

強化学習によるネットワークの状態変化予測に基づく最適リアルタイムビデオストリーム伝送システムの研究開発

代表研究者	長 田 智 和	琉球大学工学部情報工学科助教
共同研究者	玉 城 史 朗	琉球大学工学部情報工学科教授
共同研究者	谷 口 祐 治	琉球大学総合情報処理センター講師

1 はじめに

近年、通信回線の高速化を背景に、リアルタイムストリーミングシステムによるコミュニケーション通信がインターネットを用いて可能となった。最近では、人材が確保しにくい過疎地でも、十分な教育や医療が受けられるよう、リアルタイムストリーミングシステムを用いた遠隔教育[1]や遠隔医療[2]などの実用性が検討されている。

一方、これまでのインターネットアプリケーションは、確実にデータを送信することが重視されてきた。しかし、情報通信技術委員会[3]が定めた音声品質に関する JJ201.01 ガイドラインによると、片道伝送遅延は携帯電話程度の品質で 150ms とされている。リアルタイムストリーミングシステムを有効に利用するためには、ストリーミングコンテンツを確実に、かつ時間内に送信する必要がある。

また、これまでにさまざまな通信プロトコルが開発されているが、その特性には一長一短がある。TCP(Transmission Control Protocol)は到達信頼性の高い通信プロトコルであり高品質の通信が行えるが、伝送路遅延やエラーが高いネットワークではスループットが抑制される傾向がある。一方、UDP(User Datagram Protocol)は伝送路遅延が大きなネットワークにおいて高スループットの通信が行えるが、標準ではエラー訂正機能が備わっていないため、伝送路エラーの高いネットワークでは品質が低くなる。さらに、UDP に FEC(Forward Error Correction)機能を付加した研究[4]や、UDP に ARQ(Automatic Repeat reQuest)機能を付加した研究[5]がある。しかし、FEC を用いた場合には必要な通信帯域が大きくなるため、ネットワークの輻輳を助長する可能性がある。また、ARQ を用いた場合には伝送路遅延の影響を大きく受けるため、遅延の大きなネットワークでは効果的ではない。伝送路上でのエラーや遅延などのネットワーク状態は通信プロトコルの性能に影響し、ネットワークの状態により適した通信プロトコルは異なる。

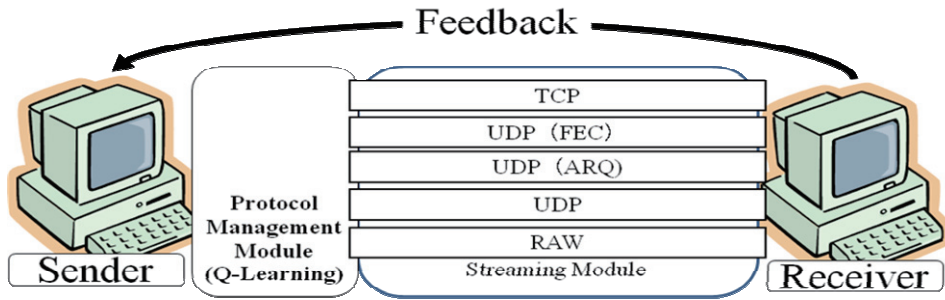
ネットワーク状態は利用環境で異なるため、全てのネットワーク状態において、時間制限と確実性を満たす通信プロトコルは存在しない。そこで、ネットワークの状態に応じて適した通信プロトコルを利用するリアルタイムストリーミングシステムを提案する。ここで、QoS(Quality of Service)制御装置や WAN 高速化装置などの通信装置は、通信プロトコルの性能に影響を及ぼすため、ネットワーク状態について閾値を設ける単純な通信プロトコル選択制御は汎用的に用いることができない。そこで本稿では、通信プロトコル管理に強化学習を用いることで、ネットワークの状態が変化しても、常に適した通信プロトコルを選択するシステムを開発し、その有効性を検証する。

2 提案システム

提案システムは、「ストリーミングモジュール」と「プロトコル管理モジュール」から構成される(図 1)。ストリーミングモジュールでは、プロトコル管理モジュールで選択されたプロトコルを自由に切り替えて、ビデオストリーミングを行う。プロトコル管理モジュールでは、強化学習を用いて、ネットワーク状態に適した通信プロトコルを選択する。

提案システムの実装には、ストリーミングコンテンツに、NTSC 規格転送速度約 35Mbps の Digital Video(以下、DV)を用いた。DV は、MPEG などと比較して、フレーム間圧縮を行わないシンプルな規格となっているため、データ転送にかかる性能の向上を測定するのに適していると考えたためである。

また、本論文では、2 種類のエラー表現を用いている。「伝送路エラー」は、ネットワーク経路上でのエラーを示し、「再生エラー」は、受信側の再生時間までに訂正できなかったエラーを表す。



(図 1: プロトコル選択型ストリーミングシステムの概要)

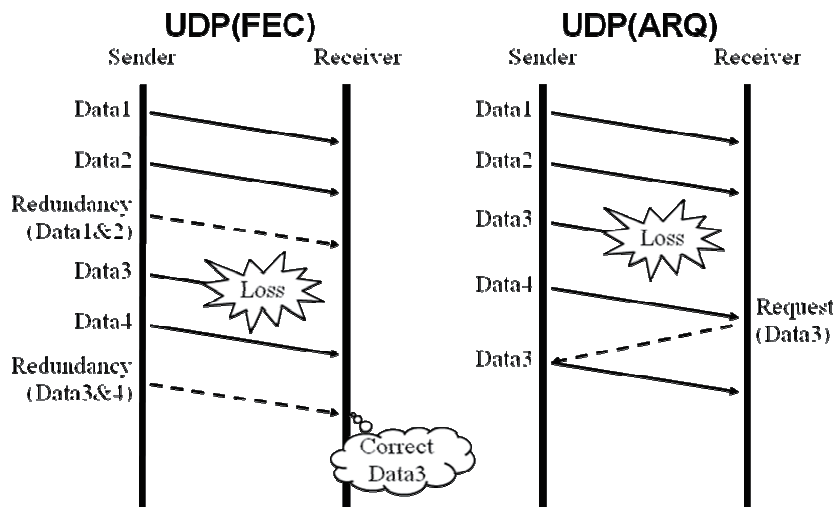
2-1 ストリーミングモジュール

ストリーミングモジュールでは、動画配送を行う。情報通信技術委員会が定めた、JJ-201.01 ガイドラインより、円滑なコミュニケーションを行うための音声通信遅延は 150ms が目安となる。これを考慮して、受信側では 4 フレーム分 (約 130ms) のバッファリングを行い、順序補正やエラー訂正を行う。

本システムでは、選択可能なプロトコルに、TCP, UDP, UDP に FEC を付加した通信方式 (UDP (FEC)), UDP に ARQ を付加した通信方式 (UDP (ARQ)), RAW データ通信の 5 つを用意した。それぞれ、送信側ホストと受信側ホストの間に仮想トンネルを保持し、プロトコル管理モジュールをからの支持に応じてシームレスに切り替えて通信を行う。なお、RAW データ通信に関しては通信装置の処理 (FEC, ARQ を組み合わせたプロトコルも研究されている) に依存し、RAW データの階層としてイーサフレームレベル (以下、RAW (ETH)), もしくは、IP パケットレベル (以下、RAW (IP)) が排他的に選択される。

UDP では RTP と組み合わせて利用し、到達順序補正や前のフレームを利用したパケット破棄対策を行う。UDP (FEC) では (15, 10) の Reed Solomon 符号により冗長化を行い、理論値で 50% のエラーが訂正できる。UDP (ARQ) では NACK ベースの ARQ を採用し、受信バッファリングの時間内に可能な限りエラー訂正を行う。伝送路エラー検知に 40ms 程度かかり、受信バッファリング時間を 130ms 程度としているため、最大再送試行回数は約 3 回となる。(図 2)

RAW データ通信では、TCP などのヘッダを付加せず、コンテンツデータをネットワークに送信できる最大サイズに分割し送信する。ただし、アプリケーション識別を行うために、RAW (ETH) ではイーサフレーム毎に、RAW (IP) では IP パケット毎に 4 バイトのアプリケーションヘッダを付加している。また、順序およびエラー補正を行わないため、再生フレーム中にエラーがある場合にはそのフレームごと破棄する。このため、伝送路エラーに対して極端な動画質の劣化を引き起こすが、選択可能通信プロトコル中で処理負荷およびネットワーク負荷は最も低くなる。



(図 2: FEC および ARQ のフローチャート)

2-2 プロトコル管理モジュール

送信側ホストは受信側ホストよりフィードバックレポートとして受信側エラー率, RTT, ジッタを取得する。これらの情報から, ネットワーク状態を判断して, それ適したプロトコルを動的に決定する。ここでは, 本システムが適用するネットワーク特性と, 強化学習の適応方法, 報酬の割り当てについて説明する。

2-2-1 ネットワーク周期

伝送路エラー率や遅延などネットワークの状態は, ネットワーク使用率の最繁時間などに大きく依存する。ネットワークにより使用率最繁時間は異なるが, 一般的にネットワーク使用率の最繁時間は決まっており, 例えば, 一般的な LAN の場合には, 出勤直後, 昼休み前後, 帰宅直前の 3 点を持つと言われている。すなわち, ネットワークの状態には周期性が存在すると考えることができる。ネットワークの状態に周期性がある場合, ネットワークの現状態とその状態が変移したという情報から, 次の状態を決定することができる。このため, 周期性のあるネットワークの状態も MDP である。

従って, そのネットワークの状態が周期性を持つならば, ネットワークの状態は MDP となり, この対となる最適な通信プロトコルも MDP となる。MDP であるならば, Q-learning を用いることで最適な行動を決定することが証明されており, 強化学習を用いて効率よく適した通信プロトコルを選択することができる。また, ネットワークの状態が突然変化した場合には, 再度学習を行うことにより, 新たに適した通信プロトコルを選択することが可能である。以上のことから, 提案システムの通信プロトコル選択には強化学習を用いることを提案する。なお, 本システムでは管理を単純化にするため, 同じ状態を 2 度以上通らないシンプルな周期性を対象とする。同じ状態を 2 度以上通る複雑な周期性を対象とする場合には, 準 MDP や部分 MDP を考慮する必要がある。このため, 本システムでは, 複雑な周期性があるネットワークには, 再学習を行うことで対処する。

2-2-2 教科学習の適用

本論文では, 通信プロトコル管理に n ステップ予測を用いた Q-Learning を用いる。環境が MDP に属する場合, Q-learning を用いることで最適な行動選択を行えることが証明されている [6]。提案システムでは, 状態として 5 つ (UDP, UDP (FEC), UDP (ARQ), RAW データ通信), 行動として現状維持および他の通信プロトコルへの切り替えの 5 つを対応させる。したがって, 提案システムでは 25 個の Q 値をもつ。この Q 値を推定することで使用するネットワーク状態に適した通信プロトコルを選択する。

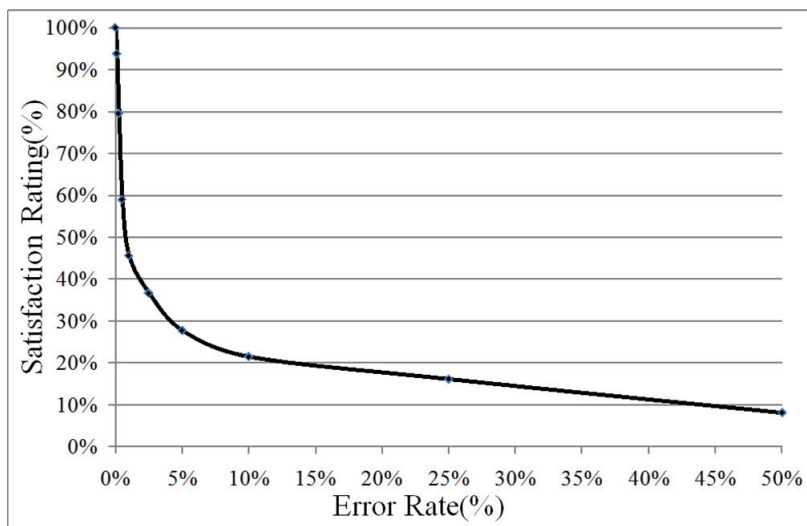
n ステップ予測 Q-learning の行動価値関数は式 (1) で求める。ここで, R_{sum} は複数ステップ間の割引報酬合計であり, 式 (2) で求める。提案システムでは, 方策にボルツマン選択を用いる。ボルツマン選択は式 (3) によりそれぞれの行動価値から次行動として選択する確率を算出する。

$$Q(s, a) \leftarrow (1 - \alpha)Q(s, a)r + \alpha [R_t + \gamma \max_{a'} Q(s', a')] \quad (1)$$

$$R_{\text{sum}} = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \dots + \gamma^{N-1} r_{t+N-1} \quad (2)$$

$$\pi(s, a) = \exp(Q(s, a)/T) / \sum_b \exp(Q(s, b)/T) \quad (3)$$

ここで, T はボルツマン温度を表し, T が大きいほどランダムな行動選択を行い, T が 0 に近いほど行動価値が最大の行動を選択するグリーディー方策を行いやすくなる。方策にボルツマン選択を用いる場合には, ボルツマン温度のスケジューリングが重要になる。そこで, 提案システムではシステムの起動直後は全ての行動価値が未知であるため, T を高い値に設定し, ランダムに行動を選択する。ある程度行動価値が推定の精度が高くなると, T を中間値に設定し, 行動価値の大きさに比例して確率的に選択する。行動価値が定まり, 行動価値の更新変化量が少なくなると, T を低い値に設定し, グリーディー方策を行う。また, 急激な RTT やエラー率の変化などが起こると, ネットワークの状態が変化したとみなし T を一時的に高く設定し, 行動価値の再学習を行う。



(図 3: ストリーミングコンテンツの満足度評価)

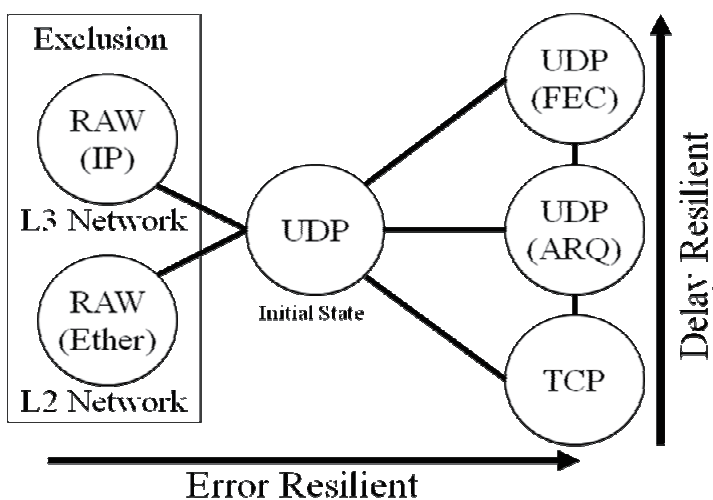
また、各通信プロトコルにより、ネットワーク負荷が異なる。ネットワーク負荷が高まれば輻輳が起きる要因となるため、同一の報酬が得られるならば、ネットワーク負荷が最も軽い通信プロトコルが最適である。従って、通信プロトコル別にネットワーク負荷を考慮する。さらに、時系列的な受信側エラー率の変動を考慮し、報酬は過去の平均報酬(式(5))を利用した、式(4)により算出する。

$$\begin{aligned}
 r &= (1 - \beta)(r_b + L) + \beta\{(r_b + L) - r_{avg}\} \\
 &= (r_b + L) - \beta_{avg}
 \end{aligned}
 \tag{4}$$

$$r_{avg} = \text{AVG}(r_b + L)
 \tag{5}$$

ここで、 r_b は基礎報酬であり、図3をもとに $-100 \leq r_b \leq 100$ の値をとる。 β は過去の状態に対する関連重みを表し、 $0 \leq \beta \leq 1$ の値をとる。 β の値が大きいほど過去の状態を重視した学習を行う。 L はネットワーク負荷に基づく調整値であり、最もネットワーク負荷の高いUDP(FEC)を-10、最も軽いRAWデータ通信を10とし、その他の通信プロトコルは0としている。

また、強化学習の収束速度を向上させるために、通信プロトコル間の親和性を考慮する。親和性の低い通信プロトコル同士は行動価値をマイナスとし、直接遷移しない設定としている。この行動価値テーブルに対応させた通信プロトコルの状態遷移図を図4に示す。



(図 4: 通信プロトコルの状態遷移図)

以上の条件にシステムを構築し、nステップTD予測Q-learningにより行動価値を推定することで、ネットワークに適した通信プロトコルを管理する。

3 実験結果

提案システムの有効性を検証するため、検証実験を行う。検証実験システム構成を図5に、検証実験の使用機器を表1に示す。送信側ホストはIEEE1394を用いてDVカメラと接続されており、ストリーミングコンテンツを取得する。送信側ホストと受信側ホストは中継器を経由してネットワークで接続されている。受信側ホストはIEEE1394を用いてDV機器と接続されており、ネットワークから取得したストリーミングコンテンツをDV機器で再生する。実験で使用するケーブルは全て100Base-Tであり、実験中に利用する通信帯域幅よりも十分に大きい。送信側ホストと受信側ホストをつなぐ中継器は、擬似ネットワークとして動作する。擬似ネットワークでは伝送路エラー率、伝送路遅延、帯域制限のエミュレーションを行う。ここで、伝送路エラー率は2項分布に従い、伝送路遅延はジッタを起こさず各設定値で一定とする。送信側ホストでプロトコル管理モジュールを動作させ、ネットワークの状態によって適切な通信プロトコルが選択されるか検証する。

検証実験に用いる3つのシナリオを表2に示す。高品質なネットワークを想定したシナリオ、帯域制限により恒常的に輻輳が発生しているシナリオ、周期的にネットワーク状態が変化するシナリオとなっている。なお、RAWデータ通信としてはRAW(ETH)を選択する。各シナリオにおいて、時間経過に伴う通信プロトコルの変化と受信側エラー率を観測する。

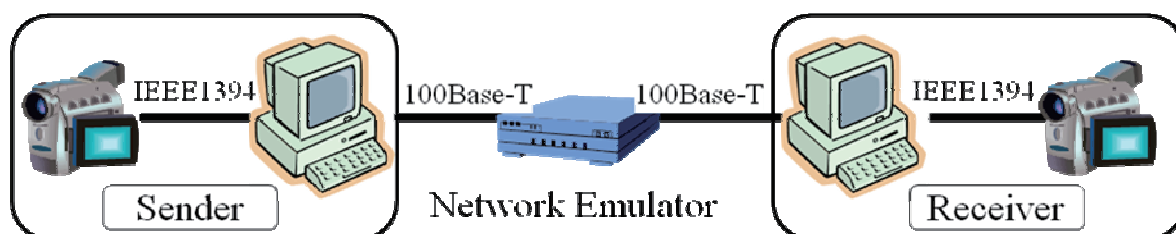
また、各シナリオにおける適した通信プロトコルを明らかとするため、検証実験で用いる全てのネットワーク状態で各通信プロトコルを用いてストリーミングを行った場合の平均受信側エラー率を表3に示す。

(表1: 実験環境)

	OS	CPU	Memory	DV Equipment
Sender	VineLinux3.2	Celeron 2.4GHz	512MB	DIGICAM(※1)
Receiver	VineLinux3.2	Pentium4 3.2GHz	1GB	HANDYCAM(※2)
Emulator	VineLinux3.1	Celeron 2.0GHz	768MB	-

(※1)DIGICAM: Panasonic DIGICAM(NV-GS100K-S)

(※2)SONY: HANDYCAM(DCR-HC40)



(図5: 実験環境)

(表2: 実験ネットワーク状態)

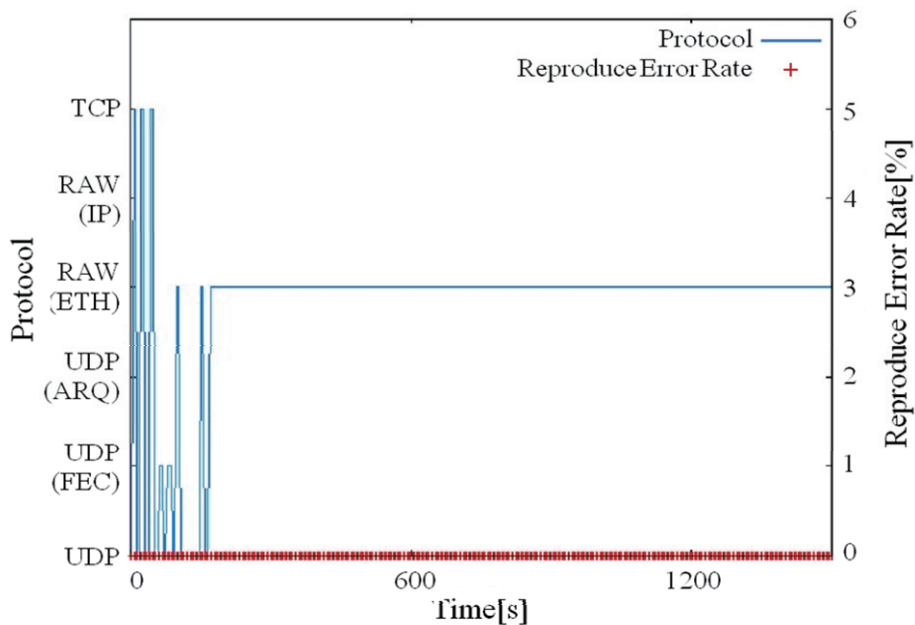
	Switch	Error Rate	Delay	BandWidth
Condition 1	None	0%	1ms	100Mbps
Condition 2	None	0%	1ms	25Mbps
Condition 3 (1)	360s Cycle	0.3%	65ms	100Mbps
(2)		0.3%	15ms	40Mbps

(表 3: ネットワーク状態ごとの受信側の平均受信エラー率)

Network Condition	UDP	UDP (FEC)	UDP (ARQ)	RAW DATA	TCP
Condition 1	0%	0%	0%	0%	0%
Condition 2	11.9%	52.8%	12.2%	78.2%	8.1%
Condition 3 (1)	0.3%	0%	0.3%	22.3%	48.2%
(2)	0.3%	10.3%	0%	22.5%	36.5%

3-1 ネットワーク状態 1 (高品質) での結果

図 6 が高品質なネットワークを想定したシナリオのシミュレーション結果となる。この図の横軸は時間経過、縦軸の主軸は通信プロトコルの状態、縦軸の第 2 軸は受信側エラー率を示している。また、図中で通信プロトコルの状態変化は実線、受信側エラー率は “+” により表されている。また、システム起動直後は行動価値を推定するため、通信プロトコルを頻繁に切り替えているが、ある程度学習が進むと RAW が選択されるようになっていく。

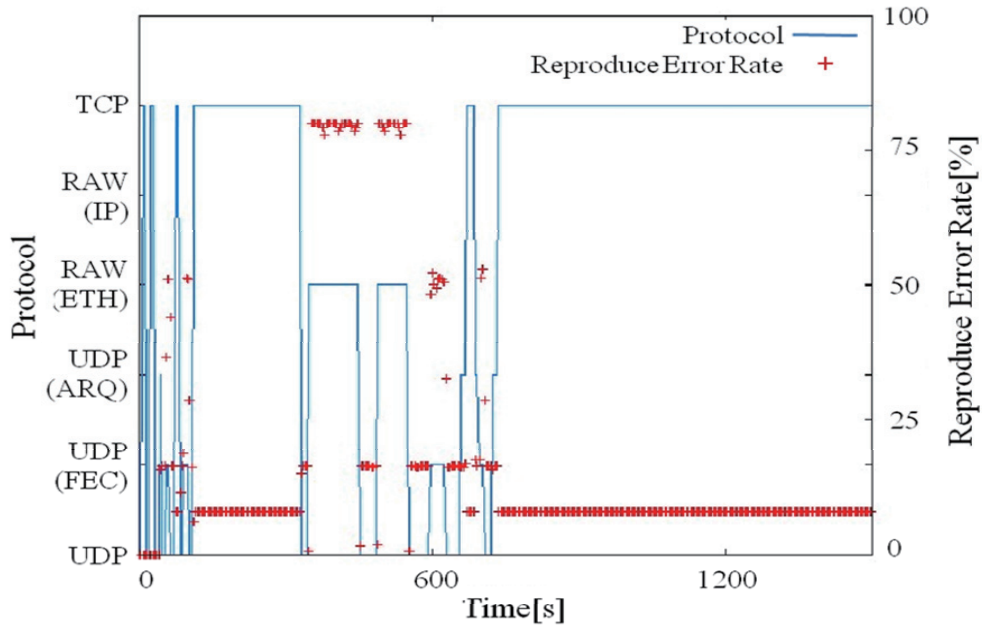


(図 6: 実験結果(ネットワーク状態 1))

3-2 ネットワーク状態 2 (輻輳状態) での結果

図 7 が輻輳により恒常的に受信側エラーが発生するネットワークを想定したシナリオのシミュレーション結果となる。表 3 から、シナリオ 2 における最適な通信プロトコルは TCP である。選択可能な通信プロトコルでは TCP は唯一輻輳制御が行える方式であり、TCP では受信側エラー率は 8% でほぼ一定となる。他の通信プロトコルでは、輻輳が発生してもそれを感知できず、データを送信し続けるため、受信側エラー率に幅が出ており、平均的な受信側エラー率が TCP よりも高い傾向となる。

これらの要因から、輻輳が発生するネットワークでは、輻輳制御機能がある TCP が適した通信プロトコルと判断される。

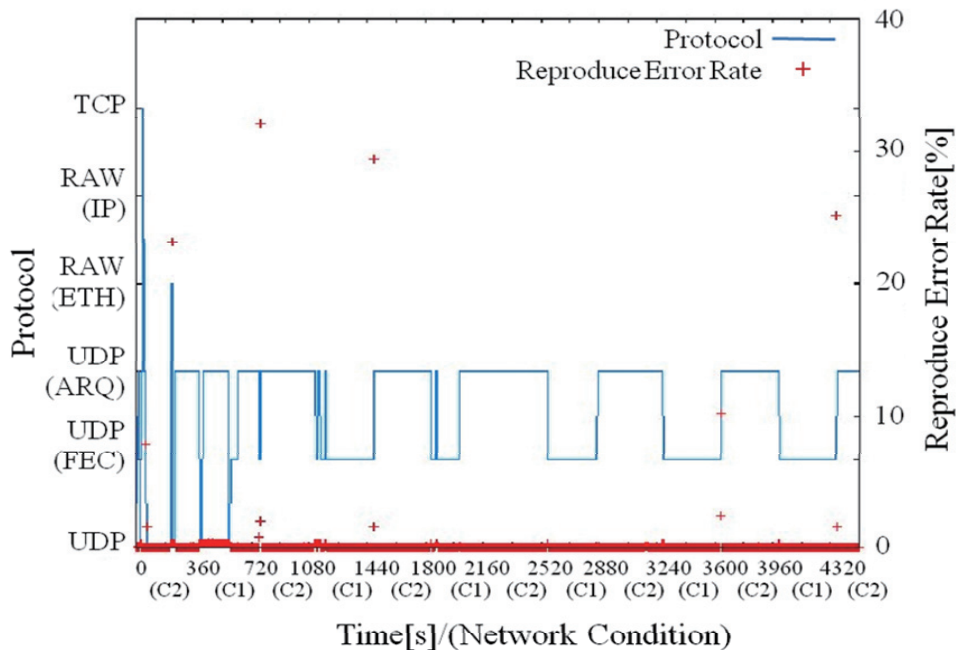


(図 7: 実験結果(ネットワーク状態 2))

3-3 ネットワーク状態 3 (周期的) での結果

図 8 は周期的に状態が変化するネットワークを想定したシナリオのシミュレーション結果となる。本シナリオではネットワーク状態が切り替わるため、横軸に時間経過とネットワーク状態の 2 つの情報を示す。なお、図中の C1 はシナリオ 3 のケース 1(表 2 の 3(1))を表し、C2 はシナリオ 3 のケース 2(表 2 の 3(2))を表す。表 3 から、ネットワークの状態が S1 の場合には UDP (FEC) が適した通信プロトコルとなり、ネットワークの状態が S2 の場合には UDP (ARQ) が適した通信プロトコルとなる。図 8 より学習が安定状態になると、ネットワーク状態が S1 では UDP (FEC)、S2 では UDP (ARQ) を用いてストリーミングが行われている。

以上の結果より、初期段階では行動価値を推定するための試行錯誤を行うが、学習が進むにつれ使用する通信プロトコルが安定し、ネットワークの状態が変化する毎に、適した通信プロトコルを選択することが確認された。



(図 8: 実験結果(ネットワーク状態 3))

3 まとめと今後の課題

本稿では、状態が変化するネットワークにおいて快適なリアルタイムストリーミングサービスが行えることを目的として、ネットワークの状態に応じて動的に通信プロトコルを変更するシステムを提案した。

提案するストリーミングシステム特有の機能は、通信プロトコルが変更可能なストリーミング機能、ネットワーク状態により適した通信プロトコルを選択する通信プロトコル管理機能である。選択可能な通信プロトコルは、TCP, UDP, UDP (FEC), UDP (ARQ), RAW データ通信の 5 種類を用意し、シームレスな切り替えが可能である。また、通信プロトコル管理機能は、強化学習の 1 つである n ステップ予測 Q-learning を用いる。通信プロトコルの選択を行動価値として、伝送路遅延, RTT, ジッタ, 伝送路エラー率に基づき、行動価値を推定することにより、ネットワーク状態に適した通信プロトコルを管理する。

最後に、提案システムの有効性を検証するため、シミュレーション環境を構築し、検証実験を行った。実験では、伝送路遅延, 伝送路エラー率, 帯域制限が行えるエミュレータを用いて、実際にネットワーク状態の変化に伴い適した通信プロトコルを選択していることを確認した。

今後の展望として、複数の有効な通信プロトコルやコーデックの選択を可能とし、ネットワークの状態に応じてコーデックや通信プロトコルを自動で管理するシステムなどが考えられる。

【参考文献】

- [1]Aiguo He, Atsushi Kara, Zixue Cheng, Kentaro Go, Akio Koyama, Tongjun Huang, Atsumi Imamiya : RIDEE-SPS: A Presentation System for Real-time Interactive Distance Education Environment, IPSJ Journal Vol.44, No.03, pp.700.708 (2003).
- [2]Kumiko Ohashi Yuichiro Gomi, Isao Okada, Mamoru Watanabe, Naoya Sakamoto, Hiroshi Nagata, Hiroshi Tanaka : Real -time Transmission of Endoscopic Motion Pictures over the BroadBand Internet, IPSJ Journal Quality Aware Internet 2003-QAI-7(2), pp.7.12 (2003).
- [3]The Telecommunication Technology Committee, [http://www.ttc.or.jp/ JJ-201.01](http://www.ttc.or.jp/JJ-201.01) A Method for Speech Quality Assessment of IP Telephony
- [4]Tohru Kondo, Kouji Nishimura, Reiji Aibara, Kaori Maeda, Tamaki Ot-suka: Performance Evaluation of FEC for High Quality Video Transmis-sion, IPSJ Journal Vol.45, No.1, pp.84.92 (2004).
- [5]Hiroaki Hagino, Yuichiro Miyazaki, Yuko Onoe, Yukio Atsumi, Masaaki Taniguchi,Hirotaka Komaki, Nagatsugu Yamanouchi: Playtime-oriented Retransmission Control Method for Multimedia Streaming IPSJ Journal Vol.45, No.2, pp.402.411 (2004).
- [6]C.J.C.H. Watkins, P. Dayan: “Q-learning”, Machine Learning , 8 , 279-292 (1992)

〈発 表 資 料〉

題 名	掲載誌・学会名等	発表年月
Protocol-Selective Streaming System based on Network State	Information Journal, Vol.11, No.1, pp.77-86, 2008.	2008/1
MPEG2 を用いたプロトコル選択型ストリーミングシステムの開発	第30回情報理論とその応用シンポジウム, Nov. 2008.	2007/11