

PIM-SM の DR Dynamic Election *

2 F - 5

松澤 智史[†] 中山 雅哉[‡]

東京大学大学院工学系研究科 電子工学専攻[§] 東京大学情報基盤センター[¶]

1 はじめに

今日のインターネットには、あらゆる種類のデータ (特にリアルタイムおよびマルチメディアデータ) を、タイムリーかつ確実にどのユーザにも配信しなければならないという課題が発生している。これらのニーズに対し、従来型のネットワークでは、多数のユーザに同じ情報同時送信すると、容易にネットワーク容量の限界に達してしまう。この伝達情報の不要な重複を減少させ、ネットワークの負荷の軽減させる目的で IP マルチキャストが誕生した。しかしながら IP マルチキャストのルーティングには、まだまださまざまな問題が多い。本稿では、IP マルチキャストルーティングの代表的なプロトコルである PIM-SM について、現在の PIM-SM の仕様では問題が発生するトポロジを明確にし、それに対して DR (Designated Router) の動的な決定法による解決案を提案する。

2 問題となる構成

現在の PIM-SM の仕様では、ある特殊な構成の場合に、通信がうまく行かないという現象が起こりうる。図 1 は、その特殊な構成であり、この場合不具合が生じる。ネットワークは 5 つあり、それぞれ 3ffe:501:ffff:1/64 ~ 3ffe:501:ffff:5/64 というネットワークアドレスを持つ。 (IPv4 では 192.168.1/24 ~ 192.168.5/24) MR1 以外はすべてマルチキャストルータであり、MR3 は xl0 側で、ff00::/8 のマルチキャストアドレスに対して RP になることを宣言しており、同アドレスに対しての Bootstrap Router (BSR) であることも宣言している。

まず送信者は MR7 の下側 Interface (fxp0) よりマルチキャストアドレス ff08::5555 宛のマルチキャストパケットを送信する。受信者は MR1 で Interface fxp0 にて ff08::5555 宛のパケットを受信する。MR1 と 7 以外の各ホストでは、pim6sd が動作しており、MR3 は xl0 側の

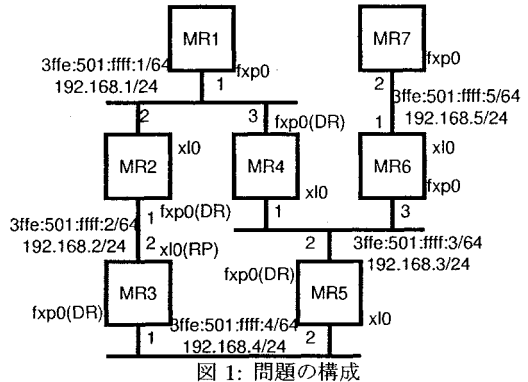


図 1: 問題の構成

Interface で ff00::/16 に置いて RP になることを宣言している。3ffe:501:ffff:1/64 の DR は MR4 (fxp0) である。

MR1 で受信を開始した場合に MR2 と 4 で実際どのようなテーブルが作られるかを見ると以下ようになる。

MR2 (*,G)	MR4 (*,G)
i o j p l	i o j p l
-----+-----	-----+-----
fxp0 i	fxp0 i 1
xl0 o j l	xl0

では次に SPT 移行時の検証を試みる。

SPT 移行へのトリガは通常 DR が役目を果たすが、この場合 DR の振る舞いをしていない MR2 がトリガとなる。このトリガとなるルータに (S,G) エントリが生成され、ユニキャストのルーティングに従い S への経路となる Interface に Join が投げられる。この場合、MR2 は MR4 側の Interface である xl0 側に Join を投げるのだが、Incoming と Outgoing が同じため o フラグも立てず Join も投げない。(o は一時的に立つがすぐ消える) 最終的に MR4 の (S,G) エントリは

MR2 (S,G)
i o j p l
-----+-----
fxp0
xl0 i

*DR Dynamic Election of PIM-SM

[†]Tomofumi Matsuzawa

[‡]Masaya Nakayama

[§]Electronics Engineering, Engineering Department, Univ. of Tokyo

[¶]Information Technology Center, Univ. of Tokyo

となる。MR4には(S,G)のエントリさえ作られない。よってこの構成ではSPTに移行できていない。

3 動的なDR決定

これらの特殊なケースは、DRが静的に決定されることから生じる。つまりDRをソース(もしくはRP)に対して動的に決定するようになれば解決する。では、実際にどのようにDRを動的に選択するかの手順を示す。

1. PIM ルータの各インタフェース情報の中にDRを情報を持たない。
2. ListenerからのMLD reportを受信した際に同一セグメントのマルチキャストルータはそのMLD reportの指定するソースへのmetricを含んだDR-Election¹を送信する。(RPT形成時のDR決定は、BSR(Bootstrap Router)よりRP情報通知パケットが届いた際にDR-Electionを発信する。)
3. マルチキャストルータはそのDR-Electionに含まれるmetricを比較し、小さい方がそのソース(もしくはRP)に対してのDRとなる。
4. metricが同じ場合は、DR-Electionに含まれるPIMルータのアドレスを比較して大きい方がDRとなる。(従来の静的DR決定法)

だが、これだけでは先程のトポロジでのSPT移行時にRP経由のパケットが流れ続けてしまうので、PIMルータには、「RPツリーでのDRだったPIMルータがMLD report(S,G)をDRだったInterface側に送信する。」と「SPT時に新DRとなったPIMルータは(S,G)RP PruneをRPに向けて送信する。」の2つの変更を追加する。

このような変更を加えることで、SPTとRPT時のDRが分離してもそれぞれRPTのパケットをprune、SPTのパケットがアクティブに流れるという状況をつくり出すことが出来る。ここで2番の(S,G)RP Pruneを出すタイミングが問題になるが、この(S,G)RP Pruneのトリガは「SPTビットのたったグループのincomingよりdataパケットが到着する。」と「DRとなるInterface側で(*,G)と(S,G)の両方のMLD reportを受信する。」によって発動する。

4 評価

DR動的に決定した場合と、従来の静的な決定の場合とでそれぞれRPTとSPTが意図するように作成されるかを評価した。ここでListenerと同一セグメントに存在するA(静的な方法でのDR)とBという2つのルータとRPとなるルータ、送信元のSという記号を用いてあらゆるトポロジでの評価をする。(A-RP(B)とはAの

¹PIM-Assertと同じフォーマットで、目的ホストへのmetric値が入る

ルーティングテーブルにおいてRPまでの最短経路上にBが存在することを示す。)

RPT生成の場合

- 状態1: A-RP(B), B-RP()
- 状態2: A-RP(), B-RP(A)
- 状態3: A-RP(), B-RP()
- 状態4: A-RP(B), B-RP(A)

SPT移行の場合

- 状態5: 状態1でA-S(B) B-S()
- 状態6: 状態1でA-S() B-S(A)
- 状態7: 状態1でA-S() B-S()
- 状態8: 状態1でA-S(B) B-S(A)
- 状態9: 状態3でA-S(B) B-S()
- 状態10: 状態3でA-S() B-S(A)
- 状態11: 状態3でA-S() B-S()
- 状態12: 状態3でA-S(B) B-S(A)

	状態1	状態2	状態3	状態4
静的	◎	◎	○	×
動的	◎	◎	◎	×
	状態5	状態6	状態7	状態8
静的	◎	×	◎	×
動的	◎	◎	◎	×
	状態9	状態10	状態11	状態12
静的	◎	◎	○	×
動的	◎	◎	◎	×

表中の◎は最適な1つの木ができあがることを意味し、○は木は作られるが2重になる、×は木が作成されないことを意味する。状態4,8,12は実環境においてありえないルーティング状態であるので、本手法による木の作成があらゆる状態で有効であると言える。

5 まとめ

本稿では、現在の静的なDR決定法によるある特殊なトポロジでの問題点の一例をあげ、解決法であるDRの動的決定法と木の生成に対する評価を述べた。

参考文献

- [1] D.Estrin, D. Farinacci, A. Helmy, D. Thaler, S. Deering, M. Handley, V. Jacobson, C. Liu, P. Sharma, L. Wei "Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification" RFC2117, June 1997.