L_004

大規模な管理専用IPネットワークのための機器負荷を考慮した経路制御プロトコル

Load Sensitive Routing Protocol for Large-scale IP Network Management

堀 賢治† 吉原貴仁† 堀内 浩規†
Kenji Hori Kiyohito Yoshihara Hiroki Horiuchi

1. はじめに

昨今のIPネットワーク大規模化はルータ，スイッチ，ホストといった管理対象機器の総数を増大させ，管理者数の逼迫や，機器故障等による通信障害箇所の増加を招いている．また機器の多様化により，全ての機器が潤沢なメモリを備えることは必ずしも仮定できなくなっている．管理者数の逼迫に対応するため，snmpやsshといったIP層の遠隔管理プロトコルを用い，管理者の操作する管理サーバから機器を集中一括して管理する手法が用いられるが，管理者がIPアドレスを設定し，経路制御プロトコルによりIP経路を設定する必要がある．しかしながら，通信障害箇所検出および迂回経路によるパケット再送可能な既存経路制御プロトコルは，機器の総数増加とともに，1) 機器間接続状態情報の増加，および2) 瞬間的なパケット欠落による多重再送用パケット保持量の増加のため，各機器のメモリ負荷が増大する傾向にある．このため本稿では i) 経路集約可能となるようなIPアドレスの自動割当てを行うことで各機器が保持する機器間接続状態情報を削減し，更に ii) 障害箇所と見なさないパケット欠落の発生率許容範囲を管理者が設定可能とすることで多重再送用パケットの保持量削減を図った，機器負荷を考慮した経路制御プロトコルを提案する．

2. 想定環境と要件

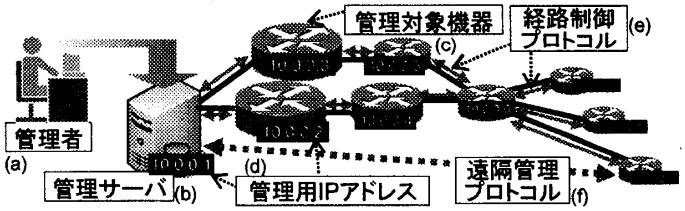

図1 想定環境

図1に本稿の想定環境を示す．管理者(図1(a))は管理サーバ(図1(b))と管理対象機器(図1(c))とに予め一意なIPアドレス(図1(d))を割当て，さらに経路制御プロトコル(図1(e))によって管理サーバと機器との間にIP経路を確立させる．その後，遠隔管理プロトコル(図1(f))によって機器を集中一括的に遠隔管理する．以下，単に「機器」といった場合，管理サーバと管理対象機器との和集合を表すものとする．

尚，遠隔管理にはスループットや低遅延より安定性・頑健性がより重要であることから，遠隔管理通信とその他の通信(例えばユーザトラヒック)とは，必ずしも同一のIPアドレス空間やIP経路を利用するとは限らない．このため本稿以下では，機器には遠隔管理専用のIPアドレス(管理用IPアドレス)が割り当てられるものとし，管理用IPアドレスによる経路制御を考える．尚，既存の充実した遠隔管理アプリケーションを利用可能とするため，TCPやUDPといった既存トランスポート層プロトコルを変更しないことが想定される．

このような想定環境において，ネットワークの大規模化に伴い経路制御プロトコルには以下の要件が求められる．

(要件1) 機器数が増加するとともに，故障機器数と，それに伴う通信障害箇所も増加する．よって遠隔管理を安定して行うために，障害箇所を検出・迂回可能であること．

(要件2) 管理対象機器の多様化により，それらのメモリ搭載量は必ずしも潤沢であるとは言えなくなるため，メモリ負荷は極力抑えられていること．

尚，機器の停止に至るような沈黙型の障害であれば，経路制御プロトコルによる経路情報の広報が行われず遠隔管理通信が当該機器を経由し得ないため，迂回処理を行う必要がない．よって上記要件1で障害箇所とは，例えば物理インターフェース故障やソフトウェアバグ等により，機器が停止することなく定常的または間欠的にパケット転送が支障を来たしているといった，非沈黙型の障害箇所を想定している．

3. 障害箇所迂回可能な既存経路制御プロトコルとその問題

既存の代表的な経路制御プロトコルの中でも，[1]は機器間の接続関係(リンクステート)情報を基に，送信元明示的経路指定(ソースルーティング)により全てのパケットを送信する．パケットには送信元機器内で固有なシーケンス番号が付与され，パケットを中継した機器がこのシーケンス番号を記述した中継確認パケット(*ACK*)を送信元へと送信することで，特定のパケットが障害箇所に遭遇したこと検出できる．また送信および再送したパケットを，*ACK*を受信できるまでメモリ上に複写保持しておくことで，障害箇所検出時には，それを迂回する経路指定を行って再送し，パケットを宛先機器へと確実に到達させることができ，上述の要件1を満たす．しかしながら[1]は，上述の要件2の観点から，以下(1)，(2)のメモリ負荷増大に繋がる問題がある．

(1) パケット送信を行う全機器が，機器総数 n に対し $O(n^2)$ 個に昇る全リンクステート情報を保持する必要がある．

(2) 一時的な信号欠落や輻輳などにより生じ得るような，瞬間的な*ACK*不到着まで全て障害箇所によるものと見なされパケットが再送される．これらは遠隔管理の安定性のために必ずしも必要な再送ではないにも関わらず，多重再送に備えるため，より多くのパケットをより長時間メモリ上に保持せざるを得なくする．

4. 提案方式

本稿で提案する経路制御プロトコルでは，以下(1)および(2)の方針により既存方式[1]の問題解決を図る．

4.1.経路集約可能となるような一時的IPアドレスの導入

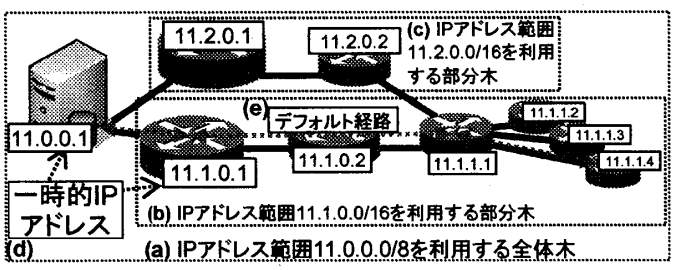

図2 一時的IPアドレス割当ての例

各機器によるリンクステート情報全体の保持を不要とするため，一時的IP自動構成方式[2]を導入する．図2および以下(I)～(III)にその概要を示す．

(I) ネットワークの全体木(図2(a))に対するIPアドレス範囲(図2では11.0.0.0/8)を予め各機器に設定しておき，図2(b)および(c)のような各部分木がそれぞれ，経路集約可能なIPアドレス範囲(図2 (b),(c)に対してはそれぞれ11.1.0.0/16,

†株式会社KDDI研究所 (KDDI R&D Laboratories Inc.)

11.2.0.0/16)に収まるように，部分木内の各機器に管理用IPアドレスとは別のIPアドレス(一時的IPアドレス，図2(d))を，機器の分散協調により自動割当てする．

(II) 機器間接続関係に変更が生じた場合にも，僅かな収束時間 ΔT にてこれに対応する新たな一時的IPアドレスへの再割当てを行い，ΔT の間を除いて経路集約可能な状態を保つ．

(III) 管理サーバと任意の機器との間には，ΔT の間を除いて常に集約された経路(デフォルト経路，図2(e))の存在を仮定できることになるため，これと障害箇所の情報さえあれば，全リンクステート情報を保持することなくソースルーティングを行うことができる．尚，ΔT の間に送信されたパケットは，ΔT 以降に行われる再送によって宛先に到達する．

4.2 パケット不到達率許容上限値の導入

多重再送に備え保持するパケットを削減するため，1) もしACKが到着しなくとも，すぐさま障害箇所に遭遇したとみなして迂回経路探索と再送を開始せず，2) 管理者の定める一定時間 *T1*[sec]毎に，*T1* 間に送信したパケットに対するACKの不到着率 *ACKDropRate*(以下で詳述)が予め管理者の定める *ACKDropRate* の許容上限値 *ACKDropTol*(以下で詳述)を超過した場合のみ，到着しなかった *ACK* に対応するパケットの再送・迂回処理を開始するように[1]を拡張する．これによって一時的な信号欠落や輻輳などにより生じ得るような僅かなパケット欠落は再送対象としないといった，管理者の方針に応じた再送対象パケットの範囲設定が可能となり，機器の再送用パケット保持メモリ負荷削減を図ることができる．以下(I)～(III)にその概要を示す．尚，以下に述べる点以外は[1]と同様な処理とし，*T1*，*ACKDropRate*，および *ACKDropTol* は管理者が予め決定し各機器に設定する．

(I) 一定時間 *T1*[sec]内に，ある宛先機器 *Dst* へと送信されたパケットに対するACKの欠落率 *ACKDropRate(Dst)*[%]，およびその許容上限値 *ACKDropTol*[%]を導入する．*ACKDropRate* は式(1)のように定義する．

$$ACKDropRate(Dst) = \left(1 - \frac{T1\text{内に}Dst\text{から受信した}ACK\text{数}}{T1\text{内に}Dst\text{へ送信したパケット数}}\right) \times 100[\%] \quad \cdots(1)$$

(II) 各機器に表1のような *ACKDropRateTable* を新たに導入する．*ACKDropRateTable* には *T1* 内に送信した各パケット送信宛先 *Dst* (表1(1)，管理用IPアドレスにて記録する)を主キーとして，送信したパケット数(表1(2))，受信した *ACK* 数(表1(3))，現時刻での *ACKDropRate*(表1(4))，*ACK* を受信できていないパケットのシーケンス番号(表1(5))を追加または更新していく．これに加え，再送用パケットも，そのシーケンス番号とともに別途メモリ上に複写保存する．尚，*T1* は各 *Dst* 毎に独立に計時する．

表1 *ACKDropRateTable* の例

(1) *T1* 内に送信したパケットの宛先	(2) *T1* 内に送信したパケット数	(3) *T1* 内に受信した *ACK* 数	(4) *ACKDrop Rate[%]*	(5) *ACK* を受信できていないパケットのシーケンス番号
10.0.0.2	*3*	*2*	*33*	*002*
10.0.0.3	*3*	*3*	*0*	
10.0.0.4	*10*	*5*	*50*	*005,006,007,008,009*

(III) 各機器は時間 *T1* 経過毎に，式(1)により *Tc-T1* から *Tc* までの時間における *ACKDropRate* を計算する．ここで *Tc* は現在時刻を表す．もし *ACKDropRate* < *ACKDropTol* であれば，再送は行わず，再送用に複写保持するメモリ上のパケットを破棄する．もし *ACKDropRate* ≧ *ACKDropTol* であれば，*Tc-T1* から *T1* の間にACKを受信できなかったパケットは障害箇所に遭遇したとみなし，迂回経路を指定して再送処理を行う．

例えば管理者が *ACKDropTol=50%* に設定していて，表1が宛先 10.0.0.4(表1で網掛けした行)について，ちょうど *T1* の計時が完了した状態であると仮定する．このとき 10.0.0.4 について *T1* の間に10個送信したパケットのうち，シーケンス番号 005,006,007,008,009 の5個について *ACK* を受信できなかったため，*ACKDropRate* はちょうど 50%となり，*ACKDropRate* ≧ *ACKDropTol* を満たす．このため上記5パケットについて障害箇所に遭遇したとみなし，迂回経路を指定して再送処理を行う．

5. 考察

5.1 パケット到達の確実性とトランスポート層との連携

4.2節の提案方式ではパケット到達の確実性(一定時間内におけるパケットの宛先到達数/送信数)が[1]に比して低下する可能性がある．しかしながら本稿ではトランスポート層の下層における経路制御プロトコルを想定している．また，アプリケーションによって，トランスポート層以下に求めるパケット到達の確実性は多様であり，例えばsshといった確実性が優先されるアプリケーションに対してはTCP，snmpといったプロトコルの単純性・軽量性が優先される場合にはUDPが用いられる．そこでトランスポート層がTCPの場合は[1]の方式を，UDPの場合は4.2節の提案方式を用いるといった実装形態とすれば，[1]によってトランスポート層から受け取った全てのパケットに対し同様に障害箇所検出・迂回再送処理を行うよりも，再送用パケット保持に要するメモリ量を削減し得る．尚，TCPパケット数とUDPパケット数との比，障害箇所発生頻度や発生時間，*T1* の大きさ等により，上記の削減量は変動し得るため，これらをパラメータとする計算機シミュレーション等を行って，適切な実装形態を検討することが望ましい．

5.2 ACKDropTol の決定方法

4.2節の提案方式は，管理者の定めた基準ACKDropTol以上の不到着率ACKDropRateが頻出する状況では，ACKの到着しなかったパケットをほぼ全て再送するようになり，そのメモリ消費は[1]を適用した場合に漸近することが予想される．このため管理者は，所望のメモリ負荷とパケット到達の確実性とを達成するようなACKDropTolを決定する必要があるが，これは必ずしも容易ではない．このためACKDropTolをACKDropRateによって適応的に上下させ，管理者はその初期値や上下限のみ指定するといった拡張が考えられる．

6. おわりに

本稿ではネットワーク規模増大に対し機器のメモリ負荷を増大させないように遠隔管理専用のIP経路を設定する，機器負荷を考慮した経路制御プロトコルを新たに提案した．最後に，日頃ご指導頂く(株)KDDI研究所秋葉所長，ならびに鈴木執行役員に感謝する．なお本研究の一部は，総務省委託研究「ユビキタスネットワーク技術の研究開発」により実施している．

参考文献

[1] I.Avramopoulous et.al., "Highly Secure and Efficient Routing," IEEE Infocom 2004.

[2] 堀，吉原，堀内，"大規模IPルータ網遠隔管理のための一時的IP自動構成方式の提案，" 電子情報通信学会2005年ソサイエティ大会予稿集 B-7-059 pp.186，2005年9月．