A Study on TCP Flow Control under Different Types of Segment Loss(セグメント損失の種類に応じたTCPフロー制御に関する研究)

氏　名(本　籍)	ザビル サラウッディン ムハマド サリム Zabir Salahuddin Muhammad Salim　　(バングラデシュ)
学位の種類	博　　士（情報科学）
学位記番号	情　　第 32 号
学位授与年月日	平成16年3月11日
学位授与の要件	学位規則第4条第2項該当
最終学歴	平成12年3月　東北大学大学院情報科学研究科システム情報科学専攻博士課程前期2年の課程
論文題目	A Study on TCP Flow Control under Different Types of Segment Loss （セグメント損失の種類に応じたTCPフロー制御に関する研究）
論文審査委員	（主　査） 東北大学教授　白鳥　則郎　　東北大学教授　曽根　秀昭 東北大学教授　加藤　　寧　　東北大学助教授　菅沼　拓夫

論文内容要旨

Recent years have experienced expansion of the Internet at a rapid pace over both wired and wireless medium. Applications running on them have been diversified, effective resource sharing techniques through mechanisms like multicasting have been coming up and the performance concerns are more engaging than any time before. About eighty percent of the Internet applications run on TCP as transport protocol. TCP issues, therefore, demand special attention for smooth Internet operations in the future. This work addresses TCP concerns for efficient performance by devising new TCP flow control mechanisms.

The major factor that influences TCP flow control is segment loss. A congestion in the network or a link error are the causes for segments to be lost. Any such occurrence affects TCP flow control resulting in a lowered data transmission rate from the sender. In our works as presented in this dissertation, we therefore propose approaches to handle different segment loss scenarios in appropriate manners. As such, better TCP flow control as well as improved TCP performance can be attained.

An important problem concerning TCP performance is associated with data loss due to network congestion. A congestion is mainly caused by over inflation of TCP transmission window on a slower link with limited buffer space. Since the network itself is the place where congestion occurs, it is imperative to handle the problem from within the network. Hence, to overcome problems associated with congestion loss, we propose network assisted approaches for efficient

TCP flow control. Network assisted approaches aim at avoiding occurrence of a congestion. Thus segment loss as well as corresponding negative impact on TCP flow control is minimized. Our proposals are concerned with detection of an incipient congestion. The objective is to notify TCP flow control mechanism at the sender for taking preventive measures when the network is about to be congested and the corresponding TCP flow occupies more share of the router buffer than it deserves. Upon receipt of this notification, TCP flow control takes appropriate actions.

Our first scheme, named Fair In-time Marking (FIM), attains this goal through Explicit Congestion Notification (ECN). The idea is to carry congestion signal by using two currently unused bits in the TCP header and two bits in the IP header of a TCP data segment and IP data packet respectively. When the network is about to be congested and the corresponding TCP flow occupies more share of the router buffer than it deserves, these bits are set in the headers by the concerned router. The TCP receiver in turn, forwards this notification to the sender using specified bits of the acknowledgements. Experiments show that FIM handles incipient congestion effectively so as to ensure high fairness among TCP flows as well as maintain high link efficiency. The improvement in fairness is in some cases, 20% higher than the most effective existing schemes while the efficiency remains high at the practically realizable maximum value.

In the next step, we evaluate effectiveness of our proposal with major TCP enhancement mechanisms. This bears significance since the processes of congestion loss handling differ to some extent in different TCP enhancements. Effectiveness with all major TCP schemes is important to generalize FIM as suitable for deployment with TCP. Results from experiments show that with all major TCP variants, FIM effectively reduces occurrence of congestion and associated data loss. It thus yields a high fairness as well as link efficiency. In addition, desired delay properties are maintained. That is, FIM ensures an average per segment delay comparable to that of existing schemes. In addition, it has the lowest variance in per segment delay than its contenders.

However, ECN is yet to be implemented in large scale. Most of the TCP end hosts are still conventional non ECN ones. Therefore, before ECN TCP gets gradual widespread deployment, normal network traffic would consist of data coming from both ECN and non ECN TCP connections. This leads to the requirement of a mechanism that can work effectively when both types of TCP traffic are present at the same bottleneck point. In keeping with this goal, for non ECN TCP hosts, we propose a strategy we named Fair In-time Dropping (FID). Here, in case of non ECN TCP connections, an early dropping of packet is deployed in order to notify TCP flow control mechanism at the sender for taking preventive measures when the network is about to be congested and the corresponding TCP flow occupies more share of the router buffer than it deserves. The ECN capable connections will be using FIM as usual for explicit incipient congestion notification. We observe that in addition to alleviating congestion, our proposals yield high level of fairness among TCP flows as well as maximum bottleneck bandwidth utilization and desired delay properties.

The problem with segment loss due to link error lies within the design of a traditional TCP end host. We therefore propose end host based solutions to tackle this problem. Traditional TCP was designed for wired networks having almost zero or negligible link error. As such, normally, there is no mechanism at the end host to identify a link error. Hence, any segment loss is inferred as a result of congestion in the network. Traditional TCP responds to network congestion by decreasing its data sending window or congestion window. Therefore, even when a segment is lost due to link error, TCP transmission rate reduces because of its misinterpretation. This results in a drastic degradation of TCP throughput when deployed over links that are prone to link errors. Wireless links are particularly prone to high link errors. Since in recent years, wireless links are being incorporated to the Internet at a high rate, such concerns deserve special considerations.

Our proposals presented in this dissertation either identify or estimate an occurrence of link error as the cause of a segment loss. Hence appropriate actions can be taken. Our devised mechanisms include TCP Identification and Revivable window (TCP I&RW) for identifying link error loss and revive TCP transmission rate accordingly. We introduce the concept of attaching a unique identification tag with each data segment or its retransmission. Identification tag is accommodated in the TCP timestamp option field in a TCP transparent manner. Therefore, TCP timestamp option's operations remain unaffected. As a segment is lost in the network due to link error, successful transmission of successive segments result in duplicate acknowledgements that trigger a retransmission of that particular segment and reduction of the transmission rate. The link layer, however, continues retransmission of the original segment. Thus, finally when that particular segment successfully reaches the receiver, corresponding acknowledgement carries the identification tag of the original segment. Upon receipt of this identification tag, TCP sender realizes the reason for segment loss and revives the transmission rate prior to the retransmission. It can also prevent an aggressive TCP receiver from occupying an unfairly high bandwidth by using a randomly generated identification tag. This randomness is maintained without violating the requirements of the PAWS (Protection Against Wrapped Sequence number) algorithm. Experiments reveal that our scheme can effectively identify the occurrence of a link error related segment loss. It also yields considerably better throughput than other existing ones over a reasonable range of link error rate.

TCP Indirect ACKnowledgement (TCP IACK), our next proposal, in addition to maintaining the performance improvement by TCP I&RW, can handle multiple segment losses and act robust against loss of acknowledgement segments. This is achieved through the introduction of one new field, packet in sequence count (pscount), in addition to the identification tag. Also a little modification in the echoing procedure of identification tags has been necessary. The purpose of pscount is to keep track of the segments that reach TCP receiver in proper sequence and inform that to TCP sender for appropriate actions in case of a segment loss. If an out of sequence segment is found that corresponds to data with higher sequence numbers than already received ones, pscount is reset to a value of zero. Receipt of each segment is acknowledged implicitly by echoing back its identification tag in the corresponding acknowledgement or duplicate acknowledgement. This mechanism leads to detect multiple segment loss within one Round Trip Time (RTT) as well as recover from the problems associated with acknowledgement loss. Through experiments, we confirm the effectiveness of TCP IACK. Our proposal has been successful in detecting multiple segment losses in one RTT and whether that is caused by a link error or not. We observe that our scheme is least affected by an acknowledgement segment loss. These improvements are reflected through a considerably better throughput yield than other schemes with similar or nearly similar functionality over a wide range of link error rates. In addition, our proposal incurs a very low or negligible overhead in comparison with its major contenders.

Both TCP I&RW and TCP IACK are meant for links, basically wireless, with link layer retransmission support. However, on links that do not facilitate a link layer retransmission like satellite links, TCP I&RW loses its effectiveness while TCP IACK loses much of its benefits. Therefore, for such cases, we propose our scheme TCP Seamless. This proposal accommodates a fast opening up of TCP connections over long delay networks as well as estimation and corrective actions upon detection of a link error. Our scheme contains two new major algorithms, namely, Fast Start and Seamless Recovery that replace Slow Start and Fast Recovery algorithms of TCP Reno respectively. The idea of using data segments of different priorities have been exploited for both the purposes. Since long delay links like satellite connectivity take quite a considerable amount of time for each round trip, conventional Slow Start algorithm becomes miserably slow when a TCP connection starts opening up. Our Express Start algorithm therefore, measures the ability of the network to carry data using probes in the form of lower priority data segments at the very beginning. The number of acknowledgements coming back to the sender leads to an estimate of data carrying capability of the network. Also, when a link error occurs, TCP sender should be able to estimate the actual reason of segment loss rather than lowering the transmission rate blindly. For this purpose, we again deploy lower priority data segments to determine or at least estimate the actual reason of the segment loss. Upon detection of a segment loss, after retransmission, the next RTT period is divided into two halves. During the first half, lower priority data segments are sent at a rate that is twice the rate of normal transmission. During the second half, the last half of lower priority data segments that were sent during the first half of RTT, are repeated with a higher priority. These actions lead to an appropriate estimation of the scenario. Corrective actions are also taken accordingly. Experiments verify the effectiveness of such a scheme over currently existing ones. Our scheme starts transmitting as much data as is possible from the beginning. In addition, even in the absence of link layer retransmission, it succeeds in estimating the occurrence of a link error. These capabilities are reflected in the overall performance as we observe that our proposals improve TCP throughput considerably than their nearest counterparts over a wide range of link error rates. For the purpose, our scheme incurs a significantly low overhead in comparison with its nearest contenders.

The contributions in this work bid fair to the advancement of knowledge in this area as well as ensure a more effective and diversified Internet for the future.

論文審査の結果の要旨

インターネットの普及と様々な利用形態の登場により，広域ネットワーク環境で効果的なデータ転送を実現するためのフロー制御技術が重要となってきている．しかし，従来のトランスポート層プロトコル（TCP）では，輻輳やリンクエラーによるセグメント損失がフロー制御に大きな影響を及ぼすため，十分なスループットなどの性能を実現することは困難であった．そこで著者は，広域ネットワーク環境におけるセグメント損失による TCP の性能劣化について分析し，これらを解決するための手法に関する詳細な研究を行った．本論文はその成果をまとめたものであり，全編 5 章からなる．

第 1 章は序論である．

第 2 章では，様々な特性を持った広域ネットワーク環境における TCP 性能劣化の原因を詳細に分析し，輻輳やリンクエラーによって生ずるセグメント損失がフロー制御に大きな影響を与え，それが性能劣化の主たる原因となることを明らかにしている．これは，広域ネットワーク環境における TCP 性能改善のための基礎となる知見である．

第 3 章では，TCP のセグメント損失が輻輳によって発生している場合の TCP 性能の改善手法として，TCP フローの公平性を考慮しつつ性能向上をはかるフロー制御方式を提案している．また，シミュレーションにより既存プロトコルとの比較を行い，提案方式の有効性を示している．この結果は，広域ネットワーク環境における公平で効率的なネットワーク利用のための有用な成果である．

第 4 章では，無線リンクを含む広域ネットワーク環境において，TCP のセグメント損失が無線リンクエラーによって発生している場合の TCP 性能の改善手法として，3 つの方式を提案している．具体的には，リンクエラーを検出した場合にのみウィンドウサイズの適切な制御を行う方式，複数のセグメント損失時に適切な制御を行う方式，および下位層でのリンクエラー訂正機能を持たない無線リンク上で効果的にフロー制御を行う方式である．また，シミュレーションによりこれらの方式と既存プロトコルとの比較を行い，これらの提案方式が従来方式よりも高い性能を実現していることを示している．この結果は，無線リンクを含んだ広域ネットワーク環境での効率的なデータ転送のための重要な成果である．

第 5 章は結論である．

以上要するに本論文は，様々な特性を持った広域ネットワーク環境上で効果的なデータ転送を実現するための TCP 性能改善手法の提案・評価を行い，その有効性を示すことによって，トランスポート層プロトコルの構成に関する基盤を与えたものであり，情報基礎科学の発展に寄与するところが少なくない．

よって，本論文は博士（情報科学）の学位論文として合格と認める．