マルチメディア通信と分散処理 71−13
（１９９５．７．１３）

ATM用確認型データ転送プロトコルSSCOPの実装と評価

長谷川 輝之　長谷川 亨　加藤 聰彦　鈴木 健二

国際電信電話（株）研究所

ATM(Asynchronous Transfer Mode)技術の進展により、ATM網を介した高速なデータ転送の実現が求められている。ITU-TがB-ISDNシグナリング用に標準化したSSCOP(Service Specific Connection Oriented Protocol)は、ウィンドウサイズが大きく、選択再送機能を有するため、ATM網における高速データ通信に適している。そこで筆者等は、OSI通信などATM上での各種データ通信に使用することを目的として、SSCOPをUNIXワークステーション上に実装した。実装したSSCOPプログラムは、選択再送や送達確認などを効率的に行うことにより、140Mbpsの回線上で、最大100Mbps以上のスループットを達成している。本稿ではSSCOPプログラムの実装方式、ならびにその性能評価の結果について報告する。

Implementation and Evaluation of SSCOP Realizing Reliable Data Transfer over ATM Network

Teruyuki Hasegawa　Toru Hasegawa　Toshihiko Kato　Kenji Suzuki

KDD R & D Laboratories

Ohara 2-1-15, Kamifukuoka, Saitama 356, JAPAN

SSCOP (Service Specific Connection Oriented Protocol), which is originally defined for B-ISDN signaling, is a promising reliable data protocol over ATM networks, because of a selective retransmission mechanism, a large window size and so on. In order to realize various data communication over ATM network, we have implemented SSCOP on UNIX workstations. This implementation has achieved the high performance by introducing data structures and algorithms for efficient selective retransmission and receipt acknowledgment. The throughput of the SSCOP library is higher than 100 Mbps over an ATM LAN. This paper describes the details of the SSCOP library implementation and the results of performance evaluation.

1 はじめに

ATM(Asyncronous Transfer Mode) 技術の進展により、ATM 網を介した高速かつ高信頼なデータ転送の実現が求められている。このためには、送達確認や誤り回復機能を有する、ATM 用の確認型データ転送プロトコルが必要となる。ITU-T においては、ATM におけるシグナリング用のプロトコルとして標準化した SSCOP (Service Specific Connection Oriented Protocol) [1] を、ATM 上の OSI 通信などの、ATM 用の各種データ通信に利用することを推奨している。そこで筆者等は、ATM 上での各種データ通信の実現を目的として、UNIX ワークステーション上に SSCOP を実装した。この SSCOP プログラムは、実装・移植が容易であり、効率的な送受信処理による高いスループットを実現している。本稿では、この SSCOP プログラムの詳細な実装方法と、ATM LAN を用いた性能評価の結果について述べる。

2 SSCOP 概要

SSCOP は AAL(ATM Adaptation Layer) タイプ 5 の上位にインタフェースする確認型データ転送プロトコルであり、以下に示す機能を有する。

1. ユーザデータの順序保存
   送信側は、上位レイヤから渡されたユーザデータに順序番号を含む SSCOP トレイラを付加し、SD(Sequenced Data) PDU(Protocol Data Unit) を作成して送出する。受信側は、順序番号に従ってユーザデータを上位レイヤに渡す(デリバーする)。
2. SD の送達確認
   以下の 2 つの PDU により SD の送達確認を行う。
   ・送信側から送られた POLL PDU に対して、受信側が送出する STAT(Solicited Status) PDU
   ・誤った順序で SD が受信された場合に、受信側が送出する USTAT(Unsolicited Status) PDU
3. 選択再送による誤り回復
   受信側は、受信した SD の順序番号を検査することで、SD の紛失を検出する。紛失した SD の順序番号は、STAT または USTAT によって送信側へ通知する。この際、連続して紛失した SD のグループ (紛失 SD グループ) とそれに続く受信した SD のグループ (受信 SD グループ) を順に書き込む。送信側は、紛失を通知された SD のみを選択的に再送する。
4. フロー制御
   送信側は、STAT・USTAT 等で受信側から通知されるウインドウの上限値に基づいて、フロー制御に用いるウインドウを決定し、その範囲内で SD を送出する。

図 1に SSCOP の通信シーケンス例を示す。SSCOP では、SD と POLL に対してそれぞれ独立な順序番号を設けており、これらを用いて以下のように確認型のデータ転送を行う。

(1) 正しい順序番号 N(S) を持つ SD(図中の SD(0)) が転送されると、受信側は上位レイヤにデリバーする。
(2) 紛失した SD(1) は、次の SD(2) の受信により検知され、USTAT を用いて送信側へ通知される。USTAT には、次に受信すべき SD の順序番号 N(R)(=1)、受信ウインドウ N(MR)(=9)(ウインドウサイズ 8 を仮定している)、紛失 SD グループを示す List elements(={1,2}) が書き込まれている。送信側はこの USTAT 受信により、送達確認された SD(0) の解放、ウインドウの更新、紛失を通知された SD(1) の再送を行う。
(3) 送信側は、一定時間毎あるいは一定の SD 送出回数毎に、POLL を送出する。POLL は次に送出される SD の N(S) と、自身の順序番号 N(PS) を持つ。POLL(4,1) を受信した受信側は、次に受信すべき SD の順序番号 N(R)(=1)、対応する POLL 順序番号 N(PS)(=1)、受信ウインドウ N(MR)(=9)、紛失 SD グループ (=[1,2):SD(1) を表す) と受信 SD グループ (=[2,4)) を示す List elements={1,2,4} を含む STAT を、即座に送出する。
送信側はこの STAT 受信に対して、紛失 SD グループ (=[1,2)) に対する再送処理、受信 SD グループに対応する SD(2).SD(3) の解放、ウインドウの更新を行う。ただしこの例では、SD(1) が POLL(4,1) 送出後に再送されているため、新たな再送は行わない。これは、SD 送出時点で保持された POLL 順序番号と、STAT 中の N(PS) を比較することで判断される。
一方、POLL(7,2) に対応する STAT 受信では、STAT の N(PS)(=2) が SD(6) 送出時の POLL 順序番号 (=1) より大きいため、要求された SD(6) を再送する。

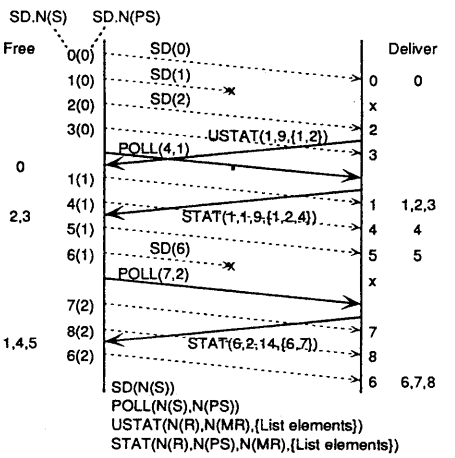

図 1: 通信シーケンス例

3 SSCOPの実装

3.1 実装方針

SSCOPを実装するにあたり、以下の方針を立てた[2]。

1. SSCOPは、実装・移植の容易でありかつ十分な性能が得られるライブラリ形式で実装する。
2. 明確なモジュール化を行ない、SSCOPのシグナル(以下プリミティブと呼ぶ)の送受信に応じた、上位レイヤとのプログラムインタフェースを持たせる。
3. 高スループット実現のため、上位レイヤとの間での不要なデータコピーを避ける方式を実現する。
4. 広帯域・高遅延のネットワークに適した広いウインドウサイズに対応可能とする。
5. 選択再送機能等のSSCOPの機能を効率的に実現するためのデータ構造や処理の流れ等を実装する。
6. SSCOPで要求される数10msecオーダのタイマ粒度を実現するため、独自のタイマルーチンを実装する。
7. 勧告で定められたSSCOPの仕様を全て実装する。
8. ネットワーク環境に応じて、ウインドウサイズ、送達確認の間隔等のSSCOPのパラメータが柔軟に変更できるよう実装する。

3.2 実装形態

図2にSSCOPライブラリを実装した通信プログラムの構成を示す。ここでは、SSCOPと上位レイヤプロトコルがライブラリとして実現され、メインルーチンを含む応用モジュールにリンクされる。本構成では、ATMコネクションの設定は上位レイヤプロトコルライブラリ(以下単に上位レイヤ)が直接ATMボード用デバイスドライバを介して行い、SSCOPライブラリは設定されたATMコネクション上でSSCOP PDUの送受信を行う。

また、高スループットを得るため、SSCOPライブラリは、応用モジュールおよび上位レイヤとバッファ制御部を共有し、レイヤ間でユーザデータをコピーしないバッファ制御方式[3]を使用している。このため、送信時は上位レイヤから渡されたバッファをそのまま使用する。また受信においては、SSCOPライブラリがバッファ制御部を通してSD受信のためのバッファを確保し、上位レイヤに渡す。

3.3 SSCOPライブラリのプログラム構成

3.3.1 概要

SSCOPライブラリのプログラム構成を図3に示す。SSCOPライブラリは以下のようなルーチン群から構成される。

1. プリミティブ送信用ルーチン群

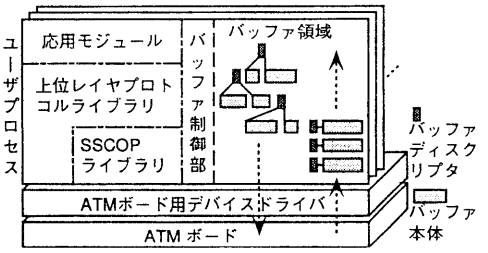

図2: 通信プログラムの構成

勧告の規定する要求／応答プリミティブに毎に、対応するSSCOP PDUの作成・送出を行なうルーチン。

2. プリミティブ受信用ルーチン群
   SSCOP PDUの受信に応じて指示／確認プリミティブの通知を行なうルーチン(受信用ルーチン)と、PDU毎の受信処理ルーチン。応用モジュールは、SSCOP PDUの受信を検知すると上位レイヤ経由で受信用ルーチンを呼び出し、全てのSSCOP PDUの受信処理を行わせる。このためこのルーチン群は、上位レイヤへのプリミティブ通知に加え、STATやUSTAT等の送出、選択再送処理等も行う。
3. タイマ管理ルーチン群
   現在時刻を定期的に更新し、必要に応じてタイマの発火処理を行わせるタイマ進行ルーチンと、タイマの起動停止などを行うルーチン。

また、SSCOPライブラリは、図3に示すように4種類のテーブルを持つ。コネクション管理テーブルは、コネクションごとの状態変数などを保持する。送信管理テーブルおよび受信管理テーブルはコネクション毎に用意されている。また、タイマ管理テーブルは起動されているタイマの情報を保持する。

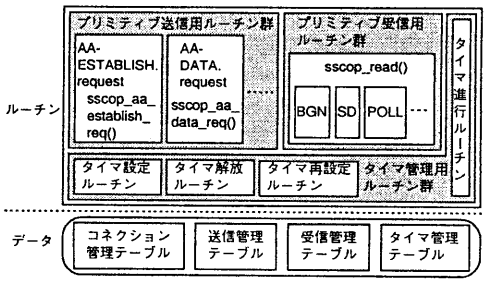

図3: SSCOPライブラリの構成

3.3.2 送受信管理テーブルの構造

送信管理テーブルは、図4に示すように送達確認されていないSDのうち最小の順序番号(VT(A))、次に送出するSDの順序番号(VT(S))、これまで送出したPOLLの順序番号(VT(PS))、ウインドウの上限(VT(MS))、

SD 管理テーブルに登録済みの SD の最大の順序番号 (VT(SH)) 等の送信用状態変数と、SD 管理テーブル、送達確認用 SD グループリストから構成される。

SD 管理テーブルは、未送出または送達確認されていない SD を格納するバッファへのポインタを保持する配列であり、参照を高速化するために、その添字を SD の順序番号に一意に対応させている。

送達確認用 SD グループリストは、選択再送を効率良く行うために、送達確認されていない連続した SD グループを管理する。個々のリスト要素はグループの先頭 SD の順序番号、末尾 SD の順序番号+1、並びに SD を送出した時点での POLL の順序番号 (N(PS)) を持つ。

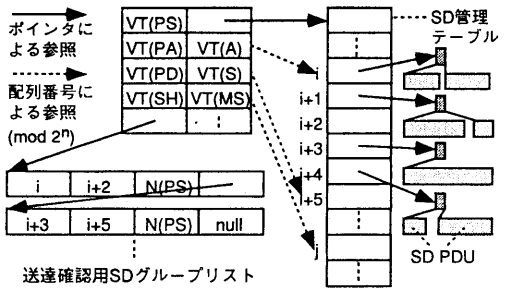

図 4: 送信管理テーブルの構成

受信管理テーブルは図5に示すように、次に受信されるべき SD の順序番号 (VR(R))、次に受信が期待される SD の順序番号 (VR(H))、ウィンドウの上限 (VR(MR)) 等の受信用状態変数と、SDU 管理テーブルおよび未受信 SD グループリストを保持する。SDU 管理テーブルは、SD 管理テーブルと同様の方式で受信済みでかつデリバーされていない SDU を登録している。未受信 SD グループリストは、VR(R) と VR(H) の間にある未受信 SD の順序番号を連続するグループ毎に管理するリストである。

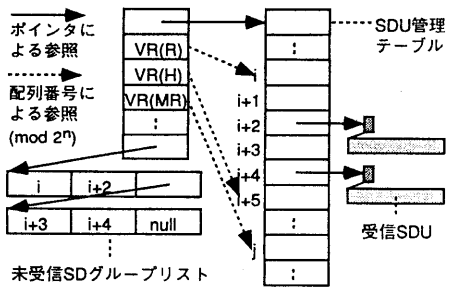

図 5: 受信管理テーブルの構成

3.3.3 プリミティブ送信用ルーチン群

プリミティブ送信用ルーチン群のうちユーザデータ送信 (AA-DATA.request) における動作は次の通りである。上位レイヤから AA-DATA.request に対応する sscop_aa_data_req() が呼び出されると、上位レイヤから受け取ったバッファ上の SDU にトレイラを付加して SD を作成し、SD 管理テーブルに登録する。SD の順序番号がウインドウの上限を越えない場合はその SD を送出する。さらに、前回の POLL 送出からの SD 送出数が MaxPD に達した場合は新たに POLL を送出し、更新された VT(PS) に対応するリスト要素を新たに作成して、送達確認用 SD グループリストの末尾に連結する。

3.3.4 プリミティブ受信用ルーチン群

受信用ルーチン sscop_read() は、PDU に応じた処理を行う。以下 STAT、USTAT、SD 受信時の動作について述べる。

(1) STAT 受信時

まず STAT が、通知する N(MR) により送信管理テーブルのウインドウの上限を更新する。そして、N(R)-1 までの順序番号を持つ SD に関する管理情報を、SD 管理テーブルおよび送達確認用 SD グループリストから削除する。さらに STAT が紛失 SD グループや受信 SD グループを持つ場合は、以下の処理を繰り返す。

紛失 SD グループに対する処理:

1. 紛失 SD グループに対応する、リスト要素を送達確認用 SD グループリストから検索する。
2. 検索したリスト要素の N(PS) と STAT 中の N(PS) を比較し、後者が大きい場合は SD 管理テーブルから SD を再送する。
3. 再送した SD に対して、対応する N(PS) の値を現在の POLL 順序番号 VT(PS) に変更する。リスト要素のすべての SD を再送しなかった場合は、新たに別のリスト要素を作成する。

受信 SD グループに対する処理:

1. 受信 SD グループに対応する、リスト要素を送達確認用 SD グループリストから検索する。
2. 受信 SD グループに含まれる SD をリスト要素から取り除く。
3. 必要に応じて新たなリスト要素を作成する。

さらに SD の再送後に必ず POLL を送出するオプション (poll after retransmission:PAR) が設定されている場合は、全ての SD 再送を終了した後に POLL を送出する。

(2) USTAT 受信時

再送を要求された SD のグループを SD 管理テーブルから無条件に再送し、送達確認用 SD グループリストの更新処理を STAT の時と同じ要領で行う。

(3) SD 受信時

VR(R) に等しい順序番号をもつ SD を受信すると、

受信した SD を SDU にして上位レイヤへデリバーする。このとき、その SD に連続した SDU が既に受信済みであるかを検査し、受信済みの場合はさらにデリバー可能な SDU があることを併せて通知する。一方、VR(R) と異なる順序番号をもつ SD を受信すると、上位にデリバーできないため、受信管理テーブルに保持しておく。そしてその順序番号が VR(H) より大きい場合、未受信 SD グループリストに新たなリスト要素を加え、その新たなリスト要素から USTAT を作成し送出する。

3.3.5 タイマ管理テーブルの構造

タイマ管理テーブルの構成を図6に示す。個々のタイマはそれぞれタイマテーブルにより表現され、各テーブルは発火時刻順に単方向リスト (タイマテーブルリスト) で管理されている。タイマテーブルにはタイマの属するコネクション (fd)、発火時刻、タイマの初期設定値 (timerval)、タイムアウト処理ルーチンへのポインタ (timeoutfp) が保持されている。タイマの識別子 (timerid) にはタイマテーブルへのポインタを用いる。

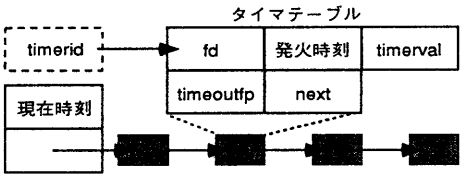

図 6: タイマ管理テーブルの構成

3.3.6 タイマ管理用ルーチン群

タイマ管理用ルーチン群は、送信モジュールおよび受信モジュールから呼び出されるタイマ設定、再設定、解放用ルーチンおよび上位レイヤから呼び出されるタイマ進行ルーチンにより構成される。タイマ設定時に呼び出される tm_settimer() は、タイマ管理テーブル内に新たなタイマテーブルを作成し、これをタイマテーブルリストの適切な位置に挿入し timerid を返す。上位レイヤから呼び出されるタイマ進行ルーチン tm_progress_timer() は、システムコールにより現在時刻を更新し、タイマテーブルリストの先頭からタイマの発火を調べる。タイマが発火していればタイマテーブル内の timeoutfp によりタイムアウト処理ルーチンを呼び出す。

4 SSCOP ライブラリの性能評価

4.1 実験形態

SSCOP ライブラリを評価するため、図 7に示す構成で通信実験を行った。ATM ボード (FORE Systems SBA-200) を有する 2 台の SUN SPARC staion20 (Solaris2.3、クロック 60MHz) を 140Mbps(TAXI) 回線により ATM スイッチ (FORE Systems ASX-200) に収容している。さらに遅延および伝送誤りを付加するため、データチャネルシミュレータ (ADTECH SX/13) を 45Mbps(DS3) 回線により ATM スイッチに接続した。

以下の条件で、SDU サイズおよびウインドウサイズ (SD 数) を変更して、100,000 個の SDU を転送し、そのスループットを測定した。

(1) 遅延および誤りの無い 140Mbps 回線のみで接続。
(2) 140Mbps 回線からの入力をデータチャネルシミュレータの接続されている 45Mbps 回線で中継し、200msec の往復遅延および BER=10^{-7}のランダムビット誤りを挿入。

なお (2) の条件においては、140Mbps 回線から 45Mbps 回線へ出力する際に ATM スイッチのバッファ溢れが起きないようにする必要がある。このため、ATM ボードのトラヒックシェーピング機能を用い、さらに送信 SDU の間隔を応用モジュールで調整することにより、送信レートを 40Mbps(45Mbps 回線におけるセルペイロード速度は約 40.7Mbps) 以下に抑制している。

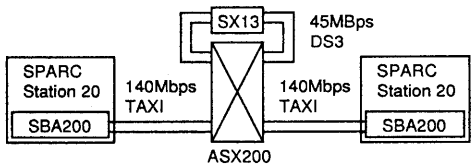

図 7: 実験構成

4.2 実験結果

実験結果を図8から図10に示す。図8は、条件 (1) におけるスループットを表しており、最大約 108Mbps のスループットを示している。図9ならびに図10は、条件 (2) において、伝送誤りを挿入しない場合と、した場合のスループットを示す。

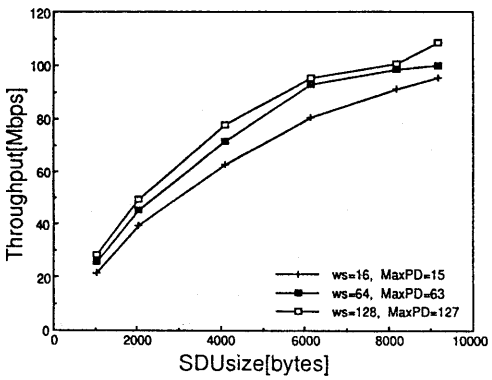

図 8: 実験結果 (遅延無し、伝送誤り無し)

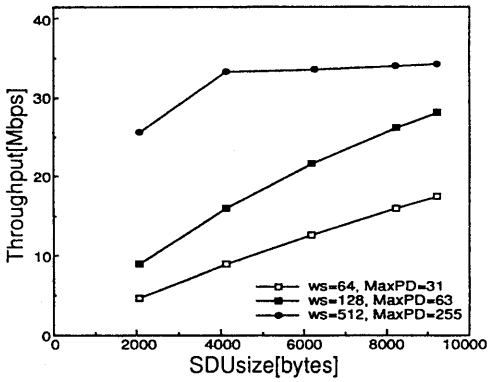

図 9: 実験結果 (200msec 往復遅延、伝送誤り無し)

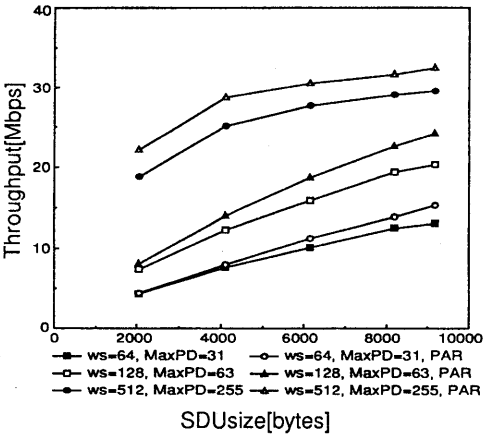

図 10: 実験結果 (200msec 往復遅延、伝送誤り有り)

5 考察

(1) 送受信管理テーブルでは、SD または SDU のバッファを管理する SD(または SDU) 管理テーブルを配列として実現し、一方、送達確認されていないまたは受信されていない SD のグループをリスト (送達確認用ならびに未受信 SD グループリスト) により管理している。この構成は、以下の理由より決定した。

- 広帯域・高遅延な網に対処するためには、大きなウインドウサイズを用いる必要がある。そこで、送達確認による SD 用のバッファの解放および選択再送を行うために、ウインドウの範囲の任意の SD 用バッファに直接アクセス可能とすることが望ましい。また、SDU 管理テーブルについても、ウインドウの範囲内の SD を受信した時、その SDU のバッファを登録する位置を SD の順序番号から直接決定できることが要求される。
- SSCOP に従った選択再送を行うためには、送達確認されていない各 SD に、送出時の POLL 順序番号を対応づける必要がある。再送を要求された場合の、POLL 順序番号の比較処理および再送後の更新処理を効率的に行うためには、送達確認用 SD グループをリストとするのが望ましい。また、SD 受信時の未受信 SD グループについては、STAT および USTAT PDU の作成処理を効率化するために、リストを用いるのが有効である。

(2) 実装した SSCOP ライブラリは、全く変更なしに、SunOS4.1.x、Solaris2.3、IRIX5.2 上で動作した。これは、シグナルなどの機種に依存した機能を使用しない設計を行ったためと考えられる。

(3) 図 8に示すように、遅延を挿入しない場合にも、高スループットを実現するには、十分なウィンドウサイズが必要である。これはプロトコル処理の遅延時間によるものと考えられる。例えば、SDU サイズ 8K バイトの場合に 100Mbps 程度のスループットを得るためには、64 個以上のウィンドウサイズが必要である。

(4) SSCOP では、ウインドウ内の最後の SD を送出するまでに、ウインドウ内の最初の SD を送達確認する STAT を受信できない場合は、送出待ちが生じる。200msec の往復遅延においては、図 9に示すように、ウィンドウサイズが 512 で SDU サイズが 4K バイト以上の場合以外は、この送出待ちが発生していると考えられる。この場合は、SDU サイズの増加に伴いスループットが向上する。

(5) 図 10に示すように、伝送誤りが存在する場合、SD 再送後に必ず POLL を送出する poll after retransmission(図中 PAR) が、スループットの向上に有効である。

6 おわりに

本稿では、SSCOP のワークステーション上への実装、ならびに ATM LAN を用いた性能評価について述べた。実装した SSCOP ライブラリは不要なデータコピーを避け、広いウインドウサイズを用いた場合にも、送達確認や再送処理を効率的に行うデータ構造を採用している。この結果 140Mbps 回線上において 100Mbps 以上のスループット、また 200msec の往復遅延および BER=10^{-7}のランダムビット誤りを挿入した 45Mbps 回線上においても 32Mbps のスループットを達成している。さらに本ライブラリは 200msec の往復遅延のある 155MbpsOC3 回線上で 95Mbps 程度のスループットを達成している。最後に日頃御指導頂く KDD 研究所浦野所長に感謝します。

参考文献

[1] ITU-T, "Service Specific Connection Oriented Protocol," Final Draft Recommendation Q.2110, December 1993.

[2] 長谷川, 長谷川, 加藤, 鈴木, "ATM 用確認型データ転送プロトコル SSCOP の実装に関する一考察" 情処全国大会, 5C-7, September 1994.

[3] 加藤, 井戸上, 鈴木, "OSI プロトコル実装のためのユーザデータをコピーしないバッファ制御方式," 情処マルチメディア通信と分散処理研究会, 62-13, September 1993.