

Progetto del Parco Scientifico della Liguria

misura 3.7, sottomisura D) dal titolo

“Studio di fattibilità per la realizzazione di uno strumento per l’analisi e la classificazione con tecniche avanzate e scalabili di traffico IP su linee ad alta velocità”

(Pos. N. 46 Avviso 2/2006)

Executive Summary

L’obiettivo del progetto è stato lo studio la definizione e la verifica sperimentale di tecniche di classificazione di traffico IP a livello di applicazione su linee ad alta velocità. La sostanziale totalità delle reti di telecomunicazioni attuali sta convergendo sulla tecnologia Internet per la fornitura di tutti i servizi. Fulcro di tale tecnologia è il protocollo IP che ad oggi veicola la maggior parte degli scambi di informazioni esistenti, dai dati alla voce e al video. I flussi di dati che viaggiano su Internet e nelle Intranet aziendali sono normalmente generati da scambi fra coppie di applicazioni che forniscono agli utenti (umani o meno) servizi di natura diversa. In molti contesti è utile, se non fondamentale, poter riconoscere i pacchetti IP appartenenti a uno stesso dialogo e associarli alla tipologia di applicazione che li ha generati. Tale operazione non è necessariamente semplice per due ragioni principali:

- 1) sempre più spesso le applicazioni tendono a “nascondersi”, esplicitamente o implicitamente, e quindi a non fornire esplicitamente indicazioni sulla natura del flusso;
- 2) su linee ad alta capacità (sempre più diffuse) i volumi di pacchetti scambiati sono molto elevati e quindi l’associazione deve essere fatta in modo molto efficiente, se si vuole operare (come accade nella maggior parte dei contesti interessanti) in tempo reale.

Partendo da questi presupposti il progetto ha perseguito i seguenti obiettivi principali:

- definizione di tecniche avanzate di classificazione del traffico IP rispettando i seguenti requisiti generali:

- capacità di classificazione del traffico TCP/IP fino a livello di applicazione;
 - scalabilità delle prestazioni, che permetta di operare su linee ad “alta velocità” (fino a 100 Mbps per soluzioni su hardware standard, fino a 1 Gbps e oltre per soluzioni con hardware dedicato);
 - integrazione con funzionalità hardware *custom* (FPGA, etc.) che siano in grado di incrementare la capacità di scalare dei meccanismi proposti a velocità elevate;
- definizione di un’architettura hardware e software integrata per la realizzazione di un apparato di classificazione modulare ad alte prestazioni;
 - realizzazione di un prototipo su architettura PC e basato su software *open source* di un classificatore che usi almeno una delle tecniche selezionate;
 - verifica delle funzionalità e delle prestazioni del meccanismo di classificazione mediante il prototipo.

Le tecniche selezionate come più efficaci sono state tre: il *pattern matching*, le Reti Neurali (RN) e le *Support Vector Machine* (SVM). La prima tecnica è quella utilizzata negli approcci più tradizionali; in questo contesto è stata affinata opportunamente, sia dal punto di vista della precisione che della efficienza ed è stata implementata nel prototipo e verificata sperimentalmente. Questa tecnologia è risultata essere importante perché rimane quella che, in assenza di traffico cifrato o di tipologie di traffico totalmente ignote a priori, permette una identificazione molto puntuale e precisa del traffico. I livelli di precisione che si riescono a raggiungere con tale tecnica sono stati accuratamente verificati e, in sintesi, risultano essere, con la specifica realizzazione fatta nel progetto, nell’ordine del 87% dei flussi identificati correttamente.

Sia le tecniche di tipo neurale che quelle SVM permettono una classificazione meno dettagliata, ossia sono in grado di distinguere categorie di applicazioni più che specifiche applicazioni, ma operano su parametri caratteristici del traffico (dimensione dei pacchetti e tempi di interarrivo) che sono invarianti alla presenza di “manipolazioni” dei *payload*, tipo la cifratura. Inoltre sono in grado di riconoscere anche protocolli e applicazioni non note a priori purché appartenenti a una classe nota (ad esempio un nuovo tipo di traffico video che usa codifiche e protocolli diversi da quelli usuali). Queste tecnologie sono risultate precise e sono in grado di riconoscere l’appartenenza di un flusso ad una classe con una precisione media di circa 87% (RN) e 95% (SVM).

Per permettere a questi meccanismi di scalare anche sulle velocità molto elevate, si è progettata una scheda acceleratrice HW basata su FPGA, per la quale si è studiata una specifica funzionalità di filtraggio che permette di eliminare dai flussi in ingresso al classificatore i pacchetti successivi all’ennesimo (soglia configurabile). Infatti il classificatore è in grado di identificare un flusso e la corrispondente applicazione generante osservando solo i primi pacchetti del flusso stesso. I pacchetti successivi hanno il solo effetto di rallentare il classificatore senza aggiungere informazione utile. Questo filtro appartiene ad una tipologia particolare, detta *Bloom filter*, molto adatta ad una realizzazione

firmware. Nel corso del progetto è stata realizzata una forma prototipale di scheda di filtraggio tramite la quale si è potuto verificare che questa tecnica può operare su flussi a 1 Gbps ed oltre di velocità e che l'efficacia del filtro stesso è elevata.

Nel corso del progetto è stato sviluppato infine un prototipo di classificatore, consistente in un software, che opera all'interno della piattaforma "Clik Modular Router", in grado di funzionare su una architettura PC standard. Tale software è modulare, realizza la cattura dei pacchetti, l'identificazione dei flussi e l'associazione del flusso a una applicazione o ad una classe di applicazioni in tempo reale. I meccanismi di classificazione realizzati sono il *pattern matching*, esteso per riconoscere anche flussi RTP (video e audio), e un classificatore basato su reti neurali. Il prototipo ha mostrato come le tecnologie studiate rispondono pienamente ai requisiti previsti dal progetto. Infatti tale prototipo è in grado di operare in tempo reale su interfacce con capacità di 1 Gbps sottoposte a carichi fino a 300 Mbps senza avvalersi del filtraggio hardware.