chi_g^2$  e che i valori di soglia di deviazione di frequenza non sono parametri critici per tale identificazione.

Nell'esempio, al fine di ridurre il numero di parametri, si può impostare  $\chi_{Rnd}^2 = \chi_{Mix}^2 = \chi_{Det}^2 = 150$ .

20 Vantaggiosamente, il metodo della presente invenzione può essere utilizzato in combinazione con il metodo per rilevare flusso dati vocale in un flusso di dati a pacchetti descritto nella domanda di brevetto italiano n. MI 2006 A 002417 qui incorporata per riferimento.

25 In breve, il metodo della domanda di brevetto italiano

Ing. Davide BONVICINI  
N. Iscriz. ALBO 1212 B  
(in proprio e per gli altri)

MI 2006 A 002417 prevede che il flusso di dati a pacchetti sia caratterizzabile da almeno due variabili misurabili  $X, Y$  e fornisce, per ciascuna variabile misurabile  $X, Y$ , una funzione  $P\{x|C\}, P\{y|C\}$  di distribuzione dei valori di  
5 ciascuna variabile  $X, Y$  in un flusso dati vocale. Successivamente, si misurano i valori  $x, y$  di ciascuna variabile  $X, Y$  per ottenere una sequenza di valori misurati  $x^{(k)}, y^{(k)}$  su un numero  $K$  di blocchi e si applica ciascun valore misurato  $x^{(k)}, y^{(k)}$  alla rispettiva funzione di  
10 distribuzione  $P\{x|C\}, P\{y|C\}$  per generare una sequenza di valori di verosimiglianza  $B_x^{(k)}, B_y^{(k)}$  dai quali si generano rispettivi valori medi di verosimiglianza  $E[B_x], E[B_y]$ . Infine, tali valori medi vengono elaborati per generare un valore di verosimiglianza di riferimento  $B$  che confrontato  
15 con un valore di verosimiglianza di soglia  $B_{\min}$  permette di rilevare la presenza di flusso dati vocale nel flusso dati a pacchetti.

Da esperimenti svolti, è emerso che l'utilizzo congiunto del metodo descritto nella domanda di brevetto  
20 italiano MI 2006 A 002417 e del metodo della presente invenzione sono estremamente efficaci nel rilevare e classificare qualunque traffico voce su IP e nel rilevare e classificare traffico voce generato da un'applicazione Skype e trasportato sia su UDP sia su TCP.

25 E' inoltre stato dimostrato che i due metodi sopra

Ing. Davide BONVICINI  
N. Iscriz. ALBO 1212 B  
(in proprio e per gli altri)

citati presentano un elevato grado di robustezza.

Come si può apprezzare da quanto descritto, il metodo secondo la presente invenzione consente di soddisfare le esigenze e di superare gli inconvenienti di cui si è  
5 riferito nella parte introduttiva della presente descrizione con riferimento alla tecnica nota.

In particolare, il metodo secondo l'invenzione consente di rilevare la presenza di un qualunque tipo di flusso vocale, anche criptato.

10 Ovviamente, un tecnico del ramo, allo scopo di soddisfare esigenze contingenti e specifiche, potrà apportare numerose modifiche e varianti al metodo secondo l'invenzione sopra descritta, tutte peraltro contenute nell'ambito di protezione dell'invenzione quale definito  
15 dalle seguenti rivendicazioni.

\*\*\* \*\*

Ing. Davide BONVICINI  
N. Iscriz. ALBO 1212 B  
(in proprio e per gli altri)

**RIVENDICAZIONI**

1. Metodo per rilevare un singolo flusso dati in un  
flusso aggregato di dati a pacchetti ed identificare  
l'applicazione generatrice di detto singolo flusso dati,  
5 detto flusso aggregato di dati a pacchetti essendo suddiviso  
in messaggi, ciascun messaggio comprendendo una pluralità di  
blocchi, ciascun blocco (g) di detta pluralità di blocchi  
avendo n bit per identificare  $2^n$  possibili valori di blocco  
(i), detto metodo essendo caratterizzato dal fatto di  
10 comprendere le fasi di:

a) fornire, per ciascun valore di blocco (i), un valore  
di frequenza atteso ( $E_i$ ),

b) misurare, per un predefinito numero (G) di blocchi  
(g) di detta pluralità di blocchi, i valori di frequenza  
15 ( $O_i^g$ ) con cui ciascun blocco (g) identifica ciascun valore  
di blocco (i) così da ottenere una pluralità di valori di  
frequenza misurati ( $O_i^g$ ),

c) elaborare, per ciascun blocco (g), detta pluralità  
di valori di frequenza misurati ( $O_i^g$ ) ed i valori di  
20 frequenza attesi ( $E_i$ ) per generare un valore di deviazione  
di frequenza ( $\chi_g^2$ ) rappresentativo della deviazione della  
pluralità di valori di frequenza misurati ( $O_i^g$ ) rispetto ai  
valori di frequenza attesi ( $E_i$ ),

d) elaborare i valori di deviazione di frequenza ( $\chi_g^2$ )  
25 generati per ciascun blocco (g) con almeno un valore soglia

Ing. Davide BONVICINI  
N. Iscriz. ALBO 1212 B  
(in proprio e per gli altri)

di deviazione di frequenza ( $\chi_{th}$ ) per rilevare la presenza di detto singolo flusso dati in detto flusso aggregato di dati a pacchetti ed identificare l'applicazione generatrice di detto singolo flusso dati.

5           **2.** Metodo in accordo con la rivendicazione 1, in cui detta fase d) comprende le fasi di:

d1) elaborare i valori di deviazione di frequenza ( $\chi_g^2$ ) generati per ciascun blocco (g) per generare un valore di deviazione di frequenza di riferimento ( $\chi_{ref}$ ) per detto

10 predefinito numero di blocchi (G),

d2) confrontare detto valore di deviazione di frequenza di riferimento generato ( $\chi_{ref}$ ) con detto almeno un valore soglia di deviazione di frequenza ( $\chi_{th}$ ) per identificare detta applicazione generatrice di detto singolo flusso dati.

15           **3.** Metodo in accordo con la rivendicazione 1 o 2, in cui detta fase c) comprende la fase di applicare la pluralità di valori di frequenza misurati ( $O_i^g$ ) ed i valori di frequenza attesi ( $E_i$ ) ad una funzione di misura statistica della deviazione di frequenza.

20           **4.** Metodo in accordo con la rivendicazione 3, in cui detta funzione di misura statistica della deviazione di frequenza è scelta tra una delle funzione entropia, media, varianza, chi quadro.

**5.** Metodo in accordo con la rivendicazione 3 o 4, in  
25 cui detta funzione di misura statistica della deviazione di

frequenza è la funzione chi-quadro:

$$\chi_g^2 = \sum_{i=0}^{2^n-1} \frac{(O_i^g - E_i)^2}{E_i}$$

in cui

$\chi_g^2$  corrisponde a detto valore di deviazione di  
5 frequenza ( $\chi_g^2$ ),

$O_i^g$  corrisponde a detta pluralità di valori di  
frequenza misurati ( $O_i^g$ ),

$E_i$  corrisponde a detti valori di frequenza attesi  
( $E_i$ ).

Ing. Davide BONVICINI  
N. Iscriz. ALBO 1212 B  
(in proprio e per gli altri)

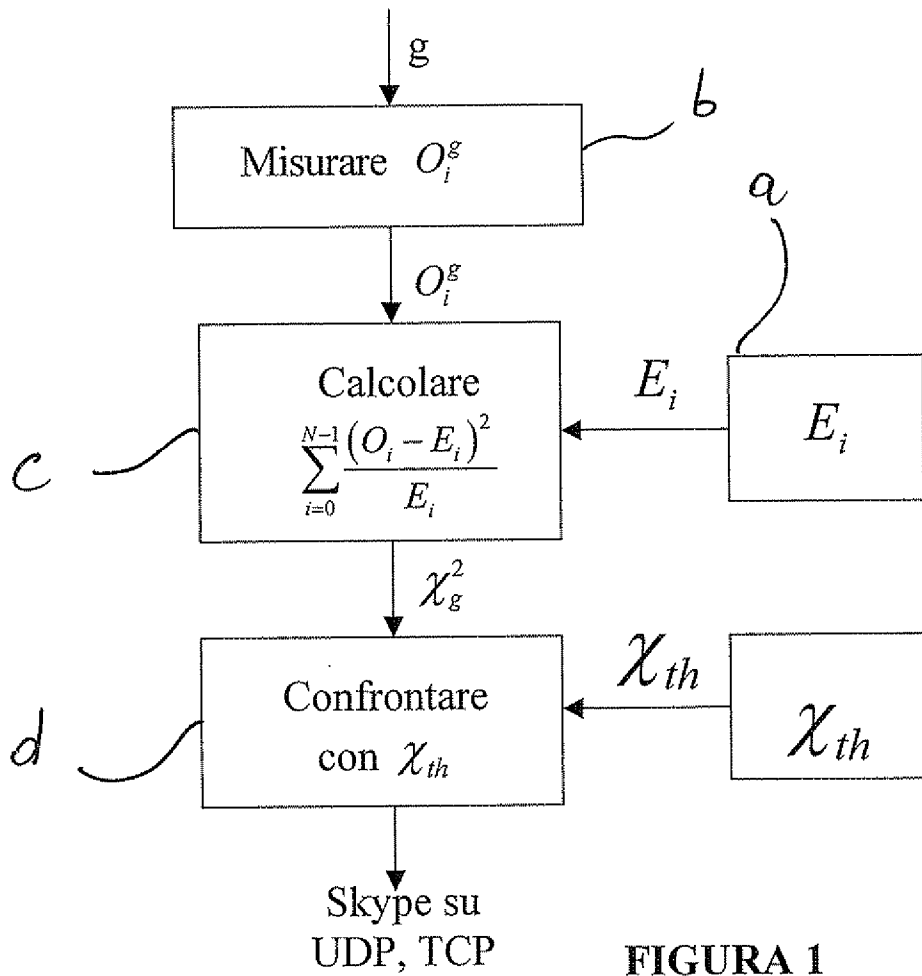


FIGURA 1

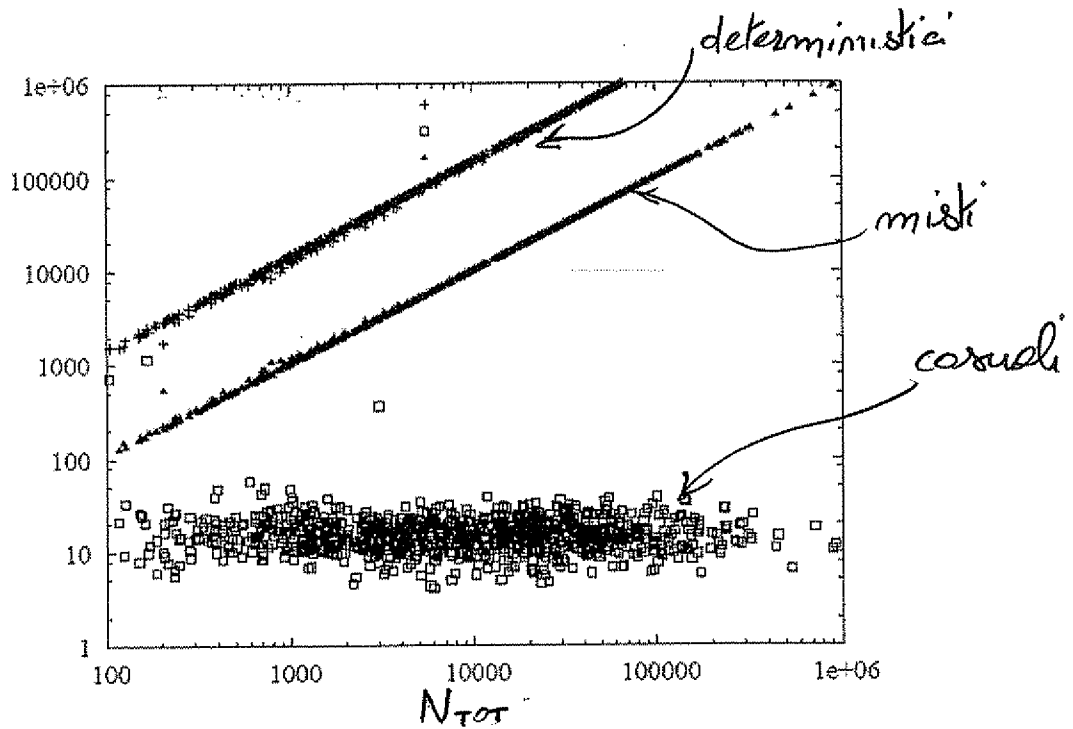


FIGURA 2