様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成　３０　年　　６　月　　５　日現在

機関番号：　１２６０１
研究種目：　基盤研究(C)（一般）
研究期間：　2014～2017
課題番号：　２６３３０１００

研究課題名（和文）帯域スケールアウト可能なネットワークアーキテクチャ

研究課題名（英文）A bandwidth scale out capable network architecture

研究代表者
小林　克志（Kobayashi, Katsushi）

東京大学・大学院情報理工学系研究科・特任准教授

研究者番号：９０２５１７１９

交付決定額（研究期間全体）：（直接経費）　　3,700,000円

研究成果の概要（和文）：インターネット帯域需要の増大に応えるため、構成部品を追加するだけでルータの帯域性能を増大(スケールアウト)できる並列化に対応したネットワークアーキテクチャとそれに必要な要素技術の実現性を検討した。
このアーキテクチャでは異なるパケット転送エンジンを経由することで発生するパケット順序交代が大きな課題となる。これを抑えるパケットスケジューリング方式として、既存の先着順(First Come First Served) に代えて、廃棄付き Earliest Deadline First を提案した。さらにその影響について評価し、既存トランスポート（TCP）への影響は軽微であることを示した。

研究成果の概要（英文）：To meet the Internet bandwidth growth, a scale-out network architecture and its elementary technologies
which performance can be increased with just adding components are discussed. On this architecture, the packet re-order due to more than one forwarding engines is one of the most significant issues. To prevent this, we proposed Earliest Deadline First with reneging packet scheduler which replaces ordinary First Come First Served one. We also demonstrated the impact to TCP transport of replacing the packet scheduler is an insignificant

研究分野：　ネットワークアーキテクチャ

キーワード：　インターネット　ルータ　パケットスケジューラ

１．研究開始当初の背景

インターネットでは年率 25%程度 (5 年間で 3 倍) トラフィック増大が見込まれ、これを支えるネットワーク基盤が求められている。半導体などの制約から、この基盤を構成するルータ・回線の並列分散化技術は欠かせない。しかしながら、ルータ・回線の並列分散化には解決すべき課題が存在している。第一の問題はコントロールプレーンの問題で、分散配置された並列コンポーネントを統合的に動作させる技術の確立である。第二はデータプレーンの問題である。我々は、スケールアウト的に性能向上を達成する並列分散ネットワークアーキテクチャとして、FAIN (Flexible Arrays of Inexpensive Networks) を提案してきた。FAIN のデータプレーンでは異なる転送エンジンに起因するパケットの転送遅延の差が避けられない。この遅延差に起因するパケット順序交替の結果発生する TCP およびアプリケーション品質の劣化である。

２．研究の目的

本研究の目的は、ルータ・回線を構成する並列コンポーネントの追加で帯域性能を向上させるスケールアウト可能なネットワークアーキテクチャの実現可能性を実践的なアプローチで示すことにある。背景で述べた二つの課題のうち、前者のコントロールプレーンに関してはここ数年で SDN (Software Defined Network) の制御技術が成熟し、課題の多くは解消されつつある。このため本研究では、後者の並列化データプレーンについて解決をめざす。

転送遅延の問題を解決する方法として、我々はパケットヘッダに許容できる遅延時間を埋め込み、残り時間の厳しいパケットから優先的に転送する Earliest Deadline First (EDF) ベースのパケットスケジューラを提案している。このスケジューラを従来の先着順、First Come First Served(FCFS)、に代え遅延差を解消する。

３．研究の方法

パケットスケジューラとして、EDF に加え許容遅延を超えたパケットを廃棄する、EDF with reneging (EDFR) を利用する。このスケジューラの有効性・実現性を２つのアプローチで検討する。

第一のアプローチは EDFR スケジューラを既存の FCFS で置き換えることで上位層、とくに TCP トランスポート、に与える影響を評価し、置き換えの可否を検討する。第二は EDF スケジューラのハードウェアによる実現性を FPGA 実装によって評価する。

４．研究成果

ここでは、(1) EDFR スケジューラが上位層に与える影響、(2)EDFR スケジューラのハードウェア実装についての研究成果について述べる。

(1) EDFR スケジューラが上位層に与える影響

①EDFR スケジューラの廃棄特性

パケットスケジューラ置き換えによる廃棄特性の変化は、TCP および上位アプリケーションの振る舞いを変えてしまう。なぜならば、TCP は廃棄を輻輳とみなし送出帯域を抑制するため転送スループットが低下するうえに、廃棄パケットの再送遅延によって利用者の体感品質を大きく損なう場合もある。

EDFR スケジューラ全体（マクロ）の廃棄特性は有限長 FCFS、すなわち既存スケジューラ、のそれと同等になることが知られている。EDFR では全てのパケットがそれぞれ異なる許容遅延を持っており、その算術平均と等価なキュー長の FCFS と同じ廃棄特性となることが示されている(参考①)。

我々は上の EDFR 廃棄率のマクロ特性に加え廃棄率の許容遅延依存性、すなわちミクロ特性について、シミュレーションによって評価した（論文③）。図では、点線がマクロレベルである廃棄率の平均許容遅延(D)依存性であり、既存研究と同じ結果を得た。

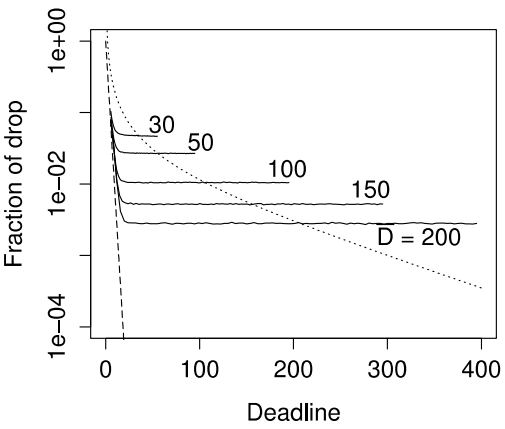

図 1　EDFR の過負荷時の廃棄特性

図 1 の実線部はミクロレベルの廃棄率の許容遅延依存性を示している。そして、廃棄率は低遅延領域を除いて許容遅延に関わらず一定となる。既存の有限長 FCFS スケジューラではパケットごとの許容遅延は全く考慮されず、到着時のキュー長で廃棄、送出が決定されるため廃棄率の許容遅延は一定とみなせる。すなわち、EDFR は廃棄率分布が一定の領域では、 FCFS と同等の廃棄特性と遅延限界に応じたパケットスケジューリングが両立している。

一方、低遅延領域の高い廃棄率は到着パケットの遅延限界が送出中のパケットの残り時間より短い場合に対応する。実際に低遅延領域の廃棄率分布は送出中パケットの残り分の送出時間、あるいはシリアル化遅延分布（破線）とよく一致している。高い廃棄率を示す低遅延領域を避けるには、遅延に対する要求を相応の値とする必要があるがその制約は特に厳しいものではない。たとえば、最

小リンク帯域 100Mbps、MTU 1500Bytes の条件で 125 ナノ秒以上の遅延限界を指定すれば完全に回避できる。

②EDFR が TCP に与える影響

EDFR パケットスケジューラが TCP に与える影響をシミュレーション(NS2)によって評価した（参考②、論文③)。評価は複数の遅延限界の TCP フローが共存したときの挙動について、Common TCP Evaluation Suite に沿って、TCP NewReno, CUBIC を対象に損失率・帯域スループット・FCT (Flow Completion Time）についておこなった（参考③)。

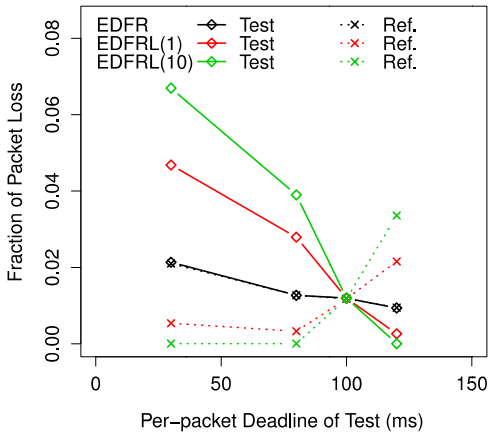

図 2 廃棄率の許容遅延依存性

図 2 黒線で示される EDFR のパケット損失率は、遅延限界の大小に関わらず同じ値を示した。これは、①EDFR スケジューラの廃棄特性、の結果と完全に一致し、TCP トラフィック環境においても EDFR の廃棄特性は維持されていることを示している。

一方で、帯域スループット・FCT に関しては、許容遅延の小さいフローが有利となる結果を得た。これは、許容遅延の小さなフローは優先的に転送されたるため、スケジューラでの待ち時間が短縮された結果であり、よく知られている TCP スループットとパケット廃棄率・往復遅延の関係も保存されている（参考④)。

以上のことから、既存の FCFS スケジューラを EDFR に置き換えても廃棄率・およびそれとスループットの関係といった TCP の振る舞いは変わらないことを示した。一方で、EDFR スケジューラの特性として許容遅延のより小さなトラフィックのスループットが有利となることから、アプリケーションが許容遅延を大きくとるインセンティブに欠けるという問題もあきらかとなった。

(2) EDFR スケジューラのハードウェア実装

現在の高速ルータのパケットスケジューラはハードウェアで実現されている。EDFR のハードウェアでの実現性を探るため FPGA への実装・評価をおこなった（論文④)。

EDFR スケジューラの実装は、Rotating Priority Queueing (RPQ) でおこなった。RPQ は、O(1) の優先キューとして知られるカレンダキューの近似方式である（参考⑤,⑥)。カレンダキューはタイムスロット毎にキューを分離し、パケット追加・削除のコストを抑える方式である。RPQ はカレンダキューのビン毎のスケジューラを優先キューから FCFS に置き換えたものである。

RPQ を NetFPGA-CML 上に実装したイーサネットスイッチ（図 3）で動作させギガビットイーサネットインターフェースにおける遅延を参照実装の FCFS スケジューラと比較した（参考⑥)。

RPQ による遅延は FCFS のそれに対し 50%程度増加した。これは RPQ ではパケットバッファが FPGA 内蔵メモリから外付け DRAM に移動したことが原因である。いずれにしろ RPQ を FCFS に置き換えても大きな性能劣化はないことを示した。

一方で、RPQ の誤差は広帯域化に伴って増大し、その特性は FCFS のそれに近づいていく。その誤差を抑えるには帯域増に伴いカレンダキューのビンあたりのタイムスロット粒度の細分化が必要となる。細分化のコストは小さなものではなく、RPQ における課題も明らかとなった。

我々は EDFR スケジューラとして Skip-FCFS も提案している（論文⑤)。Skip-FCFS は RPQ の課題であるタイムスロット粒度の細分化の解決が期待できる。

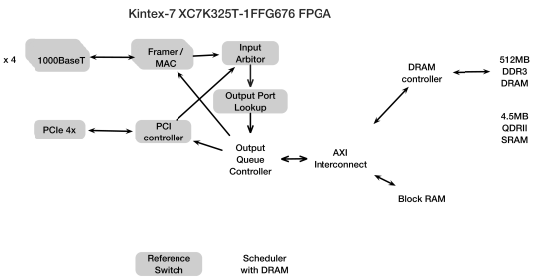

図 3 イーサネットスイッチ構成図

<引用文献>

① Kruk, Lukasz *et. al*, “Heavy traffic analysis for EDF queues with reneging”, The Annals. of Applied Probability, 21(2),484-545, 2011

② https://www.nsnam.org

③ Andrew, Lachlan *et. al*, “Towards a common TCP evaluation suite”, Proc. Int. Workshop on Protocol for Fast Long-Distance Networks, 2008

④ Mathis, Matthew *et. al*, “The macroscopic behavior of the {TCP} congestion avoidance algorithm”, ACM SIGCOMM Computer Communication Review 27(3),67-82, 1997

⑤ Brown, R. “Calendar Queues: A Fast 0(1) Priority Queue Implementation for the Simulation Event Set Problem”, Comm. ACM, 31(10), 1220-1227, 1988
⑦ https://netfpga.org

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計５件）
① 小林克志、廃棄つき Earliest Deadline First パケットスケジューラの設計および評価 - 遅延要求をサポートするインターネットにむけて -、信学技報、査読なし、114(374)、2014、37-42、https://www.ieice.org/ken/index/ieice-techrep-114-374.html
② 小林克志、並列処理とネットワークアーキテクチャ再考、信学技報、査読なし（招待講演）、114(518)、2015、81-86、https://www.ieice.org/ken/index/ieice-techrep-114-518.html
③ Katsushi Kobayashi, LAWIN: A Latency-AWare InterNet architecture for latency support on best-effort networks, 2015 IEEE 16th International Conference on High Performance Switching and Routing (HPSR), 査読あり、2015 July、Budapest、Hungary、10.1109/HPSR.2015.7483104
④ 小林克志、アプリケーションごとの遅延要求に応えるパケットスケジューラのハードウェアの設計と実装、信学技報、査読なし、116(362)、2016、29-34、https://www.ieice.org/ken/index/ieice-techrep-116-362.html
⑤ 小林克志、遅延要求をサポートするパケットスケジューラの FCFS による実現法の検討、117(354)、2017、7-12、https://www.ieice.org/ken/index/ieice-techrep-117-354.html

〔学会発表〕（計２件）
① 小林克志、遅延要求をサポートするネットワークアーキテクチャとパケットスケジューラ、JSPS インターネット技術 163 委員会研究会、査読なし、2014 11 月、倉吉市
② 小林克志、遅延要求に応えるパケットスケジューラのハードウェアの設計と実装、信学技報、査読なし、JSPS インターネット技術 163 委員会研究会、査読なし、2016 11 月、函館市

６．研究組織
(1)研究代表者
小林　克志　（KOBAYASHI, Katsushi）
東京大学・大学院情報理工学系研究科・特任准教授
研究者番号：９０２５１７１９