cmpSCTP: An Extension of SCTP to Support Concurrent Multi-path Transfer

Jingyu Wang, Jianxin Liao, Xiaomin Zhu
State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications
P.O. Box 296, Beijing University of Posts and Telecommunications, Beijing, P.R. China, P.C.: 100876
Email: {wangjingyu, liaojianxin, zhuxiaomin}@ebupt.com

Abstract—This paper introduced cmpSCTP, a transport layer solution for concurrent multi-path transfer that modifies the standard Stream Control Transmission Protocol (SCTP). The cmpSCTP aims at exploiting SCTP’s multi-homing capability by selecting several best paths among multiple available network interfaces to improve data transfer rate to the same multi-homed device. Through the use of path monitoring and packet allotment techniques, cmpSCTP tries to transmit given amount of packets at corresponding path as its ability. At the same time, cmpSCTP updates the transmission strategy based on the real-time information of all of paths. Using cmpSCTP’s flexible path management capability, we may switch the flow between multiple paths automatically to realize seamless path handover. The theoretical analysis evaluated the model of cmpSCTP and formulated optimal traffic fragmentation of real-time streaming. Extensive simulations under different scenarios using OPNET verified that cmpSCTP can effectively enhance transmission efficiency and highlighted the superiority of cmpSCTP against the other SCTP’s extension implementations under performance indexes such as throughput, handover latency, packet delay, and packet loss.

Keywords—Multi-homing, Concurrent multi-path transfer, Traffic control, SCTP.

I. INTRODUCTION

A host is multi-homed if it can be addressed by multiple IP addresses, as is the common case when the host has multiple network interfaces. Multi-homing is increasingly economically feasible and can be expected to be the rule rather than the exception in the near future. A Multi-homing host may be simultaneously connected through multiple access technologies, and even multiple end-to-end paths to increase resilience to path failure [1]. For instance, a mobile user could have simultaneous Internet connectivity via a wireless local area network using 802.11 and a wireless wide area network using UMTS.

As of today, the only transport protocol supporting Multi-homing is the Stream Control Transmission Protocol (SCTP) [2]; yet, SCTP can only exploit at most one of the available paths at any given time. That is, SCTP uses multiple interfaces only for redundancy: every host chooses a primary destination address, normally used for the transmission of the data units, “data chunks” in SCTP terminology, whereas the alternate addresses are considered as secondary, whose conditions are periodically monitored with the transmission of probe chunks called Heartbeat. The backup path is used only (i) to retransmit lost data chunks, in order to increase the probability of successful retransmissions, (ii) to transmit new data chunks when, due to the excess of the number of (e.g., five) consecutive timeouts on the primary path, the default interface is declared as “inactive”. In the latter case, SCTP transmits new data chunks toward the backup interface and Heartbeat chunks toward the primary one. As soon as theHeartbeats reception on the primary interface is confirmed, its state is toggled to “active” and the transmission is resumed.

The remainder of this paper is organized as follows. Section II surveys related work. Section III proposed the concurrent multi-path SCTP, with original contribution presented in detail, and Section IV describes analytical models to evaluate the performance of cmpSCTP and compare with other extensions based on SCTP. In Section V, simulation models are described, and numerical results are presented to investigate the performance. Finally, conclusions and possible future work are described in Section VI.

II. RELATED WORK

Researches on extending SCTP to support concurrent multi-path transfer, e.g. simultaneously sending data over multiple available paths to increase the association bandwidth, are currently in progress [3]–[7].

The work in [3], [4], by the original SCTP proposers, suggests to change the SCTP sender operation to compensate for the problems introduced by using a unique sequence-number space for tracking packets sent over multiple paths. The sender maintains a set of per-destination virtual queues and spreads the packets across all available paths as soon as the congestion window allows it. Retransmissions are triggered only when several Selective ACKnowledgments (SACKs) report missing chunks (SCTP protocol data units) from the same virtual queue.

AL et al [5] propose Load-Sharing SCTP (LS-SCTP), a mechanism to aggregate the bandwidth of all the paths connecting the endpoints and dynamically adds new paths as they become available. Path monitoring is used to stripe packets among all available paths ensuring that the association does not stall because of high loss rates or temporary path unavailability. The key idea is to introduce a per-association, per-path data-unit sequence numbering that extends the per-association SCTP congestion control to a finergrained, per-path congestion control.

Hsieh et al. [6] propose pTCP (parallel TCP) based on the Transmission Control Protocol (TCP). pTCP has two
components - Striped connection Manager (SM) and TCP-virtual (TCP-v). For each pTCP socket opened by an application, pTCP opens and maintains one TCP-v connection for every interface over which the connection is to be striped on. The TCP-v's are separate connections that are managed by the SM. This decoupling of functionality allows for intelligent scheduling of transmissions and retransmissions. Similarly, mTCP [7] also significantly modifies TCP to use multiple paths provided by a special routing layer that implements from a Resilient Overlay Network (RON), and also employs mechanisms to handle reordering side-effects. mTCP also outlines a heuristics for the detection of disjoint paths.

However, none of the previous proposals fully addresses the case in which the paths comprised in the SCTP association exhibit widely-different bandwidths and round trip times (RTTs). In such scenario, the packets sent by the source reach the destination out of order, triggering a lot of retransmissions on the underlying TCP SACK on which SCTP is based. The most promising solutions proposed to mitigate packet reordering are based on: i) estimating the available bandwidth and the RTT on each path, ii) using an appropriate packet scheduling algorithm to stripe packets across all paths, so that they reach the destination almost in order.

III. A CONCURRENT MULTI-PATH SCTP

As a general remark, we found our modifications to be similar to those recommended by other Multi-homing related papers [3]-[7]. In this section, we will refer to the features of cmpSCTP different with those techniques and present the key design elements of the cmpSCTP protocol.

A. cmpSCTP Design

Similar to LS-SCTP[5], cmpSCTP is also based on the idea of separating the association flow control from congestion control. In cmpSCTP the flow control is on association basis, thus both the sender and receiver endpoints use their association buffer to hold the data chunks regardless of the path these data chunks were sent or received. On the other hand, congestion control is performed on per path basis, thus the sender has a separate congestion control for each path. Especially, the congestion control mechanism on each path can follow the standard SCTP [2], TCP Friendly Rate Control (TFRC) [8] and other congestion control algorithms, so as to insure fair integration with other traffic in the network.

To support the decoupling of functionalities, cmpSCTP uses several novel mechanisms including multi-buffer structure, multi-state management, two-level sequence number, and cooperative SACK strategy to realize effective bandwidth aggregation. Also cmpSCTP includes an overall retransmission technique that prevents the side effects of simultaneous transmission of data on paths with different characteristics, including unnecessary fast retransmissions, which ensures fast delivery of lost data chunks to prevent stalling the association.

Through extending dynamic address reconfiguration [10], cmpSCTP keeps ongoing end-to-end paths alive and provides adaptive load sharing in multiple paths. In addition, cmpSCTP extends SCTP path-monitoring feature, through regular transmission of actual effective data chunks, to update the list of unstable paths suitable for load sharing.

B. Function Modules and Interfaces

Specifically, the structure of cmpSCTP consists of two modules: One is the Association Management (AM) which carries out flow control function for an association and the other is the Single Path Management (SPM) each of which performs single-path congestion control for its responsible path independently. Figure 1 provides an overview of the cmpSCTP architecture and key data structures. cmpSCTP as a transport layer protocol interacts with the application and IP, and acts as a container loaded one AM engine and several SPM engines. When an application opens a cmpSCTP socket, by default one SPM module corresponding to the current network interface in use is created. For each additional interface that becomes active during the lifetime of the association (e.g. when the mobile host moves into the coverage area of another network), AM module creates one more SPM module. For each interface that becomes inactive, AM closes the corresponding SPM module and removes it from the association. The number of SPM modules that co-exist in a cmpSCTP association thus depends on the number of active network interfaces used by the application.
chunk using the receive() interface, and enqueues the cmpSCTP data chunk in the recv buffer using the association sequence number, as will be discussed in Section III.E. Then the in-sequence data enqueued in recv buffer can be delivered to the application via the read() interface. When SPM module receives an ACK control chunk from IP, it passes the AM module. Then the AM module processes the ACK and updates state variables of concerned SPM module. The cmpSCTP selective ACK carries cumulative association-level ACK information and receiver window size (rwnd) for each path that can be used to perform per path congestion control as will be discussed in Section III.F and III.G. Summing all of the rwnds for each path AM can obtain the association receiver window (arwnd). APM module calculates its congestion window (cwnd) as given congestion control algorithms, and informs the AM module of a new association congestion window (acwnd) value by update_acwnd() function. SPM uses the shrink() interface to notify AM of any change in the size of its module (e.g. change in the congestion window). Upon receiving the shrink() call, AM performs dynamic reassignment. If any path becomes revived, AM uses the resume() interface to “de-freeze” the SPM module of the corresponding path, making it start requesting for transmissions as before.

Finally, AM uses the close() interface to tear down a SPM module. Once the SPM module returns with the closed() interface, AM removes it from the socket. Note that cmpSCTP binds application data to a SPM module only when the concerned module requests for transmissions. Hence adding one more module to the aggregate association has the effect of draining the AM send buffer at a faster rate, and deleting a module implies a slower rate. The dynamic addition or deletion of SPM modules does not influence the functionalities that cmpSCTP performs including reliability.

C. Multi-Buffer Structure
The single buffer architecture of SCTP has been replaced by a multi-buffer structure, meaning that each connection now has its own send buffer, and the total association has a single, shared send buffer. When a new chunk is received from the upper layer, the traffic scheduler is invoked to determine the path it will be sent over, as will be described in Section IV, the chunk is then queued into the chosen path send buffer.

However, a single receiver buffer is presented in cmpSCTP, like in standard SCTP, though each connection in the association is assigned a virtual buffer from the unique receiver buffer for per-path congestion control. In fact, all the chunks from different path are collected by a single shared send buffer. When a new chunk is received from the upper layer, the traffic scheduler is invoked to determine the path it will be sent over, as will be described in Section IV, the chunk is then queued into the chosen path send buffer.

```
Recv_Buffer_min = Bavail * RTT_{Max}
```

Where $B_{avail}$ is the available bandwidth through all available paths, $RTT_{Max}$ is the maximum $RTT$.

The multi-buffer structure guarantees connection independence as far as transmission is concerned, but it introduces the need for modifications to Selective Acknowledgement (SACK) management at the source as will be discussed in Section III.G. As a matter of fact, SACKs are generated as in AM and transmitted over the path from which the last data chunk was received, but they also carry information about other on-going connections of the same association, as is the case for delayed acknowledgments. Thus, when a SACK arrives at the source, the information is processed on each interface and the relative send buffer is refreshed.

Since the beginning of the association, each single connection proceeds without interference from other connections, handling only those packets in its send buffer. When a new data chunk needs to be transmitted, it is inserted in the send buffer of the connection indicated by scheduling algorithm (Round-Robin, Bandwidth-aware, or other more optimal algorithm). From then on, it is the connection’s responsibility to ensure that it is delivered: the data chunk remains in the send buffer until acknowledged; it causes Head-of-the-Line blocking to its own connection, but not to the other connections. When a retransmission timeout occurs, a retransmission process is declared. The cmpSCTP will use an alternative path with lowest packet drop probability. In this case, lost chunks queued in the send buffer of the failed path are shifted into the send buffer of the path chosen for the retransmission.

D. Dynamic Multi-State Management
The cmpSCTP is a multi-state transport protocol that creates and maintains one SCTP state for each network interface used by the application. As shown in Figure 2, a SPM (SCTP-singlepath) module is created for each active interface used in a cmpSCTP association, to manage the per-path SCTP state including the pair of IP addresses, SCTP ports, and congestion control parameters. While each SPM module (and the corresponding state) is addressed using the conventional connection 4-tuple, a cmpSCTP socket can be addressed through its association id of AM module. All SPM modules in a cmpSCTP association are pointed to the same association id. cmpSCTP dynamically adds or deletes states in a association depending on the connectivity (and the change in addresses) between the end hosts. Therefore, although each state can only be associated with one pair of network addresses specified when the state is created, cmpSCTP allows the addresses of the end points in an association to be changed dynamically. Moreover, as a multi-state transport protocol, cmpSCTP allows multiple SPM modules to co-exist in an association, and hence seamless handover of the transport layer states are possible.

E. Two-Level Sequence Number
The standard SCTP uses Transmission Sequence Number (TSN) as sequence number in its congestion control algorithms. However TSN might simultaneously uses only one path although it is used by entire association. Therefore we still continue to use the TSN as the sequence number that can be used for independent congestion control over each path.

The proposed cmpSCTP operates in two levels, involved
in connection level and association level. Every level has its own sequence number. At connection level, we still use the TSN, which is not used for entire association, but still represents the sequence of SCTP DATA chunk transmitted through per path, used for reliability and congestion control on each path. TSN may split the first 4-bit as path ID (PID) denoted the transmission path, the remain 12-bit part represents per-path-sequential TSN. Data chunks sent to the same path are assigned same path ID (PID) and per-path-sequential TSN. At the same time, data chunks sent to different path carry different PID. Since TSN is used only by SMTP to keep track of the states of data chunks sent through each path and per-path congestion control is activated.

At association level, the Association Sequence Number (ASN) of each data chunk, which is a per association sequence number, is used to reassemble all received data chunks from different paths to an integrated file. In other words, we use ASN to reorder the received data chunks at the receiver association buffer, regardless the path from which they have been received.

Such, we defined a new modified data chunk as Fig.3, by adding two new parameters to the standard SCTP data chunk [2]. The first parameter is a 4 bits Path Identifier (PID), which identifies the path used for the data chunk transmission. The second parameter is 16 bits Association Sequence Number (ASN), which is a monotonically increasing sequence number for the data chunks transmitted over the association. In addition, cmpSCTP continues to use the SSN for ordering the data chunks within the association streams.

<table>
<thead>
<tr>
<th>PID</th>
<th>Flags=UBE</th>
<th>Chunk Length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig.3. Modified cmpSCTP Data Chunk

F. Independent congestion control

In order to perform per path congestion control, cmpSCTP uses a number of internal variables to control the rate at which data is injected into the network. These parameters for each path are as following, all are maintained in send buffer and receiver buffer respectively. In addition, these buffers keep track of the TSNs sent and received on each path.

- **rwnd** – this corresponds to a sender’s view of the receiver’s incoming buffer space. Summing all of the rwnds can obtain the arwnd, which represents the available space in the receiver’s association buffer is kept on an association basis.

- **cwnd** – this corresponds to the sender’s view of network conditions. The initial value of cwnd is less than or equal to twice the path maximum transmission unit (MTU), corresponding to the more aggressive value recently adopted for SCTP.

- Slow-start threshold (ssthresh) – the sender uses this to distinguish between slow-start and congestion avoidance phases.

- Partial bytes acked – the sender uses this to limit cwnd adjustment to only once per RTT during congestion avoidance phases.

The congestion control algorithms of slow start, congestion avoidance, fast retransmission, fast recovery and timeout retransmission may used by the standard SCTP will all be applied to each path, on the other hand, other congestion control scheme, for example TFRC [8] and wireless TCP etc, may also be used and enable congestion control to special applications like multimedia streaming and various scenarios.

The cmpSCTP receiver of per-path still uses TSN to facilitate their operations. When a receiving DATA chunk advances the TSN for a path and current receiver window of that path is being fully utilized, the receiver window of that path is allowed to be increased by at most the lesser of 1) the total size of the previously outstanding DATA chunk(s) acknowledged at that path, and 2) the destination’s path MTU. Then the sender receivers per-path receiver window reports from the receiver through modified cmp-SACK as described in Section III.G, all per-path congestion window are refreshed.

G. Handling of cmp-SACK chunk

In order to acknowledge the received data chunks, cmpSCTP defines concurrent multi-path SACK (cmp-SACK) as Figure 4. The cmp-SACK chunks, which can also be called ASN-based SACK, received by a cmpSCTP source endpoint actually reflect the reception of ASNs sent through all the paths used by a cmpSCTP association. The cmp-SACK chunk carries several pieces of information includes four different parameters than the standard SCTP SACK, Cumulative ASN ACK, Time Stamp, Path ID and Advertised Receiver Window Credit.

- The Cumulative ASN Ack is a per association cumulative acknowledgement instead of Cumulative TSN ACK in the standard SCTP SACK, which records the highest sequential ASN received.

- Time Stamp is used to order the cmp-SACKs received from the different paths.

- Every Path IDs and their correspondent Advertised Receiver Window Credit, each of them reflects the current capacity of each receiver’s inbound virtual buffer.

- A sequence of Gap Ack Blocks, which record any out of sequence ASNs received.

- A sequence of Duplicate ASNs, which record any ASNs for which duplicates have been received.

<table>
<thead>
<tr>
<th>Type</th>
<th>Flags=0</th>
<th>Length=variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cumulative ASN Ack</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time Stamp</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PID</td>
<td>Advertised Receiver window credit(rwnd)</td>
<td></td>
</tr>
<tr>
<td>PID sk</td>
<td>Advertised Receiver window credit(rwnd)</td>
<td></td>
</tr>
<tr>
<td>Num of Fragments=N</td>
<td>Num of Dups=N</td>
<td></td>
</tr>
<tr>
<td>Gap Ack Bk #1 start</td>
<td>Gap Ack Bk #1 end</td>
<td></td>
</tr>
<tr>
<td>Gap Ack Bk #N start</td>
<td>Gap Ack Bk #N end</td>
<td></td>
</tr>
<tr>
<td>Duplicate ASN #1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duplicate ASN #M</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig.4. Modified cmpSCTP SACK

Data is not considered fully delivered until the cumulative ASN Ack point advances past its ASN. Thus the information in the Gap Ack Blocks corresponds to SCTP SACK blocks. Under normal conditions cmpSCTP uses a delayed acknowledgement scheme which sends one cmp-SACK for every mean RTT (RTTmean) within the association incoming packet which contains one or more new data chunks.

Considering the chunks of the same ASN is possible to
receive from different paths on a certain redundant transmission strategy, acknowledgement mechanism should be design based on ASN to prevent unnecessary retransmissions unless data chunk losses occur on every transmission path. Fast retransmission is triggered by four consecutive duplicate SACKs at an association. Whenever a data chunk needs to retransmit, cmpSCTP will use an alternative path with packet lowest drop probability. Since in our concurrent multi-path transfer, current cwnds of all paths are available, such modified retransmission scheme is more efficient than defined in RFC 2960: selected alternative path contain the four new modified and appendant request parameters for the paths. These parameters signal:

-0xC001-AddPath: the path specified is to be added to the Monitored_Set;
-0xC002-DeletePath: the path specified is to be removed from the Monitored_Set;
-0xC007-ActivePath: the path specified is to be added to the Active_Set;
-0xC008-DeactivePath: the path specified is to be removed from the Active_Set.

In a handover situation, MH sends CT ASCONF chunks with these four different types of parameters. To add new path, for example, the MH should send ASCONF chunk of 0xC001 type, which should be acknowledged by ASCONF-ACK chunk. Such through using the concurrent multi-path feature, cmpSCTP could reduce latency and increase throughput during performing vertical handover. This means that packets are not lost during the handover and there is no interruption to service, making it suitable for handover of real-time traffic.

Commonly, the monitored path is assigned the probing data have been send to the active path originally. Such, a data chunk can be carried by multiple paths to minimize the impact of delayed loss recovery. Specifically, the data carried by the first path that recently suffered from blackouts will be bound to another path, such that the association can continue receiving data while the concerned path probes for the duration of blackouts.

The cmpSCTP handover support can be achieved by using modified Address Configuration Change (ASCONF) and Address Configuration Acknowledgement (ASCONF-ACK) control chunks of mSCTP, which may contain the four new modified and appendant request parameters for the paths. These parameters signal:

-0xC001-AddPath: the path specified is to be added to the Monitored_Set;
-0xC002-DeletePath: the path specified is to be removed from the Monitored_Set;
-0xC007-ActivePath: the path specified is to be added to the Active_Set;
-0xC008-DeactivePath: the path specified is to be removed from the Active_Set.

In a handover situation, MH sends CT ASCONF chunks with these four different types of parameters. To add new path, for example, the MH should send ASCONF chunk of 0xC001 type, which should be acknowledged by ASCONF-ACK chunk. Such through using the concurrent multi-path feature, cmpSCTP could reduce latency and increase throughput.

**H. Flexible Multi-Path Management**

The mobile SCTP (mSCTP) [9] is the SCTP with the ADDIP extension. It utilizes dynamic address reconfiguration [10] to manage the possible changes of IP addresses such as adding new addresses and deleting obsolete addresses while keeping ongoing end-to-end connections alive. However, the current SCTP specification designates only one path at each destination host as the primary path, and all new data is transmitted to the only primary path. If ever the primary path fails, new data transmission fails over to an alternate reachable destination path. Furthermore, the SCTP association experiences a reduction of traffic rates immediately after handover regardless of available traffic state of the new primary path.

In order to mitigate these negative effects, we refer to this new extension by cmpSCTP. It proposes to achieve higher throughput and seamless handover in an SCTP association by concurrently using all independent paths between a sender and receiver for data transfer.

By the cmpSCTP, we mean that the mobile host is maintaining connections with more than one path. These paths are classified by following two set:

1. Assume MH is connected to any mobile network, then MH find out another mobile networks
2. MH enters concurrent multi-path transfer
   - MH keeps an association to several of paths simultaneously
   - AddPath (the new path);
   - MH connects k available paths including an Monitored_Set(n candidate paths) and Active_Set (m stable paths)
3. MH measures the path rate received from the Monitored_Set and Active_Set paths periodically.
4. For i = 1…m
   - If (Monitored_Set.Path.state==CON_AVOID &&& Expectation of recently Monitored_Set.Path-rate) >= Rate threshold &&
     - Variance of recently Monitored_Set.Path-rate <= Variance threshold)
     - Activepath(Monitored_Set.Path);
5. For j = 1…k
   - If (Monitored_Set.Path.state==CON_AVOID &&& Expectation of recently Active_Set.Path-rate) < Rate threshold ||
     - Variance of recently Active_Set.Path-rate <= Variance threshold)
     - Deactivepath(Active_Set.Path);
6. Update Fragmentation_Strategy (Active_Set);
7. If Receiver(Path.state==LOSS) DeletePath(Path);
8. When MH only connected one mobile network, MH leaves concurrent multi-path transfer

The **Active_Set** includes the paths that form a cmpSCTP connection to the mobile host, which allows to transfer data packets for the MH.

The **Monitored_Set** is the list of candidate paths that the mobile hosts continuously measure, but their bandwidths is not enough to be added to the **Active_Set**.

The cmpSCTP handover support can be achieved by using modified Address Configuration Change (ASCONF) and Address Configuration Acknowledgement (ASCONF-ACK) control chunks of mSCTP, which may contain the four new modified and appendant request parameters for the paths. These parameters signal:

-0xC001-AddPath: the path specified is to be added to the Monitored_Set;
-0xC002-DeletePath: the path specified is to be removed from the Monitored_Set;
-0xC007-ActivePath: the path specified is to be added to the Active_Set;
-0xC008-DeactivePath: the path specified is to be removed from the Active_Set.

In a handover situation, MH sends CT ASCONF chunks with these four different types of parameters. To add new path, for example, the MH should send ASCONF chunk of 0xC001 type, which should be acknowledged by ASCONF-ACK chunk. Such through using the concurrent multi-path feature, cmpSCTP could reduce latency and increase throughput.

Commonly, the monitored path is assigned the probing data have been send to the active path originally. Such, a data chunk can be carried by multiple paths to minimize the impact of delayed loss recovery. Specifically, the data carried by the first path that recently suffered from blackouts will be bound to another path, such that the association can continue receiving data while the concerned path probes for the duration of blackouts.

The MPM (Multi-Path Management) algorithm determines the **Active_Set** that is likely to yield the maximum throughput for upper application. The rest of the candidate paths are kept in a Monitored_Set, from which replacement paths will substitute failed or degraded paths from the **Active_Set**. The unstable paths are put into the Monitored_Set to detect their available network conditions. These paths are periodically checked against the so called “triggering conditions”. If a triggering condition is fulfilled, the MH decides if a path should be added to the Active_Set. Then, MH creates a report which is sent to the CH The rest of the candidate paths are kept in a Monitored_Set, from which replacement paths will substitute failed or degraded paths from the Active_Set. The sender assigns a sending rate to each

---

**Fig.5. Multi-Path Management Algorithm**
of the active paths. The strategy continues as far as there is no need to switch to a different Active Set. A switch is needed if a path fails or the network path becomes congested. At that time, the topology is updated with new values. The algorithm is described in Figure 5.

IV. ANALYTICAL MODELING

For the sake of illustration, consider two network hosts, host A and host B, where host A is sending data to host B. Assume there are K distinct paths from A to B, where the time needed to send a packet along path j is a linear function of flow size.

It is assumed that the path in Active Set has reached a steady state, that is, its transport protocol can be modeled with a Markov chain for reference [11,12]. We consider the arrival traffic rate at a link queue (segments/sec) as queue with Poisson arrivals with rate $\lambda$. The service time of packet $i$ is assumed to have an exponential distribution with rate $\mu_i$, where $\mu_i$ are independently and identically distributed (i.i.d) random variables with an arbitrary distribution. The service times of packets are assumed to be independent. Once a packet has entered the queue, it does not leave until it completes service. The queue can be analyzed by considering a single cycle consisting of a busy period and an idle period.

As the utility for each path in the flow control problems, we use the power for the path. Since the throughput is equivalent to the rate of the arrival flow (i.e., the arrival rate), from [13], the overall power $P$ of the system and the power $p_i$ of path $i$ are given as

$$P = P(\lambda) = \begin{cases} \frac{\lambda}{n\mu} & \text{if } \lambda \leq n\mu \\ 0 & \text{otherwise} \end{cases}$$

and

$$p_i = p(\lambda) = \begin{cases} \frac{\lambda}{n\mu} & 0 \leq \lambda \leq n\mu - \sum_{j<i} \lambda_j \\ 0 & \text{otherwise} \end{cases}$$

Where $T$ denotes the mean delay time in the system and $i=1, 2, \ldots, n$, respectively. Note that $\lambda = \sum \lambda_i$. We have $P = \sum p_i$. $P$ and $p_i$ are positive for $\lambda < n\mu$ and zero for $\lambda = 0$ and $n\mu$.

We can formulate, for the system based on cmpSCTP described above, two typical optimal flow control schemes: the noncooperative scheme (I) and the overall cooperative scheme (II). The two schemes in different queuing system are presented as follows:

(I) The noncooperative scheme: Each path strives to maximize unilaterally its own power. That is, the noncooperative scheme is to find $\lambda^*_i$ for each $i = 1, 2, \ldots, n$, that satisfies

$$p^*_{i} = \max_{\lambda_i \geq 0} p_i(\lambda^*_i)$$

(II) The overall cooperative scheme: A single agent maximizes the overall power of the system, that is, it strives to find $\lambda^*$ that satisfies

$$p^* = p(\lambda^*) = \max_{\lambda \geq 0} p(\lambda)$$

Based on the above definitions, we consider two queuing system models as shown in Fig. 6: $S_n$—a system consisting of $n$ separated M/M/1 queues (top) and $S_{M/M/n}$—an M/M/n queue (bottom). In the former, the flow control schemes are concerned with each separated M/M/1 queue. On the other hand, in the latter, the flow control schemes are concerned with the M/M/n system, where $K$ is the number of active paths that are serviced at the same time over specific network. We recall that the latter is the result of grouping together the former.

For $n$ separated M/M/1 queues, an arrival traffic rate that is mutually independent and forms a Poisson arrival process with rate $\lambda_i$, $i=1, 2, \ldots, n$. Note that $\lambda = \sum \lambda_i$ and that the transmission time is independent of the player that sends the job.

Assume the average arrival rate of every queue is $\lambda^*$. From M/M/1 queuing theory, System utilization is

$$\rho = \frac{\lambda^*}{\mu}$$

and Probability of no packet in the system is

$$P_0 = 1 - \rho.$$ The average number of packets in the system is

$$L_q = \frac{\lambda^*}{\mu(\mu - \lambda^*)}.$$ Further, by Little’s law, the average delay in multi-queues system is

$$T = \frac{1}{\mu - \lambda^*}.$$ For an M/M/N queue, the system utilization is $\rho = \frac{\lambda}{n\mu}$, and Probability of no packet in the system is

$$P_0 = \frac{1}{n! (1 - \rho)}.$$ the mean number of segments in queue can be calculated as:

$$L_q = \frac{(\lambda / \mu)^n \rho}{n!(1 - \rho)^2} P_0.$$ By Little’s law, we can model the average delay in system as:

$$T = \frac{1}{\rho} L_q + \frac{1}{\mu}.$$ According the queuing theory, the system power of the noncooperative scheme consisting of $n$ separated M/M/1
queues is not as good as the cooperative scheme of M/M/n queue, that is \( p^* \leq p^* \). Thus from the view of the whole system flow model, cmpSCTP needs to use the following mechanisms to achieve the overall cooperative scheme avoiding the noncooperative scheme:

1) Delayed Fragmentation: In cmpSCTP, data will be bound to a path only when the data is ready to transmit. cmpSCTP does not allow data to be queued up inside each path as soon as possible.

2) Quicken Update: when a data segment can be lost through one path, cmpSCTP will assign another most reliable path to carry the retransmission data segment in order to minimize the impact of data loss.

3) Dynamic reassignment: If a path reports losses or bandwidth fluctuations, cmpSCTP immediately unbinds the data that is lost or overflows, such that other paths ready for transmission can timely deliver the concerned data.

4) Cooperative retransmission: when a data segment can be lost through one path, cmpSCTP will assign another most reliable path to carry the retransmission data segment in order to minimize the impact of data loss.

V. SIMULATION AND RESULT

The proposed cmpSCTP protocol has been implemented in the network simulator OPNET [14], and tested with various network configurations. The purpose of the extensive simulations is two-fold: first to investigate the performance of the proposed cmpSCTP with various network parameters, and second to compare the cmpSCTP protocol with other multi-path transport protocol.

In our simulation, we created the network topology consisting two hosts. To investigate the impact of various network parameters on the performance of the cmpSCTP, the multiple overlapping cells are also varied by using different simulation configurations including the number of overlapping cells, and available bandwidth in the cell the mobile host is entering. Available bandwidth in a cell is varied by changing the average of the Poisson distribution used to generate background traffic in all of cells.

To investigate the impact of various network parameters on the performance of the cmpSCTP, the network is only considered concluding two cells consisting with the two same or different base stations. The following network parameters are varied in the simulations: the number of overlapping cells and available bandwidth in the cell the mobile host is entering. Available bandwidth in a cell is varied by changing the average of the Poisson distribution used to generate background traffic in all of cells.

A. Effect of Path Number

In this preliminary experiment, we tested the capability of the cmpSCTP protocol to provide load sharing in multiple paths. We created the network is considered including multiple overlapping cells consisting with the same base stations.

In Figure 7, referring to the four-path case, we can see that using cmpSCTP as the transport layer, the traffic load is sharing among the available paths. Along with the number of path increases the per-path throughput lessens little due to the overhead and network congestion, which the association throughput reaches maximum value.

Moreover, Figure 8 shows the delay of packet at receiver depending on the number of paths between the two nodes. In all cases, the delay stays below 160 ms. This, together with measured average delays in the 40-160 ms range, lead to excellent performances of real-time applications. The more the number of path introduces, the less every packet brings delay. That is to say, we can achieve the more stable performance through concurrent multi-path transfer of cmpSCTP.

V. SIMULATION AND RESULT

B. Impact of Bandwidth Disparity

In this experiment we tested the robustness of the cmpSCTP protocol in dynamic conditions. We assumed that we have two paths, namely path 1 and path 2 in order to examine the performance of cmpSCTP under the condition of paths with diverse capacities. The speed of mobile host movement is assumed to be 15 m/s (meters per second) and the RTT is assumed to be 60 ms. In case 1, the mobile host moves from a cell with a larger amount of available bandwidth (2Mbps) to a cell with a smaller amount of available bandwidth (1Mbps). In case 2, a mobile host moves between two homogeneous cells with the same amount of available bandwidth (2Mbps). In case 3, a mobile host moves from a cell with a smaller amount of available bandwidth (2Mbps) to a cell with a larger amount of available bandwidth (5Mbps). In our performance study we used the association throughput as a performance metrics, which is defined as the amount of data delivered to the receiver’s application layer per second. Simulation results are presented to evaluate the performance of the cmpSCTP.

In Figure 9 shows the throughput during handover between heterogeneous cells. Figure 9 shows the throughput during handover. The x-axis in the figure 9 represents the time, while the y-axis
represents the effective association throughput excluding the duplicate packets. As can be seen from figure 9, that despite the difference in the bandwidths of the paths, the association throughput achieved by cmpSCTP is close to the ideal throughput during handover. The high throughput achieved by cmpSCTP is due to its striping mechanism that is based on the rate of the bandwidth of the paths.

C. Sensitivity to the Schedule Strategy

In our simulation, we created a cmpSCTP association consisting two paths between the two multi-homed cmpSCTP hosts: cmpSCTP source and destination, which available bandwidth is 1Mbps and 2Mbps individually. As the above analysis in the section IV, we simulate the LS-SCTP or pTCP as n independent M/M/1 queues, and the cmpSCTP approaches to an M/M/n queue model, which the n is 2. The simulation result represents the association throughput and the average delay of every packet or packet’s sojourn time.

![Fig.10 Throughput performance using various multi-path transport protocol](image)

As can be seen from Figure 10, all three schemes show relatively small fluctuations in the throughput. The proposed cmpSCTP shows more high association throughput due to a series of efficient improvement mechanisms

![Fig.11 Delay performance using various multi-path transport protocol](image)

Fig.11 shows the delay performance by performing the concurrent multi-path transfer feature. We can see that the end-to-end delay during performing cmpSCTP is much lower than the other multi-path transport protocol, while in most situations it brings noticeable improvements to jitter, packet loss percentage and reordering delays at the receiver.

VI. CONCLUSION AND FUTURE WORK

In this paper, we proposed a cmpSCTP protocol which highly coupled with Mobile IP to keep two or more end-to-end paths concurrent transferring new data from a source to a destination host. The cmpSCTP distributes the data on the available paths based on an estimation of the available bandwidth of each path. We presented the design and details of the proposed approach, and evaluated its performance through simulation experiments. Our simulation results demonstrated that the cmpSCTP can lead to satisfactory performance which is able to utilize the available bandwidth efficiently. We compared the performance of cmpSCTP with LS-SCTP and pTCP. Results present that cmpSCTP dramatically outperforms LS-SCTP and pTCP in terms of throughput and delay especially in heterogeneous network environment.

Further investigation is planned to address some of the issues associated with impact studies on the other path factors such as packet loss rate, security and cost, and mechanisms to mitigate it. Also, the assumption of independent paths is being dropping off, we then plan to enable the sender to dynamically decide from either shared or distinct congestion control across paths through incorporating an end-to-end bottleneck detection mechanism [15]. The analysis and evaluation of these issues are our future work.

ACKNOWLEDGMENTS

This work was jointly supported by: (1) National Science Fund for Distinguished Young Scholars (No. 60525110); (2) National 973 Program (No. 2007CB307100, 2007CB307103); (3) Program for New Century Excellent Talents in University (No. NCET-04-0111); (4) Development Fund Project for Electronic and Information Industry (Mobile Service and Application System Based on 3G); (5) National Specific Project for Hi-tech Industrialization and Information Equipments (Mobile Intelligent Network Supporting Value-added Data Services). We express our gratitude to the Research Institute of China Mobile. Thanks also to Yan Zhang, Cong Liu and the anonymous reviewers for their helpful suggestions which helped to improve the paper.

REFERENCES