HARDWARE ACCELERATION OF PROTOCOL IDENTIFICATION

Petr Kobierský
Master Degree Programme (2), FIT BUT
E-mail: xkobie00@stud.fit.vutbr.cz

Supervised by: Jan Kořenek
E-mail: korenek@fit.vutbr.cz

ABSTRACT

A dynamic grow of computer networks encourage rapid development of network applications and services. For providing sufficient network service quality is important to shape network flows based on their application protocol type. This work analyse current best protocol identification methods for use on multigigabit networks. Based on analysis the hardware architecture, which accelerate computationally intensive algorithms parts, is proposed. Proposed solution is able to work on 10 Gb/s networks and export application protocol type using Netflow protocol.

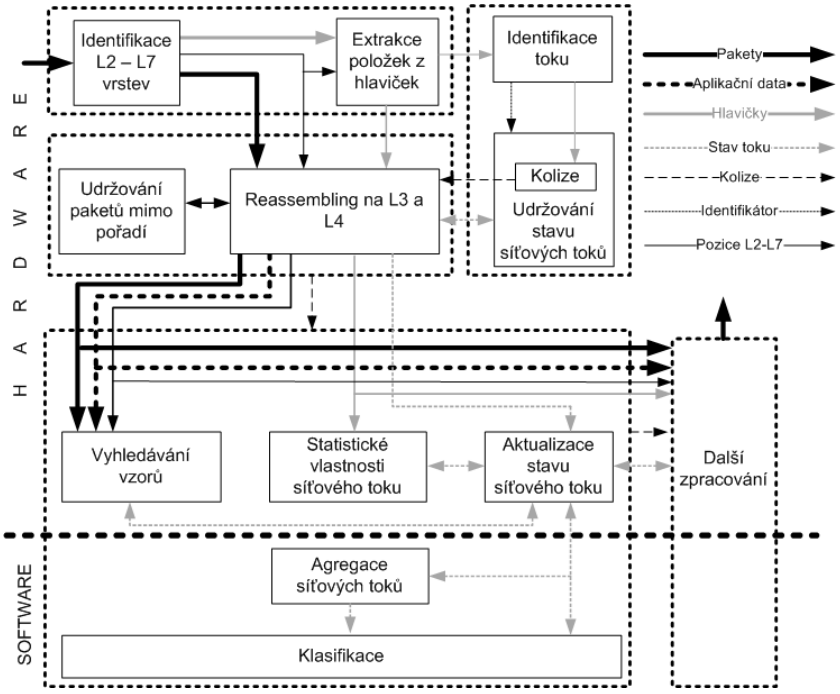
1 ÚVOD

Internet v současné době využívá stále větší množství služeb. Různé služby kladou na síť rozdílné požadavky, které musí být sítí zajištěny. Některé služby mohou naopak představovat pro danou síť vysokou zátěž a proto musí být omezeny nebo případně zakázány (typicky P2P služby pro sdílení obsahu). Pro zajištění různé kvality služeb je nutné v síťových tocích identifikovat aplikační protokoly, což vzhledem k jejich různým typům (šifrované, otevřené a uzavřené standardy) představuje zvlášť na multigigabitových sítích vážný problém.

Klasické metody identifikace provozu založené na číslech portů jsou vzhledem k dynamicky voleným portům a tunelování aplikací nedostatečné. Proto se používají přístupy založené na identifikaci aplikačních protokolů s využitím vyhledávaní signatur (regulární výrazy) v aplikačních datech. V řadě případů i tyto metody selhávají a je vhodné využít metody založené na statistických vlastnostech síťových toků. Pro identifikaci P2P sítí je vhodné informace o síťových tocích agregovat pro jednotlivé koncové stanice a následně je použít k identifikaci provozu.

Žádná z uvedených metod není dokonalá a hodí se pouze pro identifikaci určitých druhů provozu. S využitím kombinace těchto metod, je ale možné dosáhnout velmi vysoké přesnosti pro všechny typy provozu. Bohužel některé činnosti, které je v tomto případě potřeba řešit jsou výpočetně velmi náročné. Tato práce analyzuje výpočetní náročnost dílčích činností identifikace aplikačních protokolů a rozděluje je mezi hardware a software. Na základě tohoto rozdělení jsou vybrané činnosti mapovány na technologii FPGA, s cílem akcelerovat současné nejlepší metody identifikace aplikačních protokolů na 10 Gbit sítě.

2 MODEL IDENTIFIKACE APLIKAČNÍCH PROTOKOLŮ



Obrázek 1: Model identifikace aplikačních protokolů

Na základě analýzy existujících metod identifikace aplikačních protokolů byl vytvořen obecný model, který v sobě zahrnuje nejlepší známé metody identifikace protokolů. Tento model byl sestaven z důvodu analýzy výpočetní náročnosti jednotlivých činností prováděných v rámci identifikace aplikačních protokolů. Vytvořený model znázorněný na obrázku 1 se skládá z několika kooperujících bloků, které řeší zpracování nižších vrstev protokolů, udržování stavových informací k síťovým tokům, reassembling na L3 a L4 vrstvě, identifikaci síťového provozu a libovolnou uživatelskou aplikaci.
Simulací modelu byly identifikovány činnosti, které omezují současné použití všech známých metod identifikace aplikačních protokolů na sítě do 1 Gb/s. Jedná se zejména o softwarové vyhledávání regulárních výrazů (propustnost okolo 300 Mb/s) a sbírání statistických informací o síťových tocích (propustnost do 1 Gb/s). Získávání agregovaných statistik a výsledná klasifikace již pracuje s informacemi o síťových tocích. Tento datový tok však představuje průměrně 20x méně dat než paketový tok, proto je možné tyto činnosti s využitím vhodných klasifikačních algoritmů (např. rozhodovací stromy) realizovat na obecném procesoru.

3 AKCELERACE IDENTIFIKACE APLIKAČNÍCH PROTOKOLŮ

Vytvořený model byl namapován na technologii FPGA s využitím existující architektury [1] sondy pro sběr Netflow statistik. Dále byly vybrány statistiky vhodné k statistické identifikaci aplikačních protokolů (průměrná mezipaketová mezera, průměrná délka paketu, doba trvání síťového toku a další). Vybranou Netflow sondu lze o sběr těchto statistik relativně snadno rozšířit pomocí změny v XML schématu, ze kterého se generuje odpovídající hardware. Stávající

architekturu sondy bohužel nelze využit k identifikaci protokolů na základě vyhledávání signatur, protože sonda pracuje pouze na L2-L4 vrstvě modelu ISO/OSI. Z tohoto důvodu byla sonda rozšířena o jednotku pro vyhledávání regulárních výrazů v aplikačních datech.
Regulární výrazy pro identifikaci jednotlivých aplikačních protokolů byly použity z projektu L7-filter [2], který obsahuje více než 100 signatur aplikačních protokolů. Některé z těchto signatur byly přepsány do formátu PCRE a převedeny [3] na NKA. Výsledný automat lze poměrně jednoduchým způsobem mapovat na logiku FPGA a následně použít k vyhledávání vzorů [4]. Pokud se ale tento automat převede do formátu, kdy je s každým přechodem přijímáno více znaků (lazy FSM) lze s technologií FPGA dosáhnout propustnosti vyhledávání až 10 Gb/s.
Do výsledné architektury nebyla zařazena jednotka pro skládání toků na L3-L4 vrstvě. Tato činnost nelze na síťových prvcích uspokojivě řešit a také je velmi náročná na hardwarovou realizaci. Současné FPGA implementace umožňují spojovat pouze jeden paket mimo pořadí. Rovněž existují metody jak lze celkem jednoduše tyto jednotky obejít (např. vhodné nastavení TTL). Vzhledem k tomu, že většinu aplikačních protokolu lze identifikovat s prvním 1-4 kB dat, absence této operace nepředstavuje výrazný nárůst neidentifikovatelného provozu.
Získané statistiky a nalezené signatury jsou exportovány přes sběrnici k SW zpracování. Informace o tocích jsou dále agregovány a použity k získání statistik k jednotlivým koncovým stanicím (počet spojení, počet TCP a UDP relací). Nalezené reg. výrazy, statistiky o síťových tocích a agregované statistiky jsou použity jako vstup klasifikace.
Po úspěšné klasifikaci je typ aplikačního protokolu předán spolu s informacemi o síťovém toku ve formě standardizovaného Netflow v.9 záznamu k dalšímu zpracování. Záznamy jsou posílány na kolektor, kde mohou být vizualizovány nebo použity k nastavení routeru/firewallu a následné prioritizaci/omezení určitého typu provozu.

4 ZÁVĚR

Na základě analýzy nejnovějších přístupů identifikace aplikačních protokolu byl vytvořen model pokrývající současné metody identifikace. Vytvořený model byl využit k nalezení výpočetně náročných částí algoritmu identifikace provozu. Tyto části byly mapovány na technologii FPGA a díky tomu bylo možné navýšit propustnost metod identifikace aplikačních protokolů z 300 Mb/s na 10 Gb/s. Vytvořená architektura síťové sondy umožňuje exportovat informace o síťových tocích a detekovaných aplikačních protokolech ve formě Netflow protokolu. Tyto informace lze dále použít k omezení a blokování síťových toků, pro plánování propustnosti linek a serverů nebo také v systémech IDS.

REFERENCE

[1] Kořenek J., Žádník M., Kobierský P., Lengál O.: Network Probe for Flexible Flow Monitoring. accepted to IEEE DDECS 2008 Workshop

[2] WWW stránka projektu L7-filter, http://l7-filter.sourceforge.net (únor 2007)

[3] Hank A.: Detekcia narušenia počitačovej siete. Bakalářká práce, FIT VUT v Brně, 2007

[4] Kořenek J., Kobierský P.: Intrusion Detection System Intended for Multigigabit Networks. IEEE DDECS 2007 Workshop: 361-364