

マルチキャスト： IGMP

2024/08/22

小原泰弘

要点

- マルチキャスト（グループ形成）はタイマーです
 - IGMPデバッグ：キャプチャーすれば分かる
- マルチキャスト動作原理
 - IGMP動作原理：タイマーベースのグループ形成
 - PIM-SM動作原理：共有木：配送木の単一化
 - SSM：Source-Specific Multicast

マルチキャストの近年の隆盛

- テレビ局、IP映像転送、8K高精細映像転送
- SFPがマルチキャスト（IGMP）喋る
- さっぽろ雪まつり：「8K + IP Multicast + SRv6」
- インタロップ：「Media over IP」

マルチキャストが難しい理由

- ユニキャストは終点ベースでシンプル
- マルチキャスト
 - 終点が地理的に複数あって複雑
(複数の分散した終点をまとめてグループとして扱いたい)
 - ネクストホップが複数
 - 始点ごとに配送経路が変わる (くせにグループとしては同一)
 - 複雑すぎる、消費多すぎなので、効率性無視して、配送経路を単一化
 - 効率性悪すぎなので、効率化 (最適化)
 - 動的配送木変更
 - SSM

技術資料 (1)

- RFC 988: Host Extensions for IP Multicasting, 1986 (IGMPv0)
- RFC 1112: Host Extensions for IP Multicasting, 1989 (IGMPv1: Appendix)
- Stephen E. Deering Dissertation: Multicast Routing in a Datagram Internetwork, 1991
- Core Based Trees (CBT): An Architecture for Scalable Inter-Domain Multicast Routing, SIGCOMM '93
- RFC 2236: Internet Group Management Protocol, Version 2, 1997
- RFC 2710: Multicast Listener Discovery (MLD) for IPv6, 1999
- RFC 3376: Internet Group Management Protocol, Version 3, 2002
- RFC 3810: Multicast Listener Discovery Version 2 (MLDv2) for IPv6, 2004
- RFC 4541: Considerations for Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) Snooping Switches, 2006

技術資料 (2)

- RFC 3569: An Overview of Source-Specific Multicast (SSM)
- RFC 4604: Using Internet Group Management Protocol Version 3 (IGMPv3) and Multicast Listener Discovery Protocol Version 2 (MLDv2) for Source-Specific Multicast
- RFC 4607: Source-Specific Multicast for IP
- RFC 7761: Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised), 1998->2006->2016

- RFC 3618: Multicast Source Discovery Protocol (MSDP)
- RFC 4610: Anycast-RP Using Protocol Independent Multicast (PIM)
- RFC 5059: Bootstrap Router (BSR) Mechanism for Protocol Independent Multicast (PIM)
- RFC 4760: Multiprotocol Extensions for BGP-4
 - SAFI 2 (multicast) が定義されているだけ

IPマルチキャストの基本

- マルチキャストとは？ユニキャストとは？ブロードキャストとは？
- IPにおけるマルチキャストの基本（Steve Deeringの成果）
- データグラムネットワークでのマルチキャスト（IP Multicast）を規定。
- Best-Effortである。not guaranteed。
- Host Group: 単一のIP終点アドレスで表されるホスト群。メンバーシップはダイナミック（join/leave）。メンバーの場所や数に制限がない。ホストは複数グループに所属できる。マルチキャスト送信者はホストグループメンバーである必要はない。
- インターネット上全ホストというグループは定義なし
- ホストのオペレーション：
 - JoinHostGroup (group-address, interface) → (Unsolicited) Membership Report メッセージの送信
 - LeaveHostGroup (group-address, interface) → Leave Group メッセージの送信
- RPF : Reverse Path Forwarding

クラスD IPアドレス

- Class D IP address: Section 4 of RFC 1112
- アドレスのクラス (→CIDR) 。
 - クラスA: 0.0.0.0/1: /8: 0.0.0.0~127.255.255.255
 - クラスB: 128.0.0.0/2: /16: 128.0.0.0~191.255.255.255
 - クラスC: 192.0.0.0/3: /24: 192.0.0.0~223.255.255.255
 - **クラスD: 224.0.0.0/4: --: 224.0.0.0~239.255.255.255**
 - クラスE: 240.0.0.0/4: --: 240.0.0.0~255.255.255.255
- 224.0.0.0/4
 - **1110** **** . **** **** . **** **** . **** **** / 4
 - $128 * 1 + 64 * 1 + 32 * 1 + 16 * 0 = 224$
 - 224.0.0.0 ~ 239.255.255.255
- ALL-SYSTEMS (224.0.0.1 / FF02::1)
- ALL-ROUTERS (224.0.0.2 / FF02::2)
- SSM: 232.0.0.0/8, FF3x::/96

個人的感想

- ホストグループとは、ホスト群を単一のアドレスで表しましょう、という発明
 - ラベルスイッチングや、BGPコミュニティとかと同じような発明
 - グループかどうかは、終点アドレス見れば分かる：224/4 or FF00::/8
- IGMP: avoid implosion: random delay timer
 - 処理性能が低く、かつ同程度の処理性能の装置が大量にある場合、コリジョンが多く起きるため、相乗効果で事態を悪化させる。これを回避するための昔の手法。今は要らない気がする。コリジョン？スイッチング？
 - 誰か一人が返せば良い。そのためのランダムタイマーは頭良い。

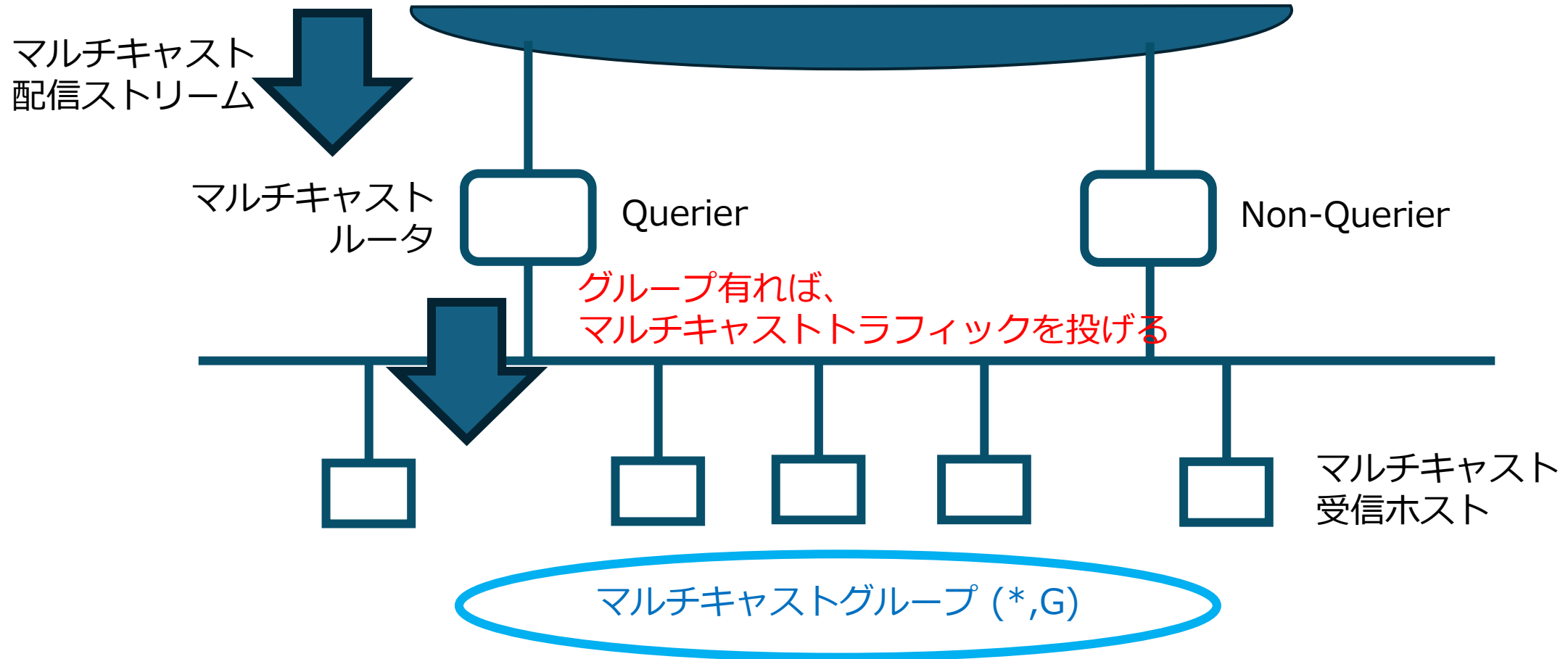
IGMPシリーズ

- RFC 988: Host Extensions for IP Multicasting, 1986 (IGMPv0)
 - create groupメッセージとかある。version (typeの上位4bit) は0.
- RFC 1112: Host Extensions for IP Multicasting, 1989 (IGMPv1)
 - v0から、プロトコルっぽい部分がゴソッと削られている
- RFC 2236: Internet Group Management Protocol, Version 2, 1997
 - 一番ポピュラーな（使われている）IGMP。
- RFC 3376: Internet Group Management Protocol, Version 3, 2002
 - ソースフィルター
 - 浮動小数点
 - メッセージサイズの克服 : Current + Change

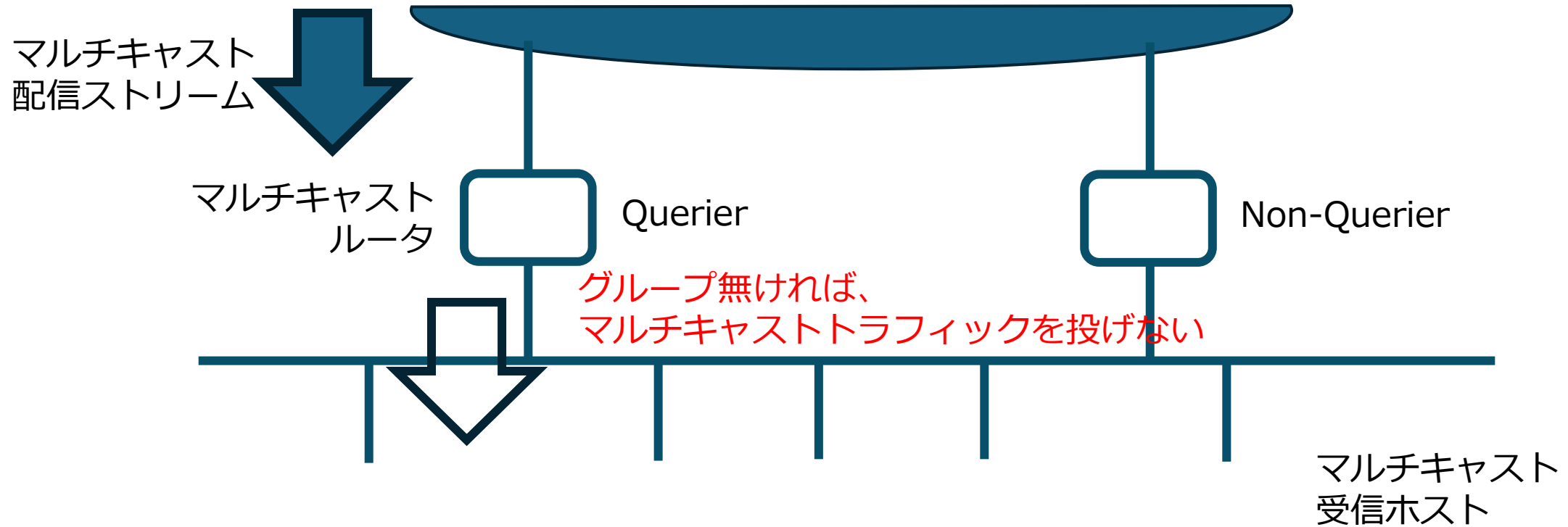
IGMPの原理

- ローカルLANで、特定グループに一つでもホストが属しているか（つまりグループがあるか？）、を調べるプロトコル
- 基本的には全部タイマー。
 - 一つでもグループありますかメッセージ : Membership Query: General Query
 - 特定のグループありますかメッセージ : Membership Query: Group-Specific Query
 - 少なくとも1ホストは居ますよメッセージ : Membership Report
- ルータがQueryで聞いて、ホストがReportで答えたら、グループあり。ランダムタイマーしてどれか一個が答える。他は返答を聞いていて、タイマーをリセットする。だから (robustness variable分以外は) 1個ずつしか流れない。
- 一定時間返答が無ければタイマーが時間切れになってグループ削除。
- Querier : 代表で聞くルータ。Non-Querier : それ以外のルータ。Querier Election
- Queryがあれば、所属するホストはランダムタイマーして時間切れになったらReportを投げる。大抵どれか1つのホストが答えて終わり。
- v2 では Query も Report も 8バイトの小さなメッセージ。
- v2 で Leave ができた。(v1から比べて) グループ削除が数十秒早くなるだけ。
- v3 ではソースごと。ソースフィルタ記述のためにメッセージサイズが増大。

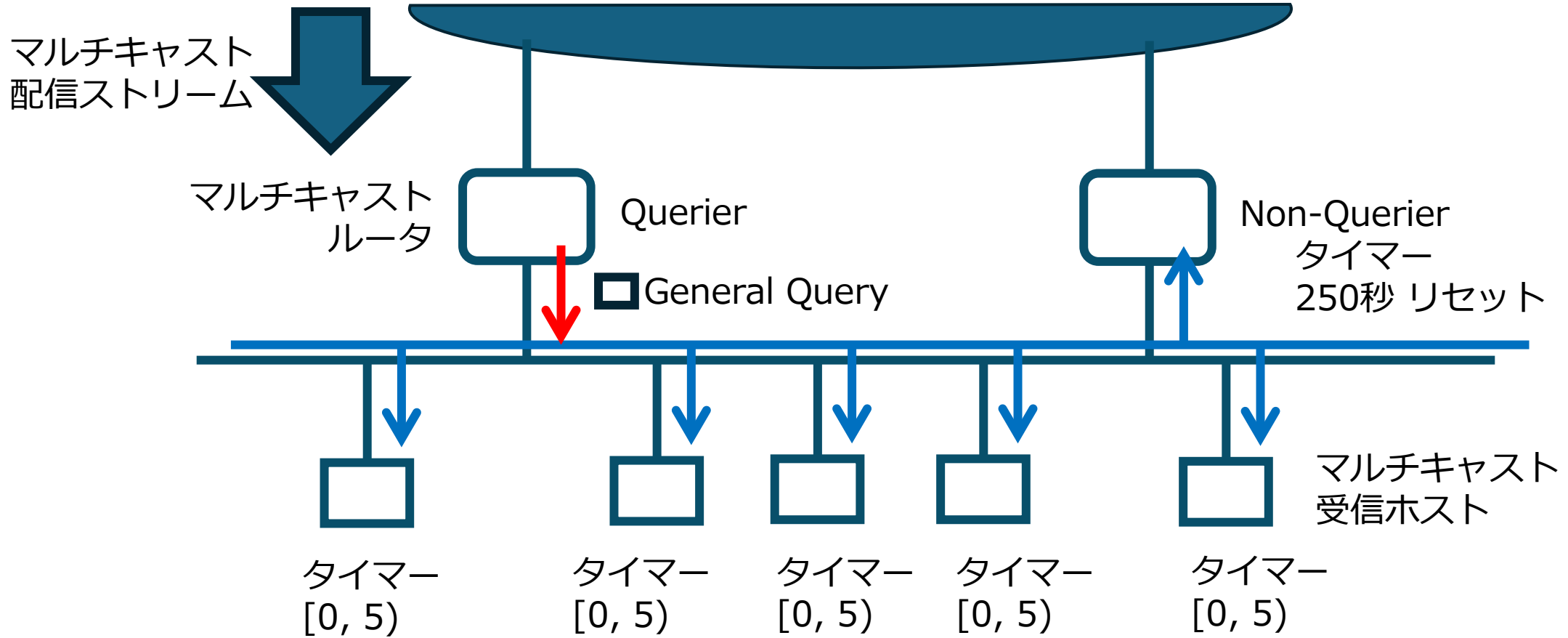
IGMPv1/v2の目的



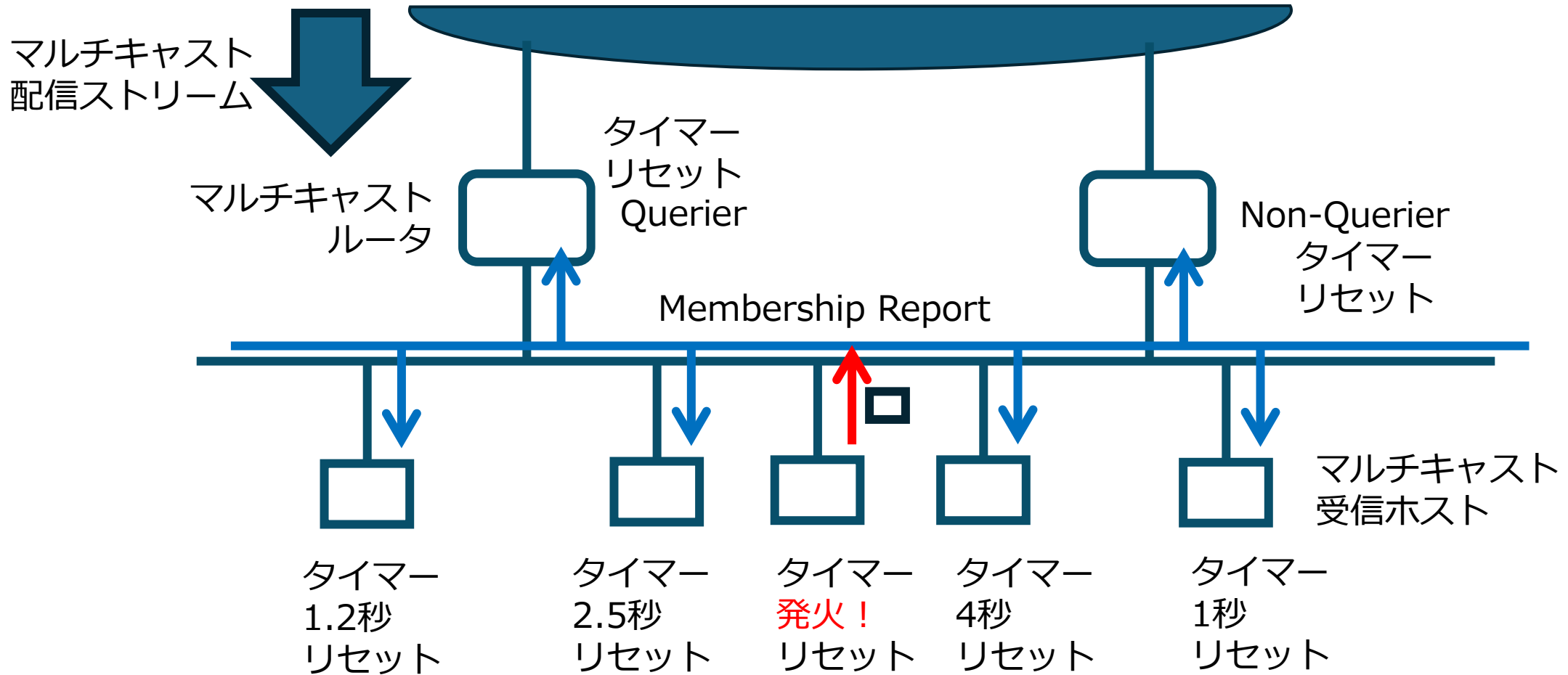
IGMPv1/v2の目的



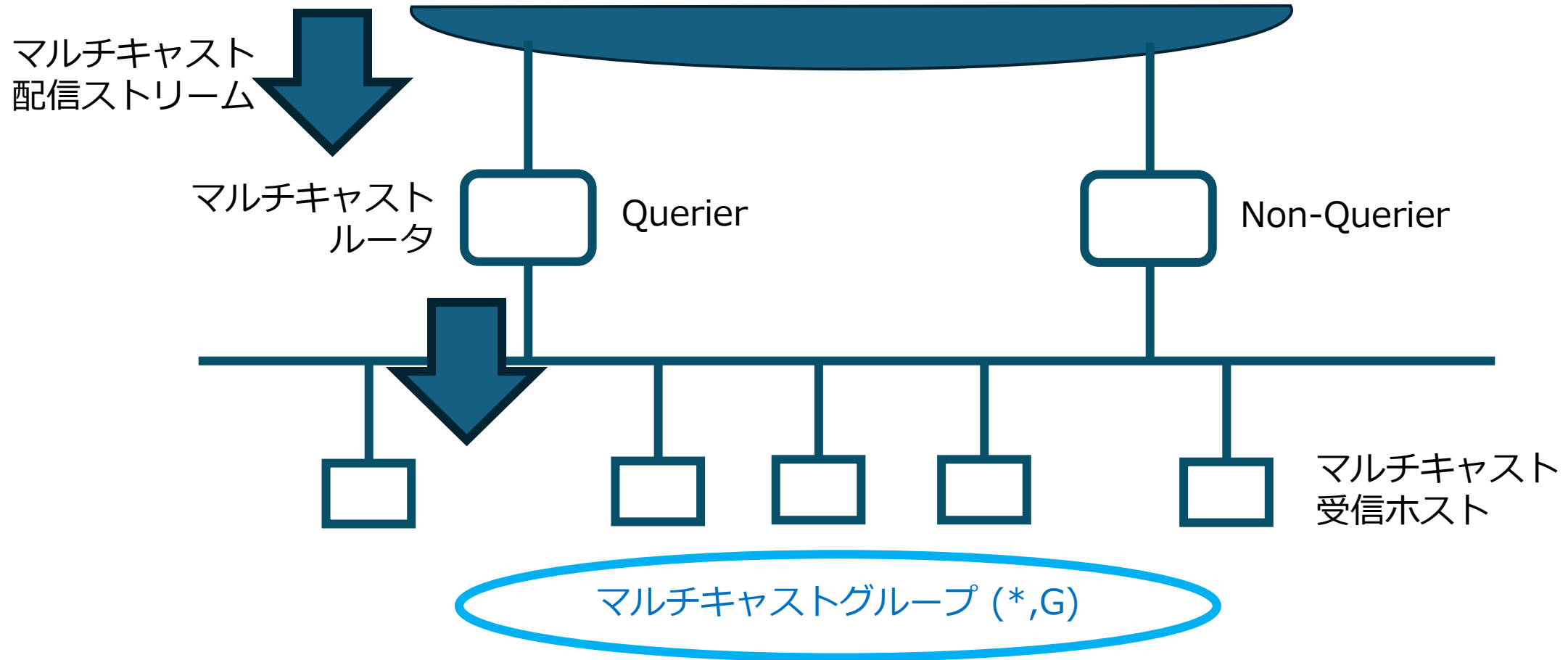
IGMPv1/v2 : Query



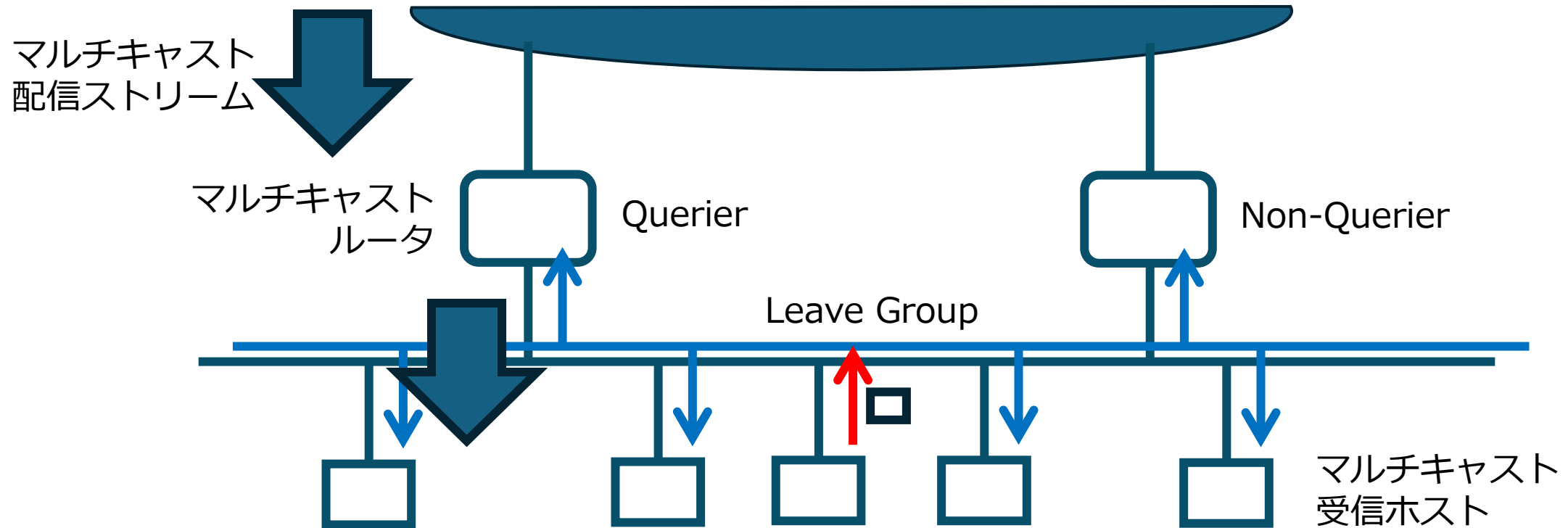
IGMPv1/v2 : Report



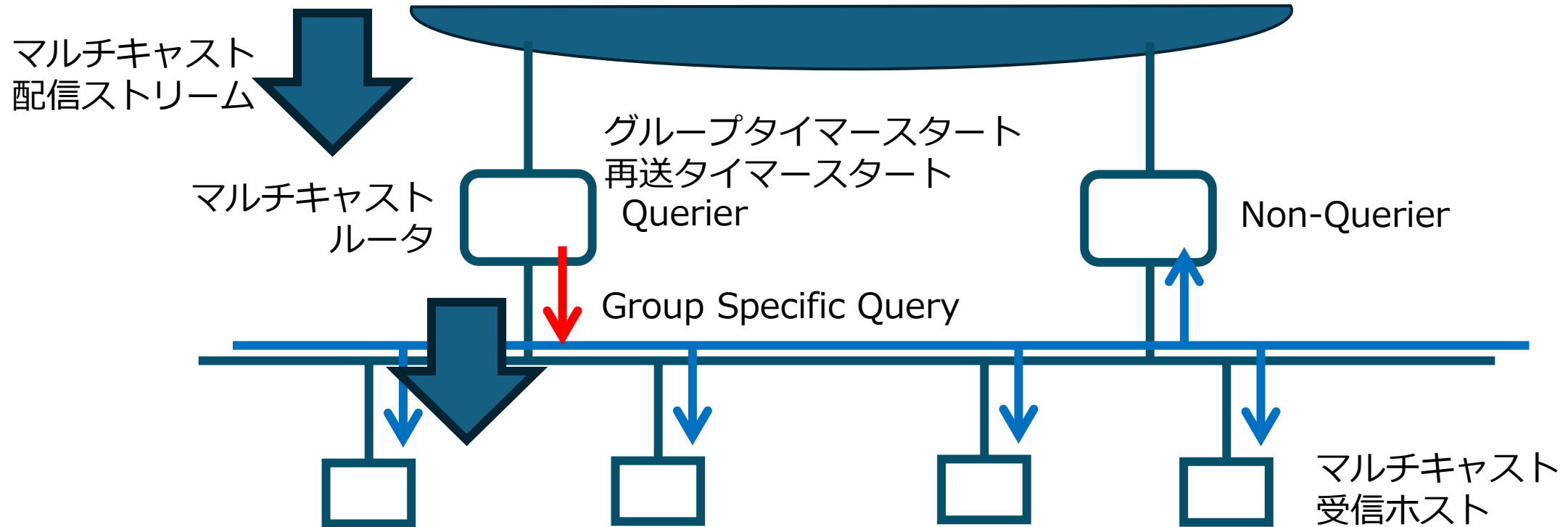
IGMPv1/v2 : グループ有るので投げる



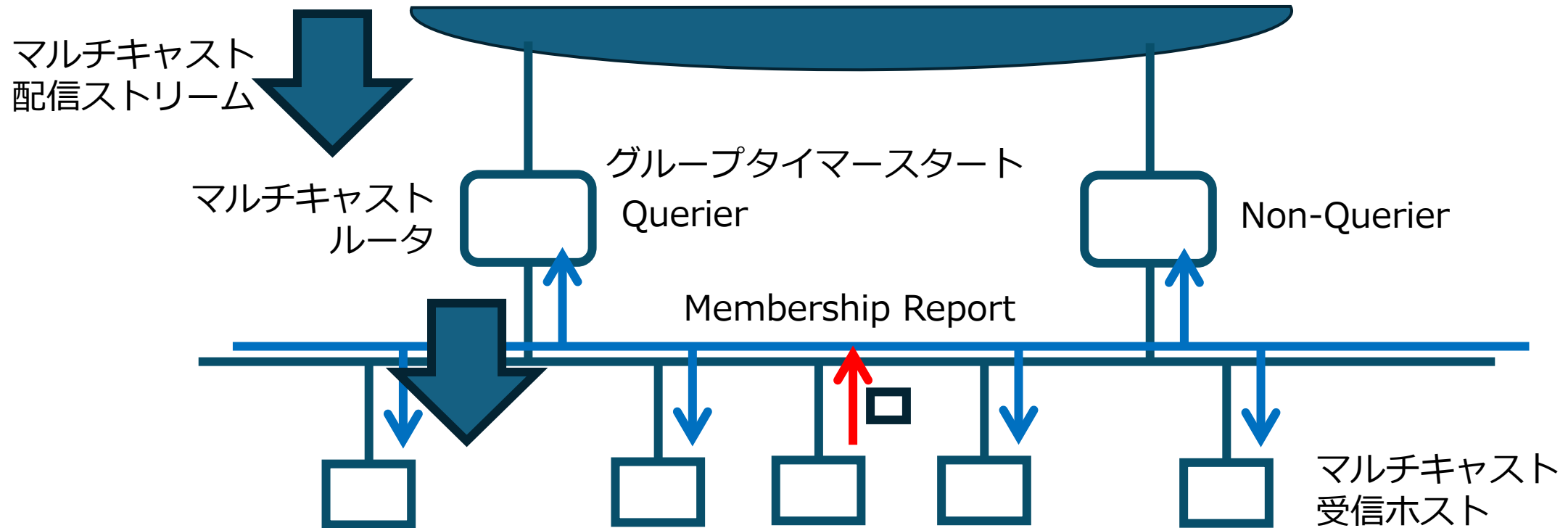
IGMPv2 : Leave Group Sent



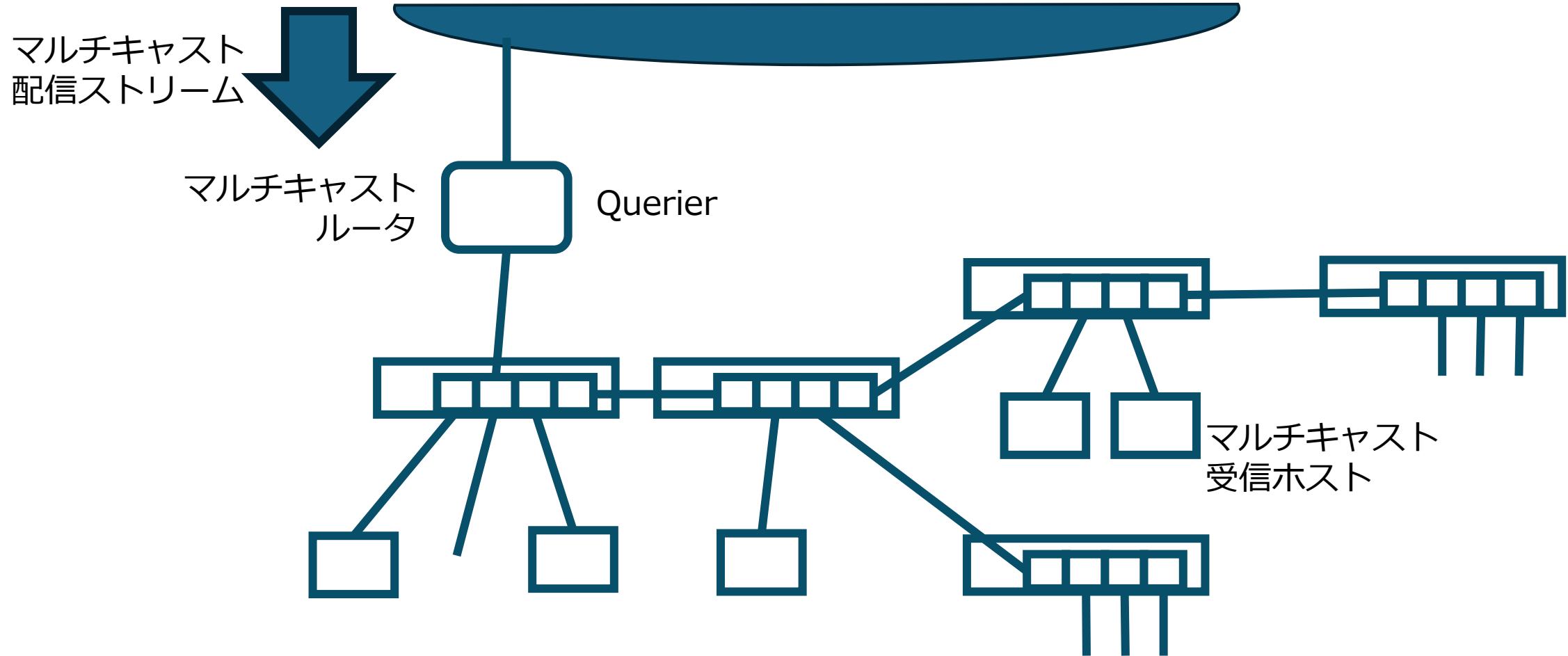
IGMPv2 : Leave Group Received



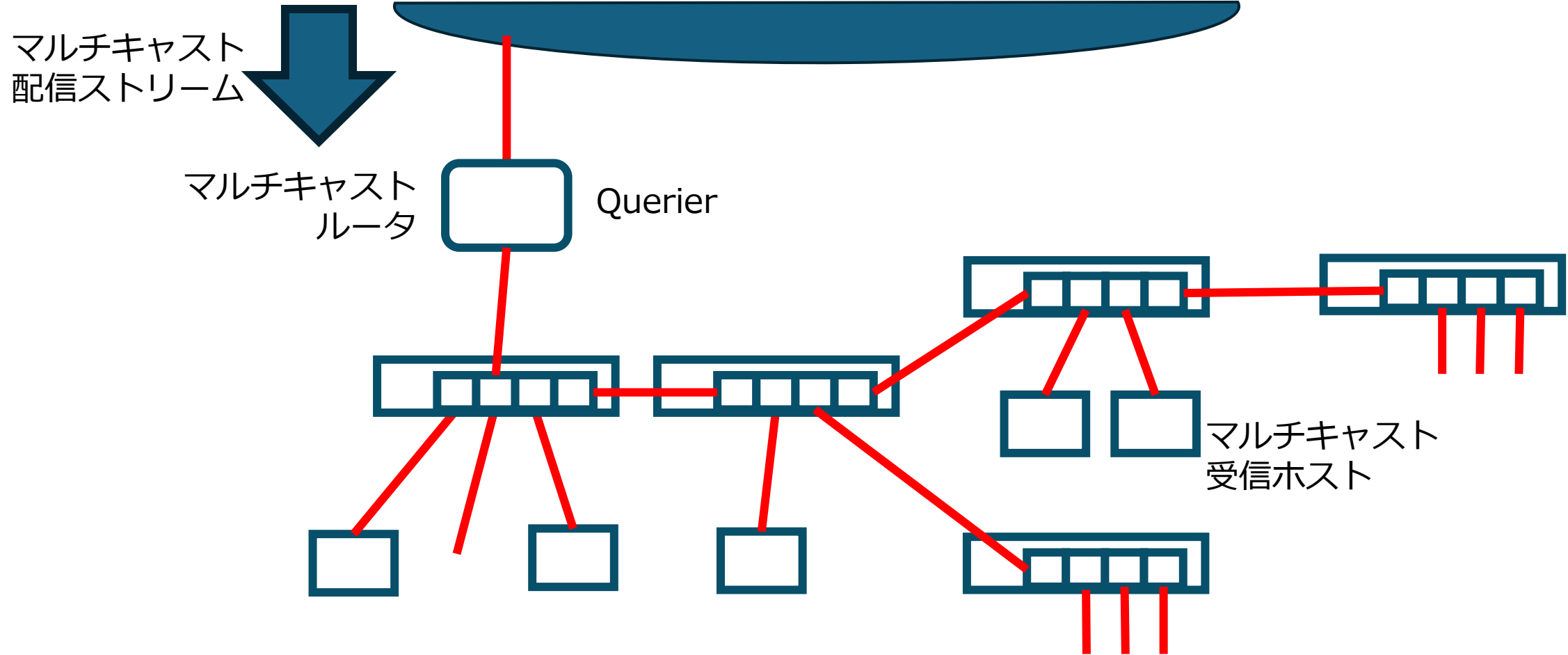
IGMPv2 : Unsolicited Report



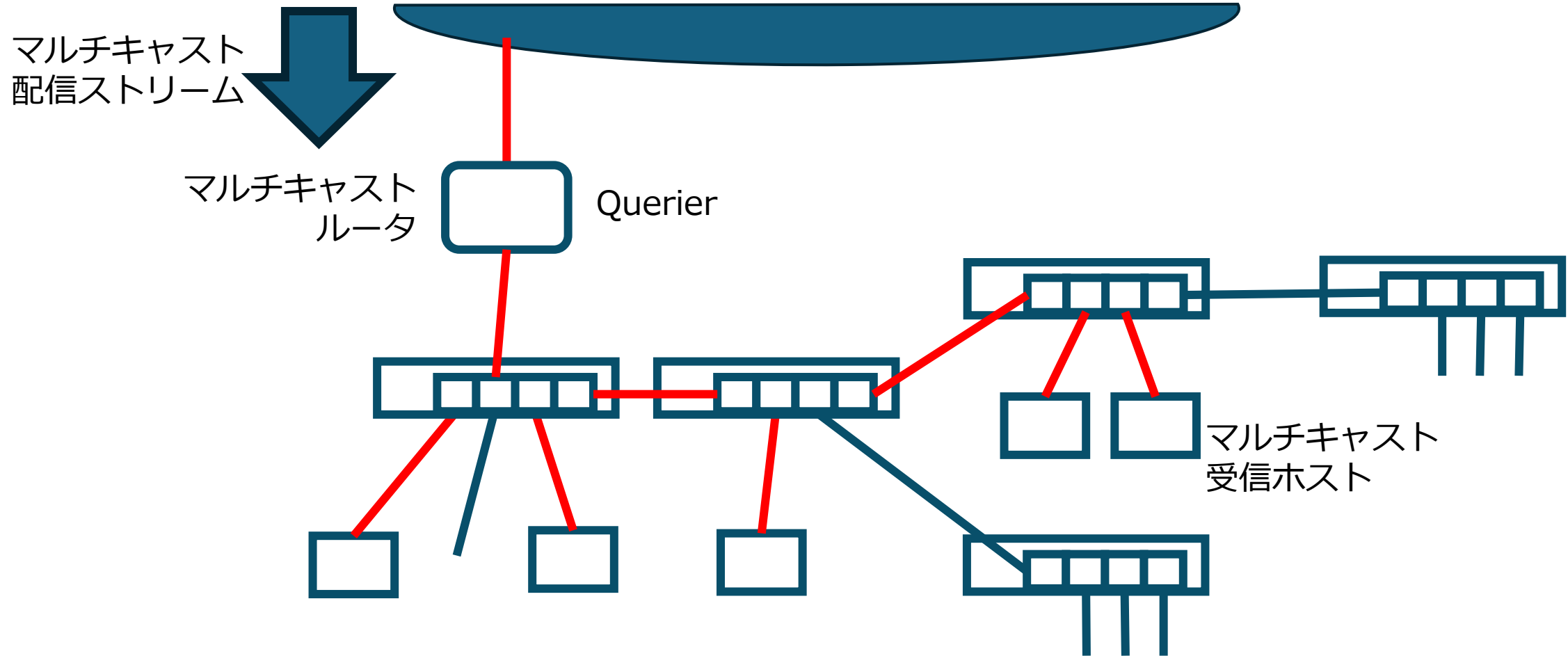
IGMP snooping



IGMP snooping (without)



IGMP snooping (with)

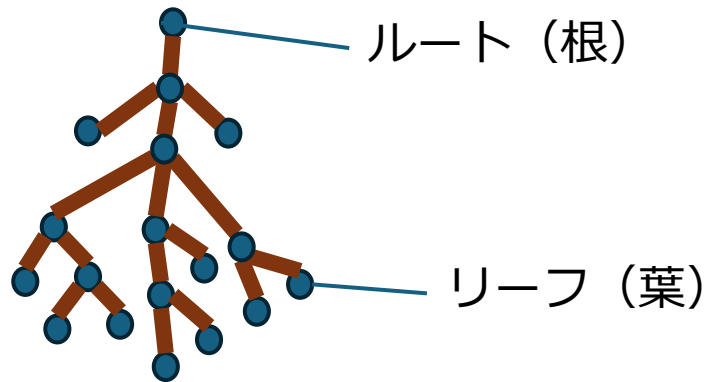
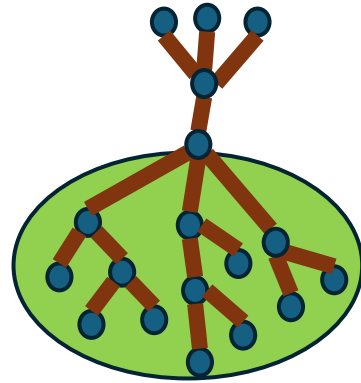
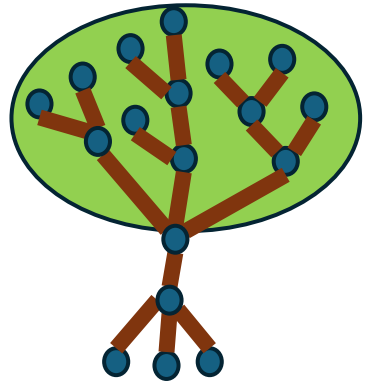


IGMPv3

- 余計な機能がいっぱい
- 指数部と仮数部、floating point での数の指定： 1.23e+999
 - 粗い（粗く変動する）粒度で、小さな数や大きな数を（同じビット幅で）表せる。（例：8-bit：256 v.s. 32,768）
- ソースフィルター要る？
 - ホストやアプリがどうやってソースの違いを知るのか？
- PIM-SSM では IGMPv3 INCLUDE だけ使う

- Current + Change Report：ソースフィルターの追加など
- EXCLUDE > INCLUDE：基本ルールとして、EXCLUDEのホストが居たらルータもEXCLUDE。
- INCLUDEとEXCLUDEでタイマーが別の意味。
- EXCLUDE (X, Y) (X := timer > 0, Y := timer = 0)
 - timer >0 の場合転送する、timer =0 の場合は転送しない
- INCLUDE (X) (X := timer >0)
 - INCLUDEの場合はtimer >0 の場合だけソースリストを保持、これらについては転送する。

木とは？

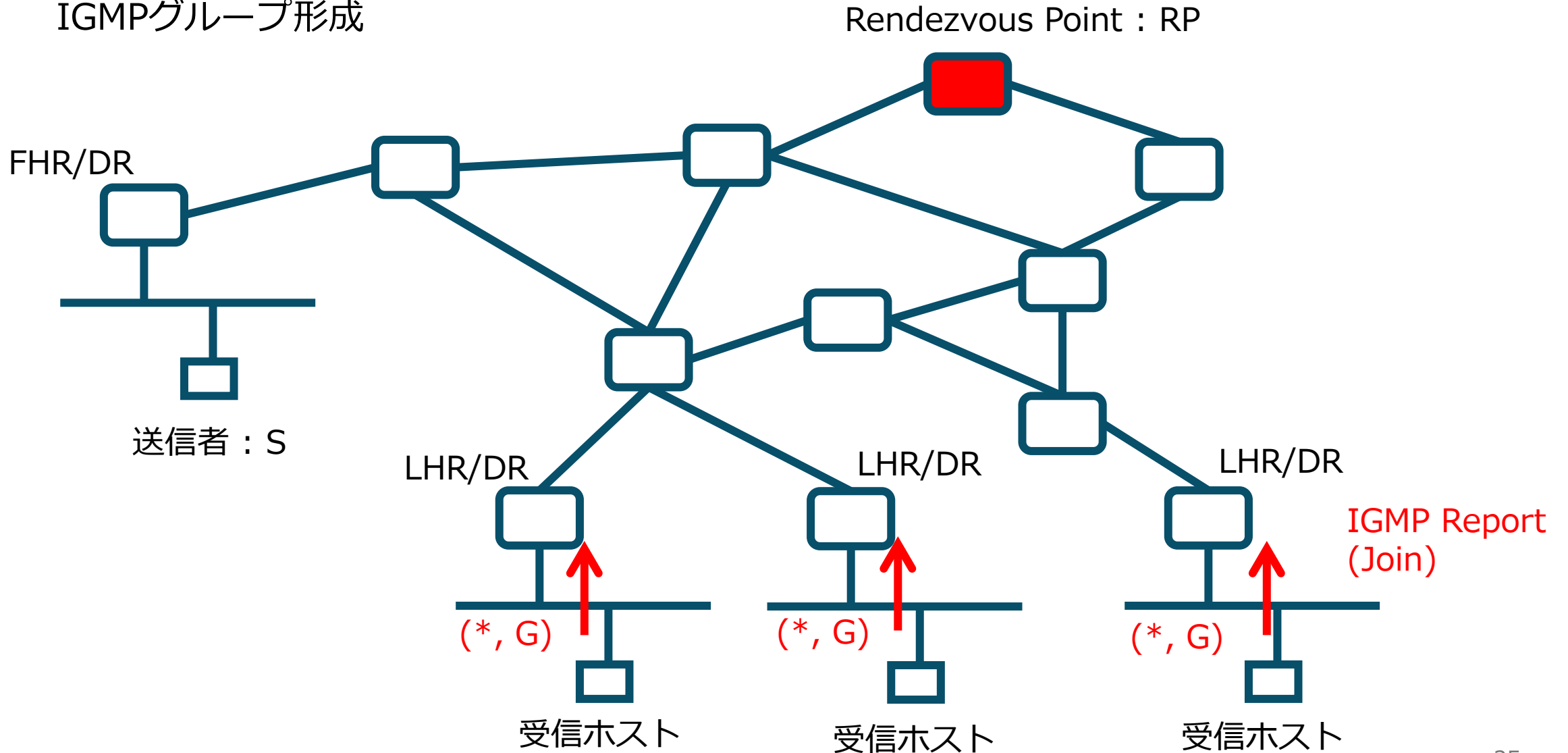


- グラフ理論 $G=(V,E)$
- ノード、接点、頂点、vertex
- リンク、枝、辺、edge
- ツリー・木構造
 - サイクル（閉路）を持たないグラフ
- 有向・無向
- DAG: Directed Acyclic Graph

- 全域木 spanning tree
- 最短木 shortest path tree

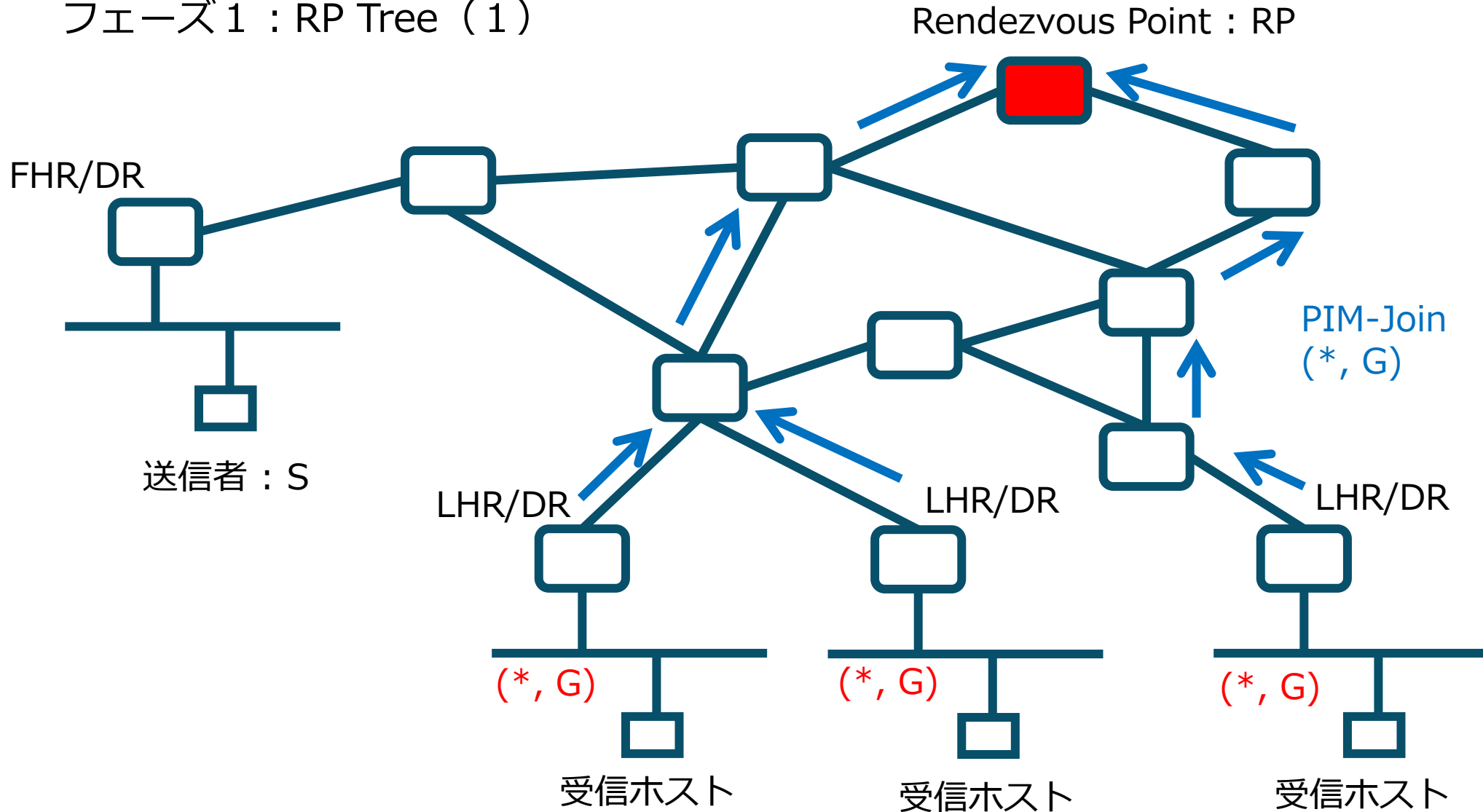
PIM-SM

IGMPグループ形成



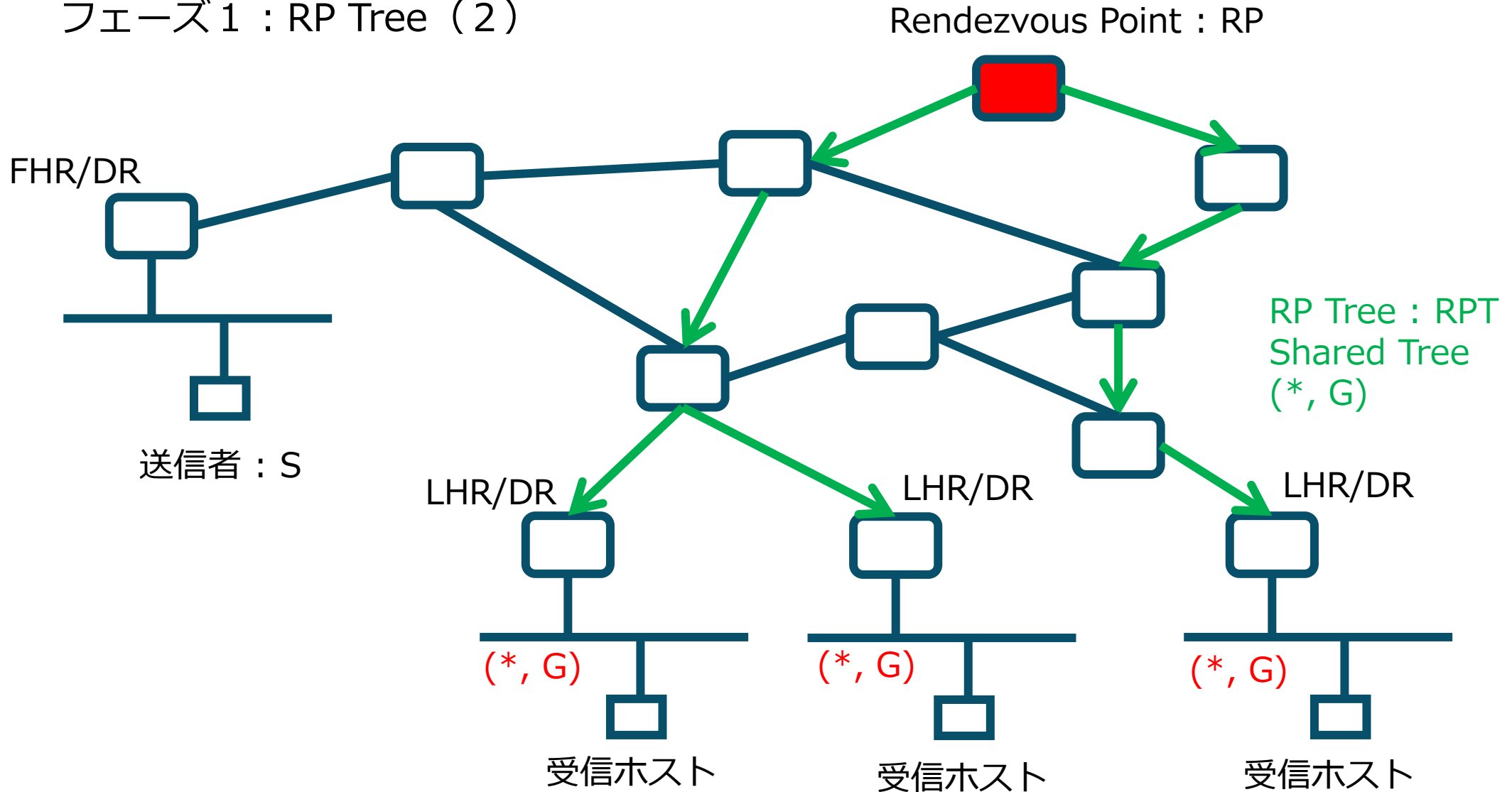
PIM-SM

フェーズ1 : RP Tree (1)



PIM-SM

フェーズ1 : RP Tree (2)

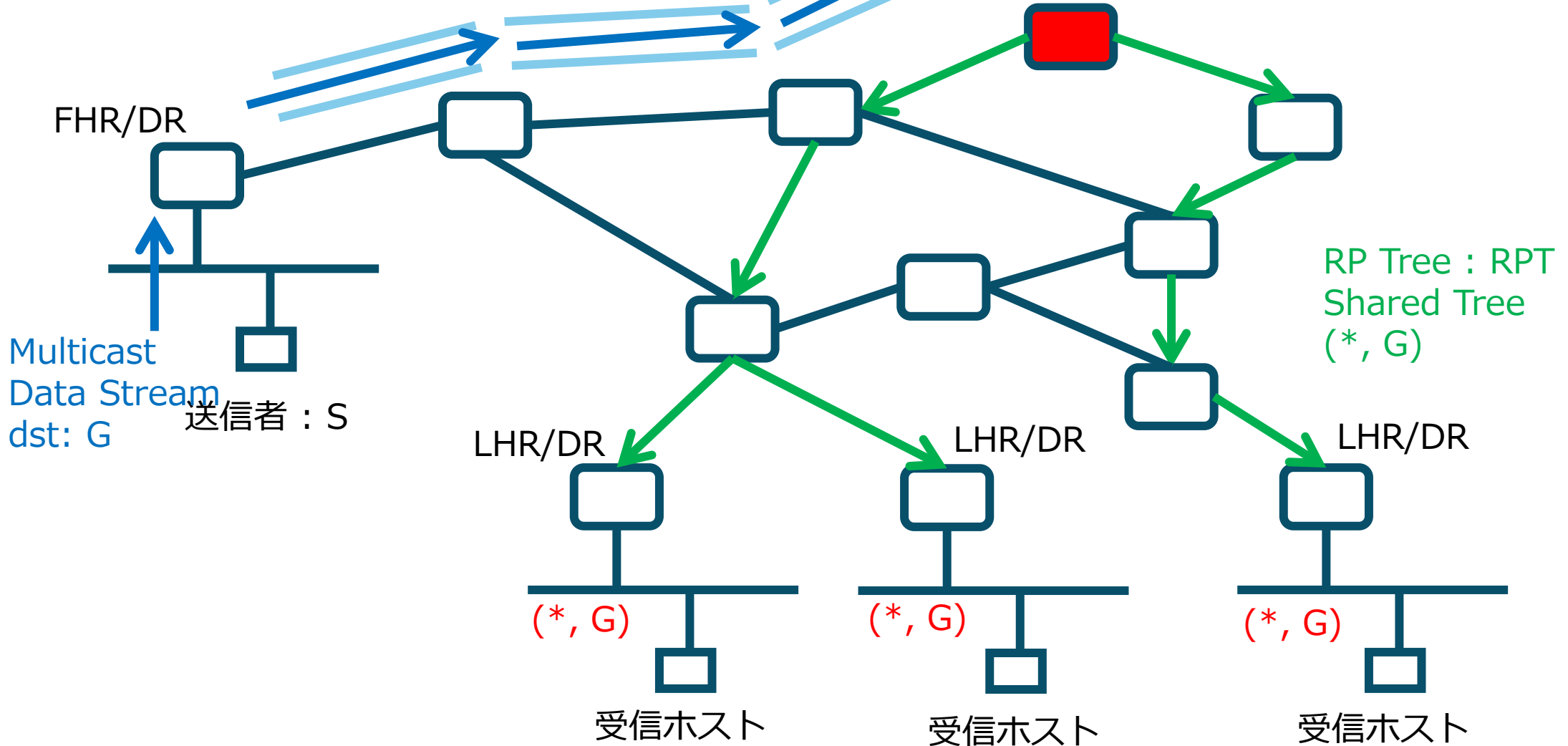


PIM-SM

フェーズ1 : データ送信 (1)

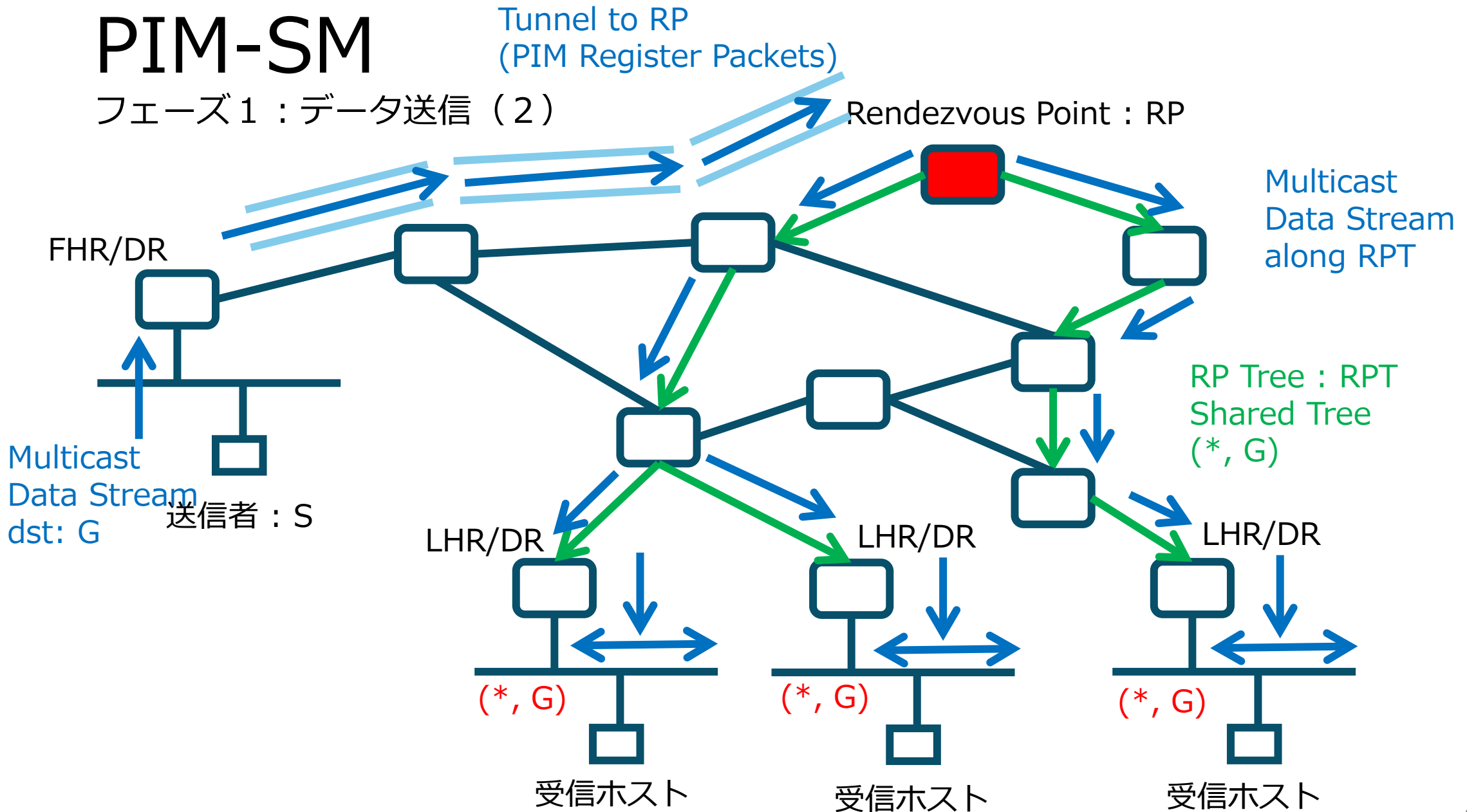
Tunnel to RP
(PIM Register Packets)

Rendezvous Point : RP



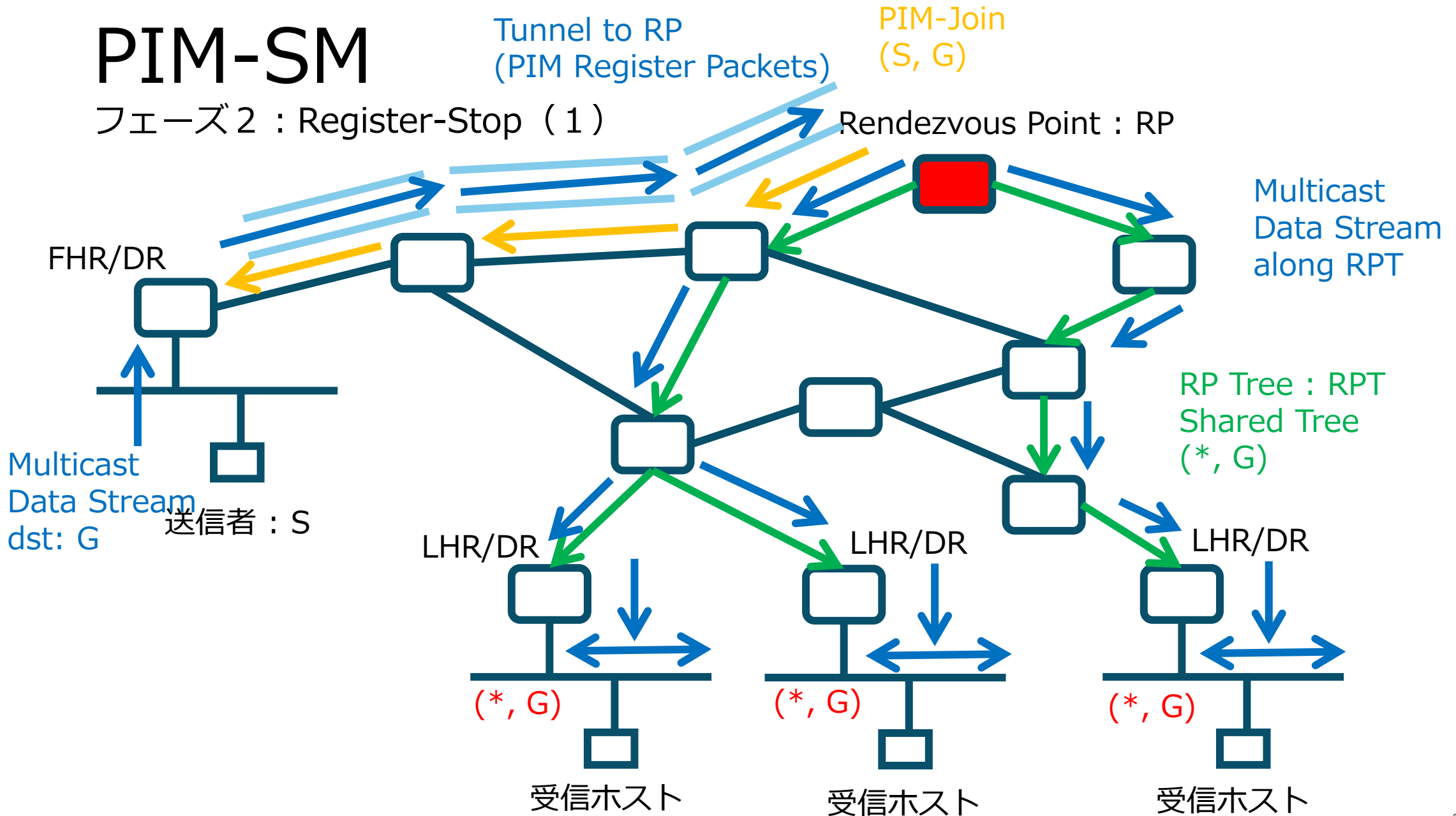
PIM-SM

フェーズ1 : データ送信 (2)



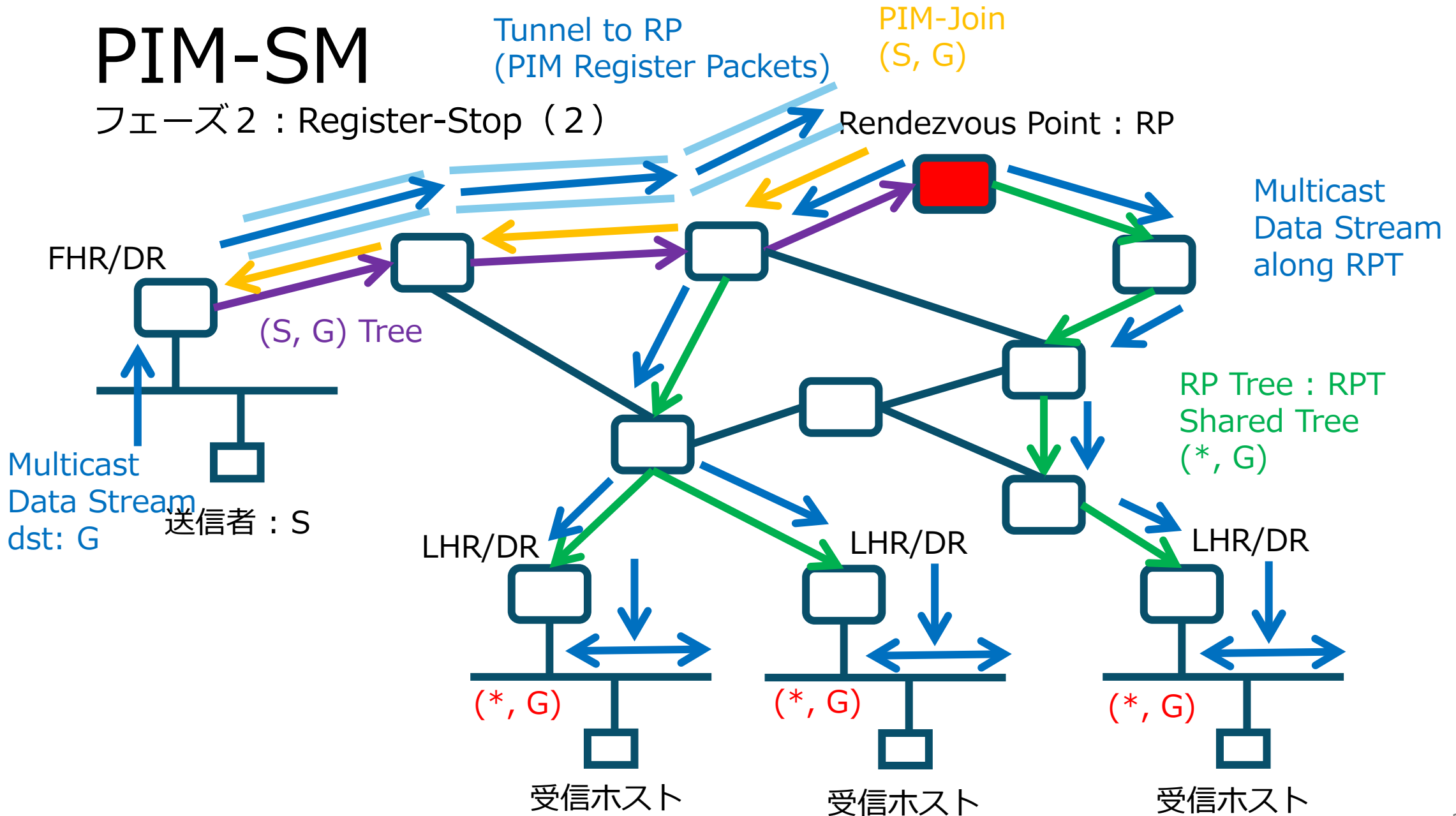
PIM-SM

フェーズ2 : Register-Stop (1)



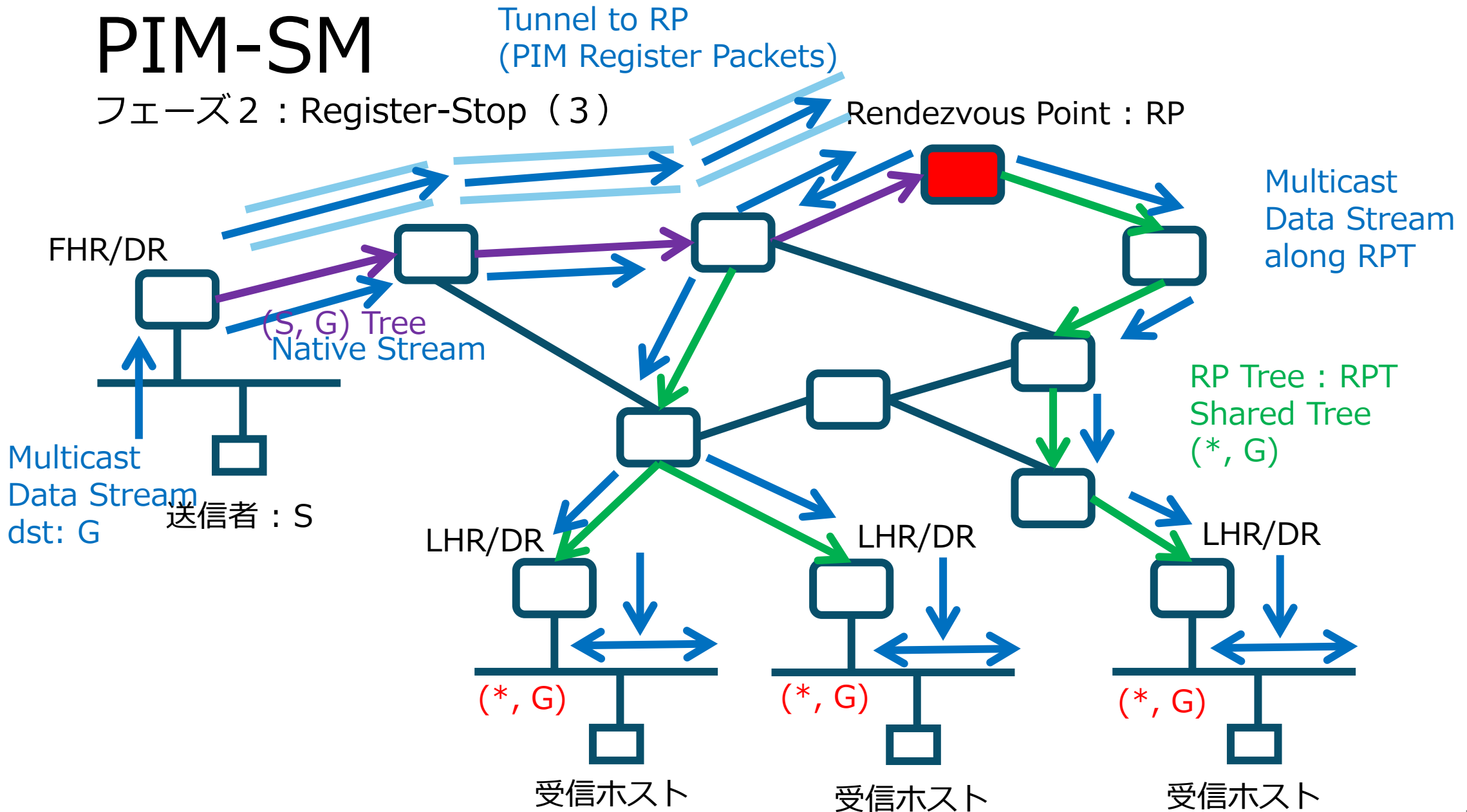
PIM-SM

フェーズ 2 : Register-Stop (2)



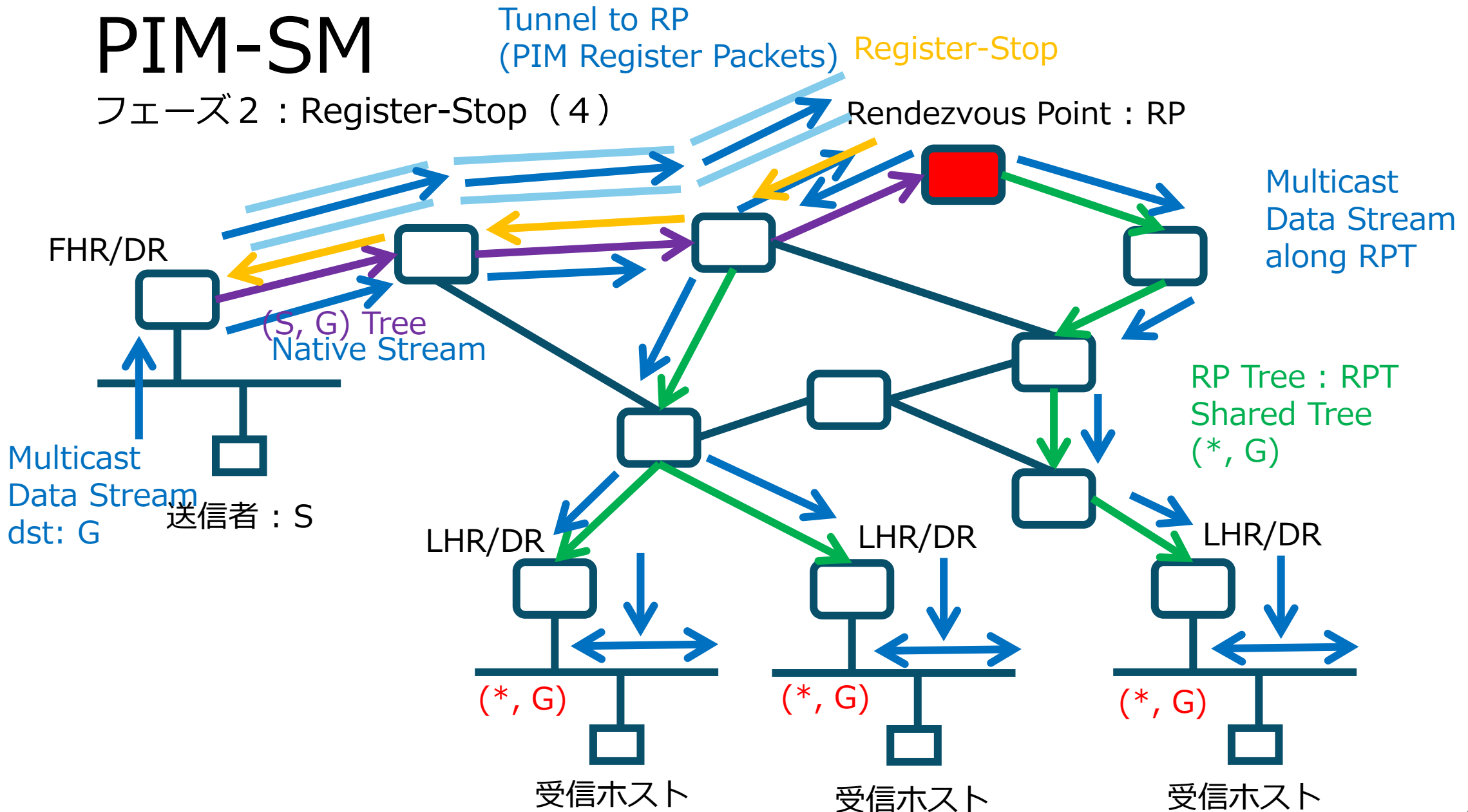
PIM-SM

フェーズ2 : Register-Stop (3)



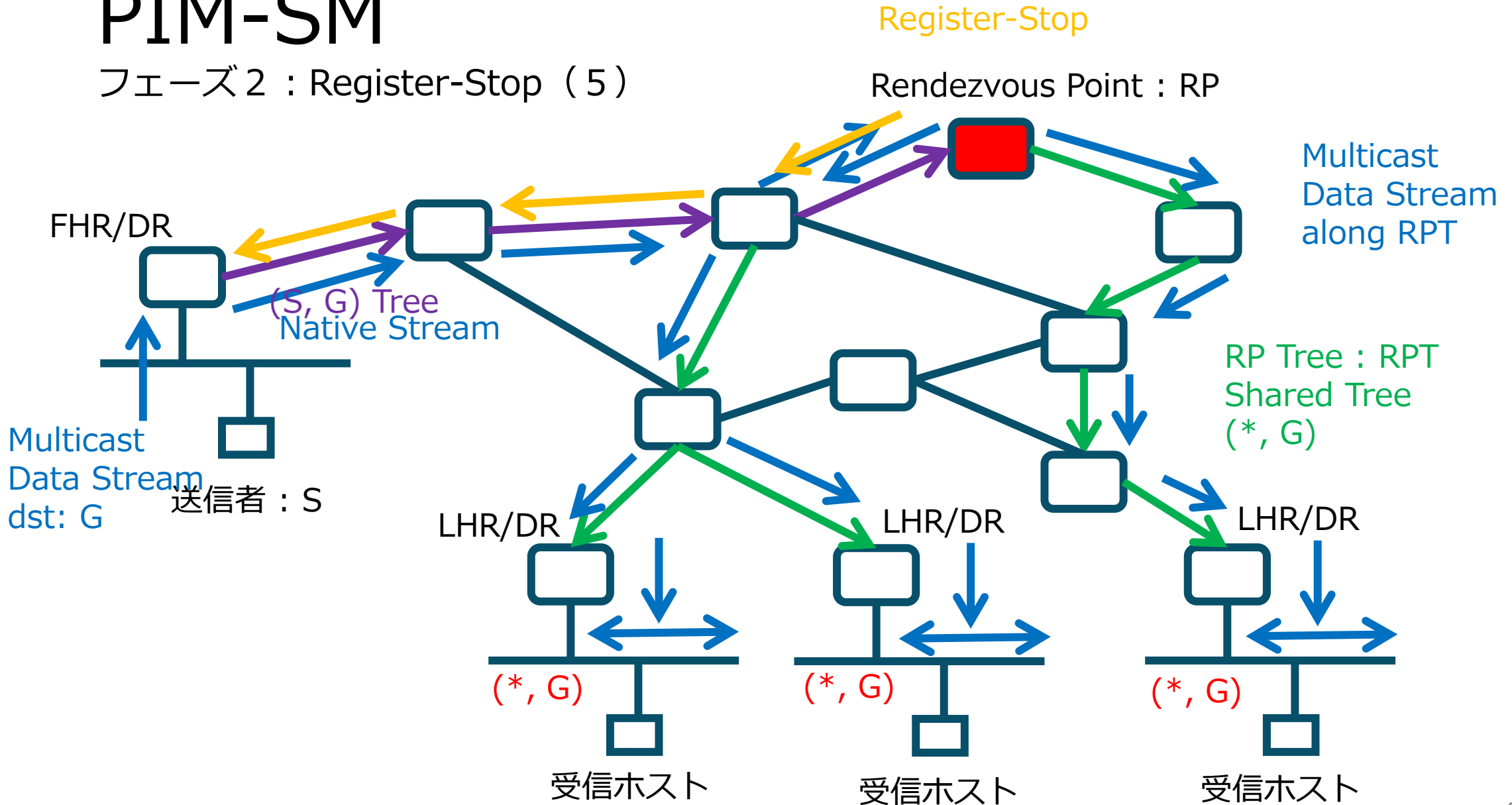
PIM-SM

フェーズ 2 : Register-Stop (4)



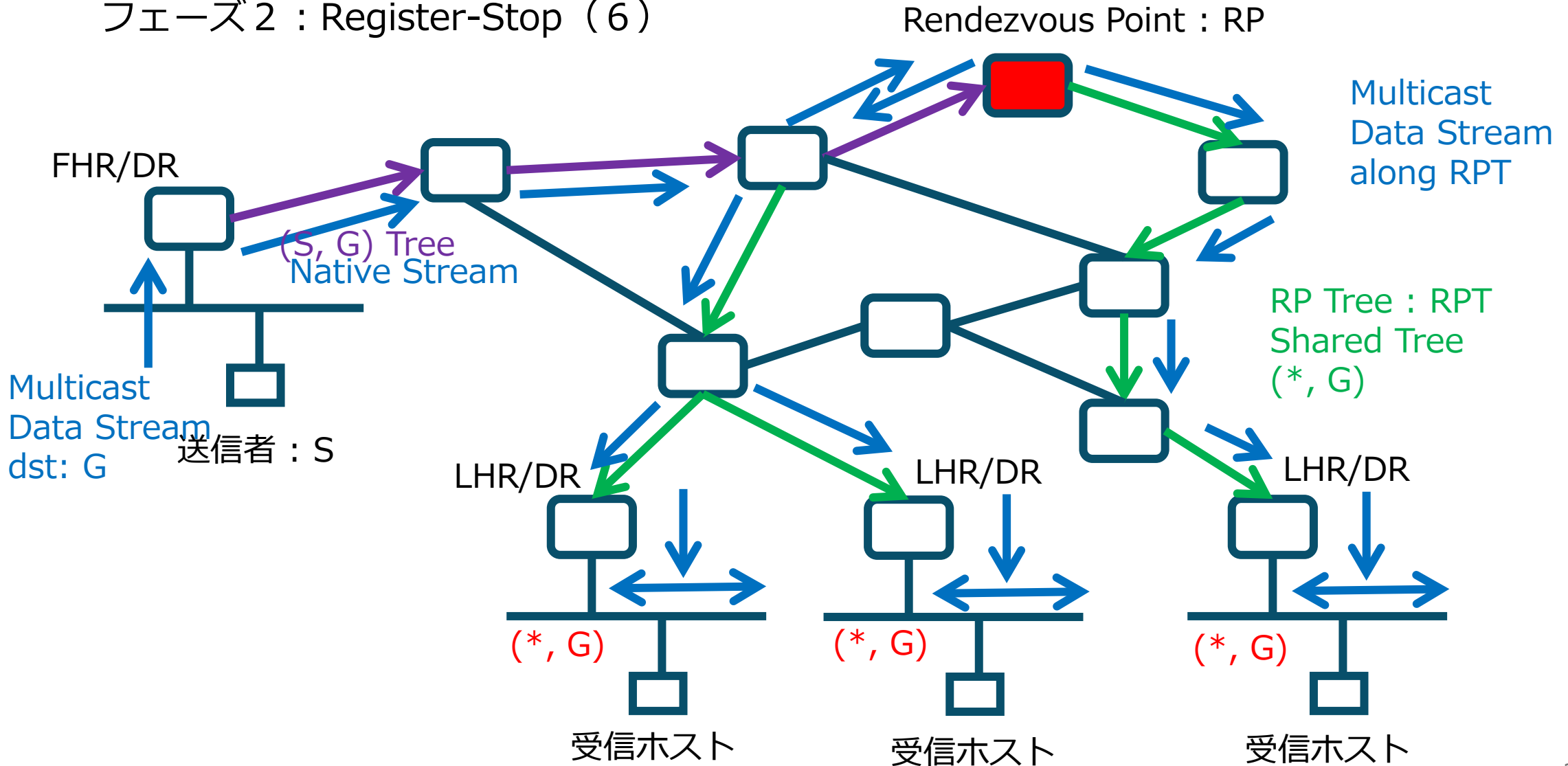
PIM-SM

フェーズ2 : Register-Stop (5)



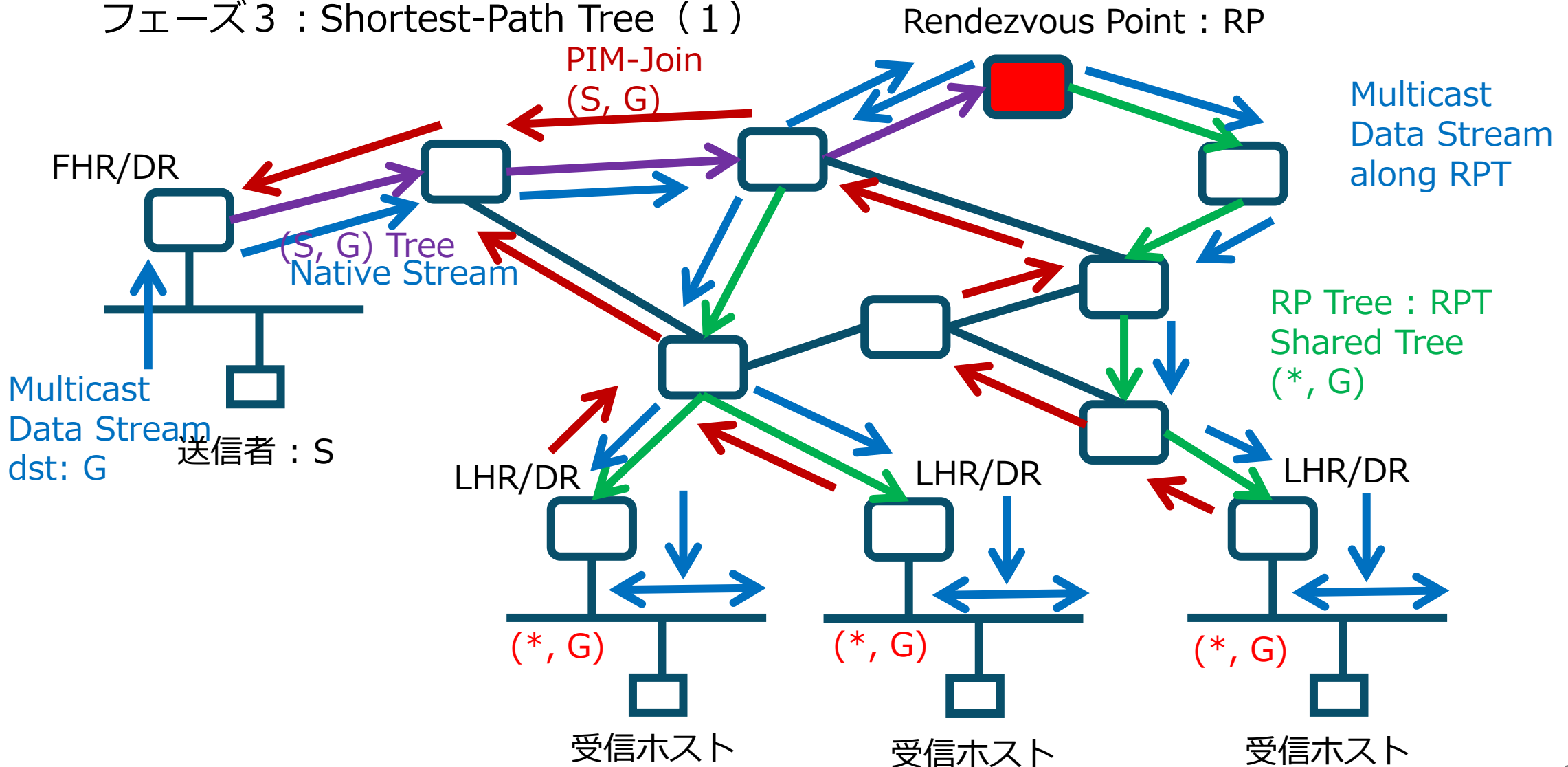
PIM-SM

フェーズ2 : Register-Stop (6)



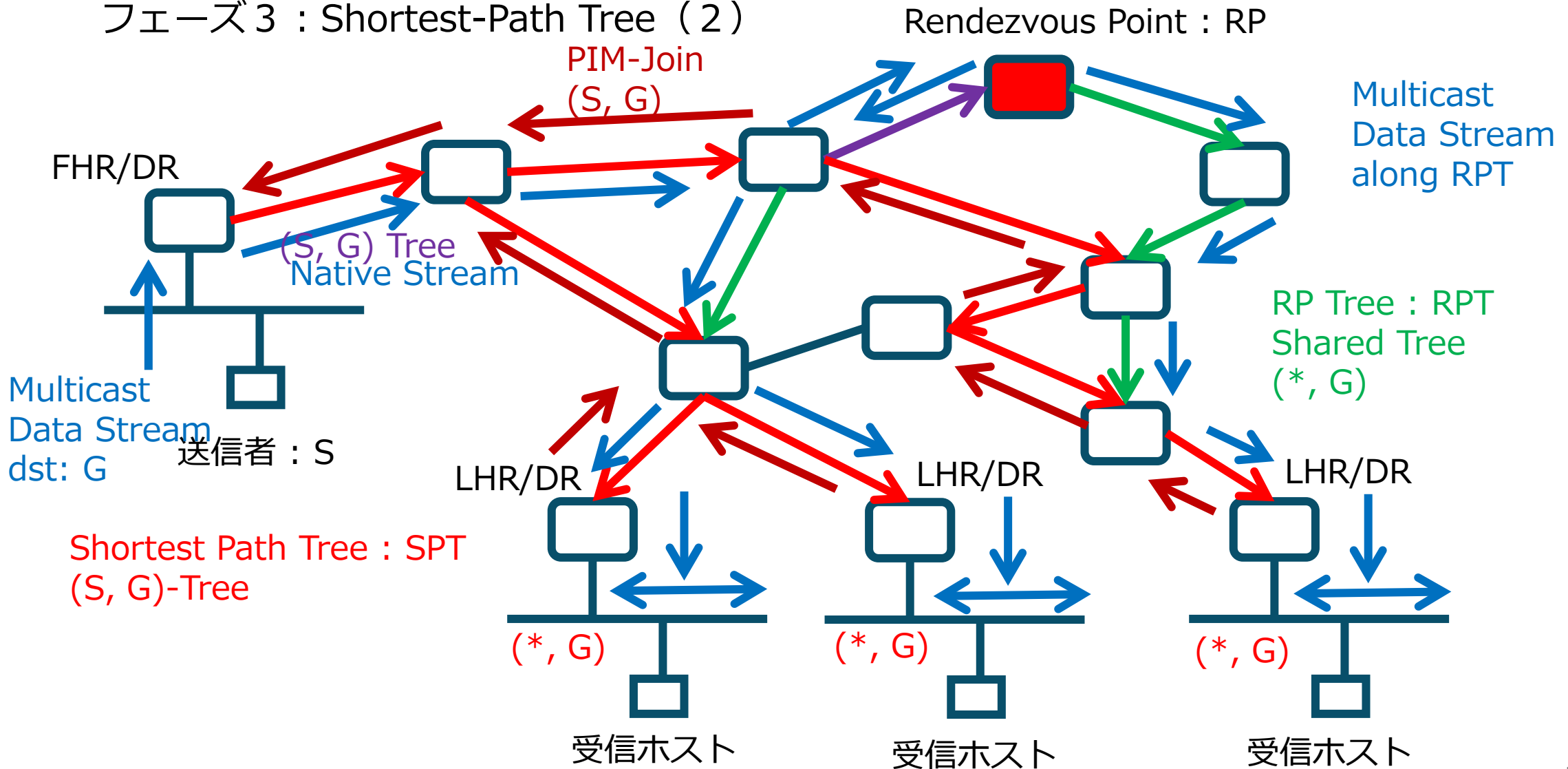
PIM-SM

フェーズ3 : Shortest-Path Tree (1)



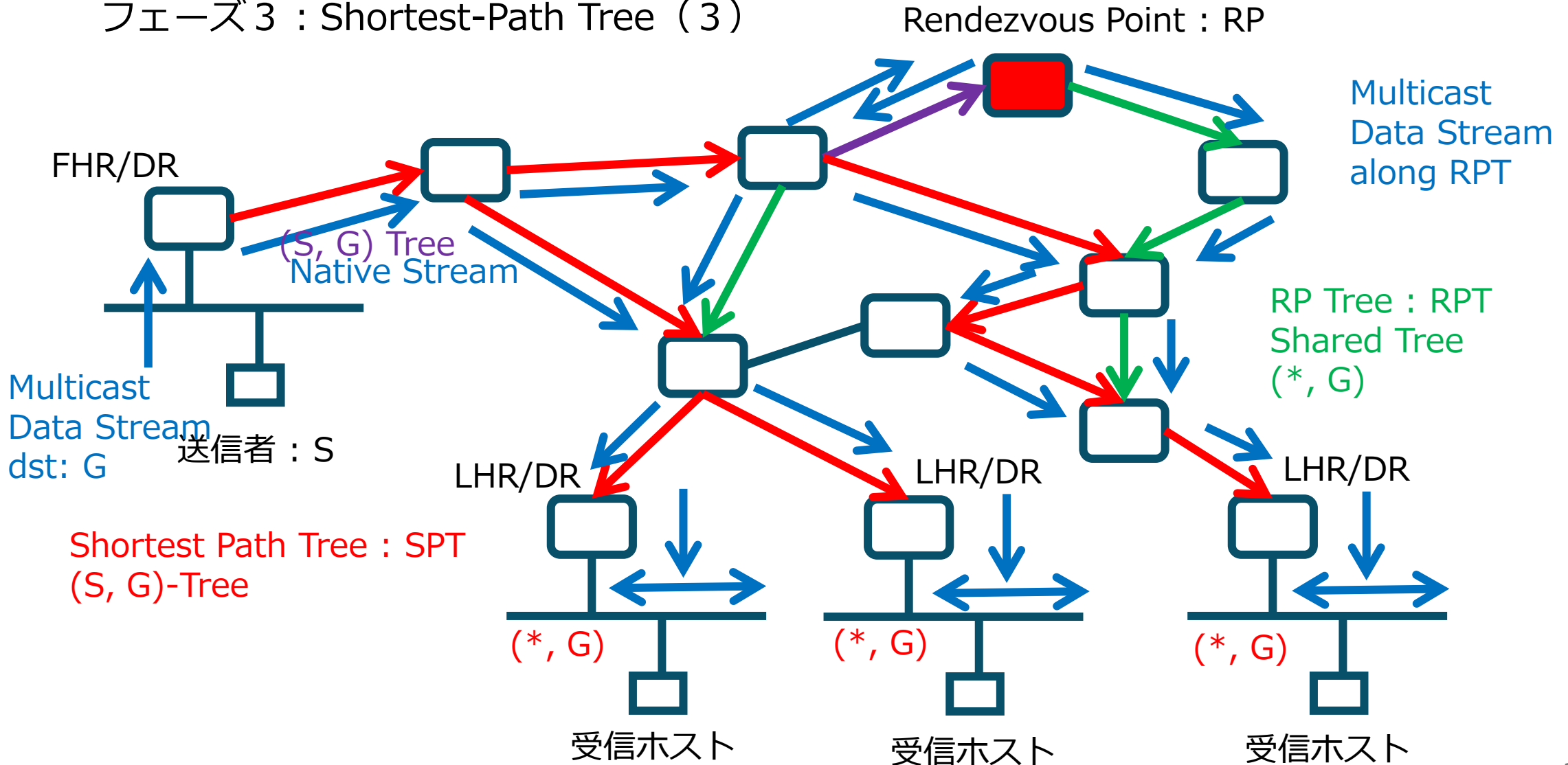
PIM-SM

フェーズ3 : Shortest-Path Tree (2)



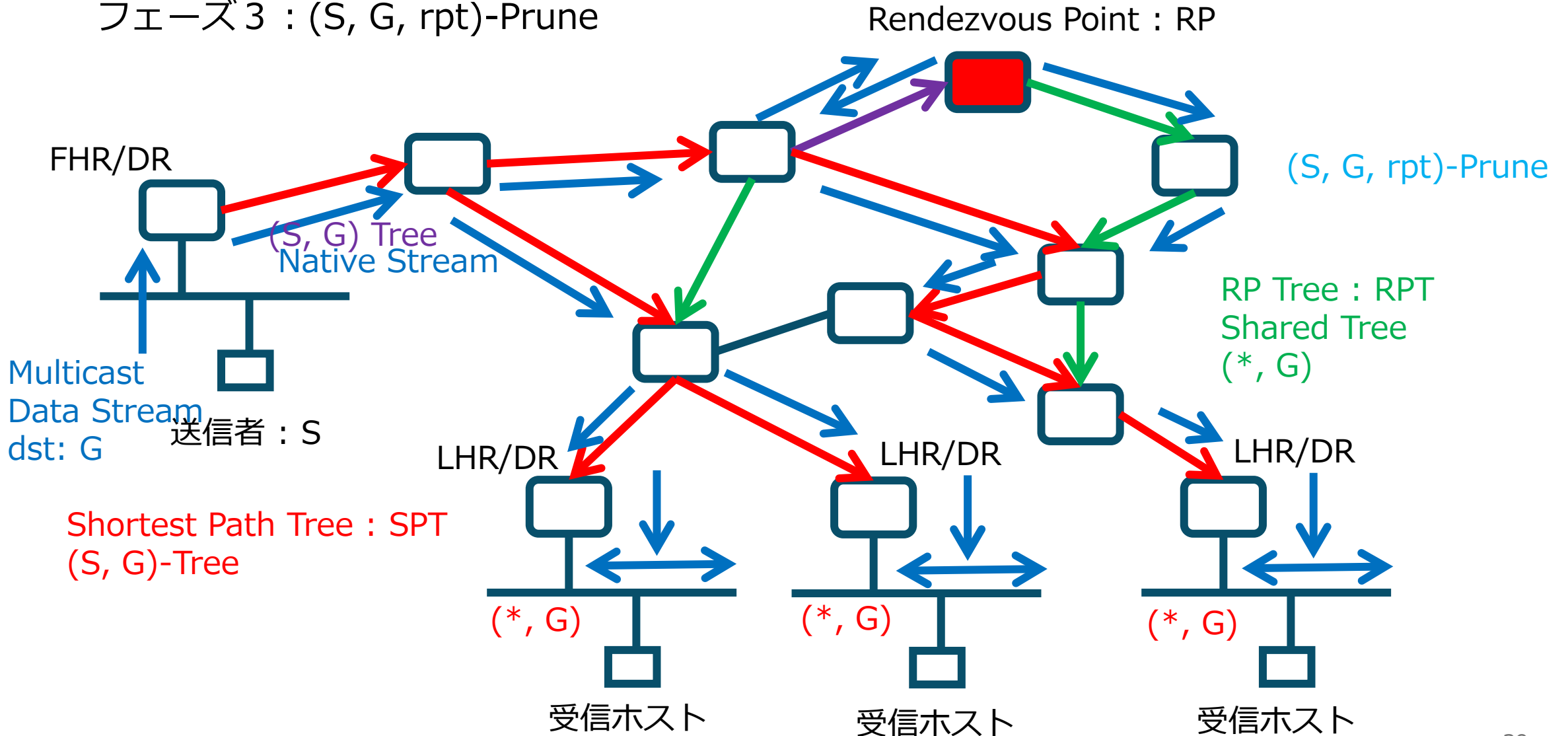
PIM-SM

フェーズ3 : Shortest-Path Tree (3)



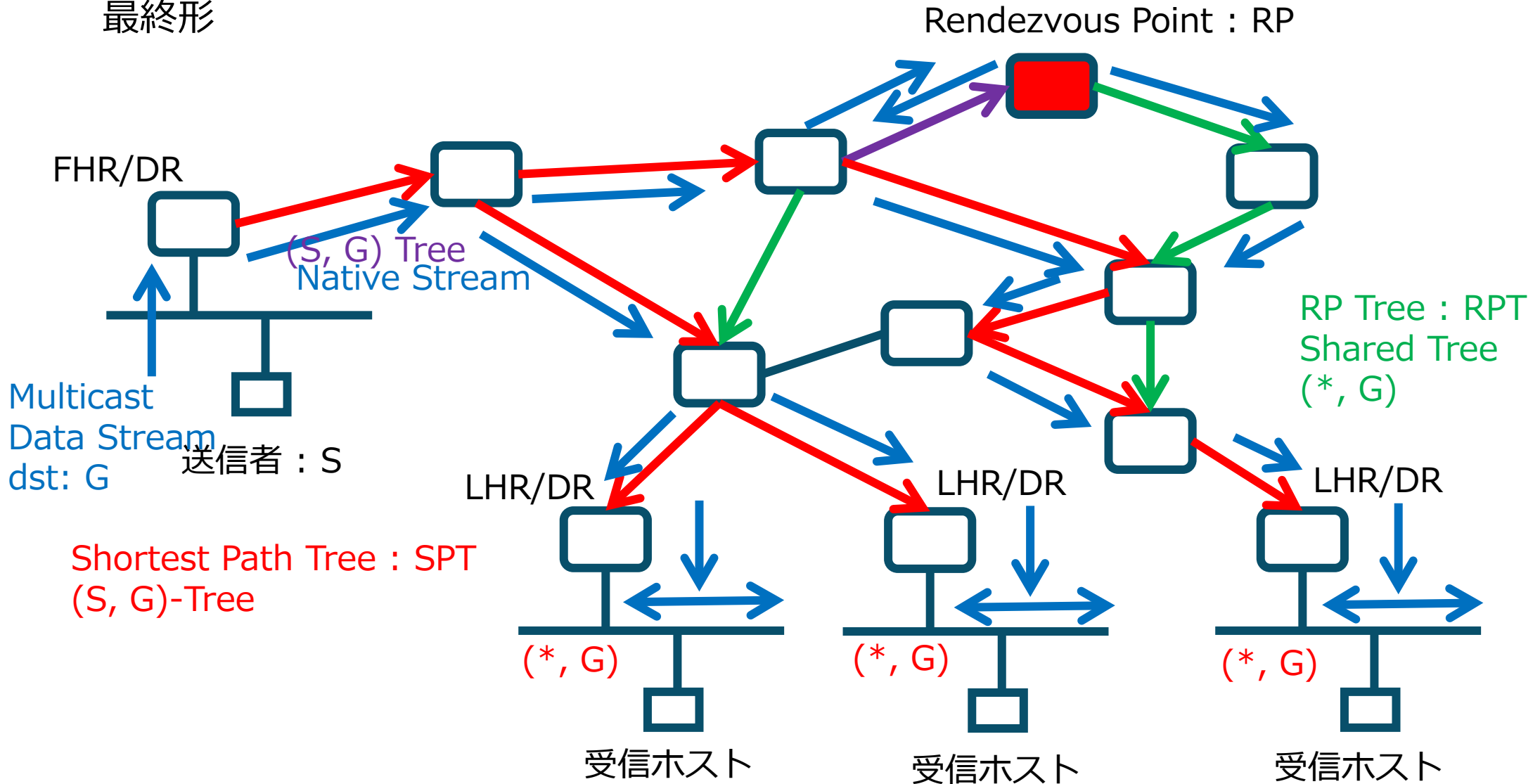
PIM-SM

フェーズ3 : (S, G, rpt)-Prune



PIM-SM

最終形



SSM

- ASM: Any-Source Multicast
 - (*, G) = "Group"
- SSM: Source-Specific Multicast
 - (S, G) = "Channel"
- SSM:
 - IGMPv3, MLDv2, PIM-SSM (Subset of PIM-SM)

その他

- Query Election
- Auto-RP
- BSR
- RPF
- Inter-domain

まとめ

- IGMP動作原理
- PIM-SM動作原理