

Haute école d'ingénierie et d'architecture Fribourg
Hochschule für Technik und Architektur Freiburg

Réseaux IP

448. Multicast & Routage Multicast

Réseaux IP

448. Multicast & Routage Multicast

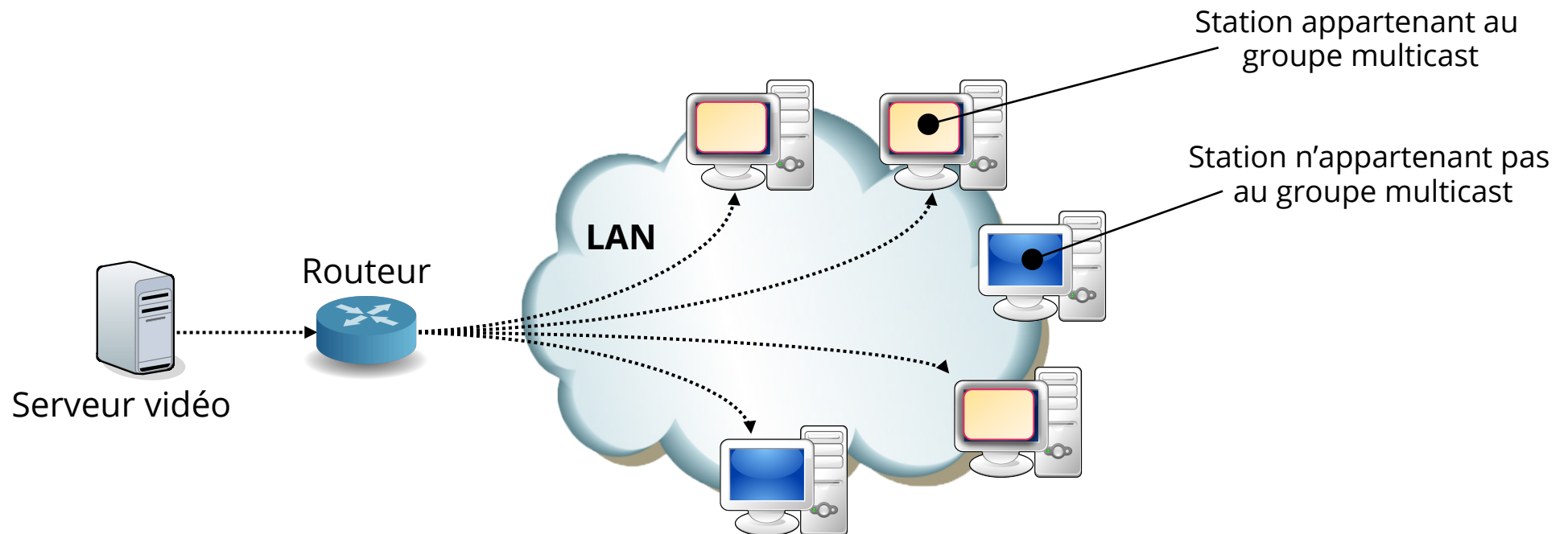
Multicast IPv4, Adressage Multicast, Arbre de distribution, IGMP, Algorithmes de routage multicast, DVMRP, PIM

Références:

- IP Multicast, Cisco Press, 2017, ISBN-13 978-1-58714-459-2
- D. Comer, **Internetworking with TCP/IP, 4th. Ed.**, Prentice-Hall, 2006.
- L. Toutain, **Réseaux locaux et Internet**. Lavoisier Hermes, 2003
- University of Wisconsin, **CS640 Course**, 2006
- RFC 1112, 3376, 2974
- <http://www.routage.org>, <http://www.crir.univ-avignon.fr>

Multicast IPv4

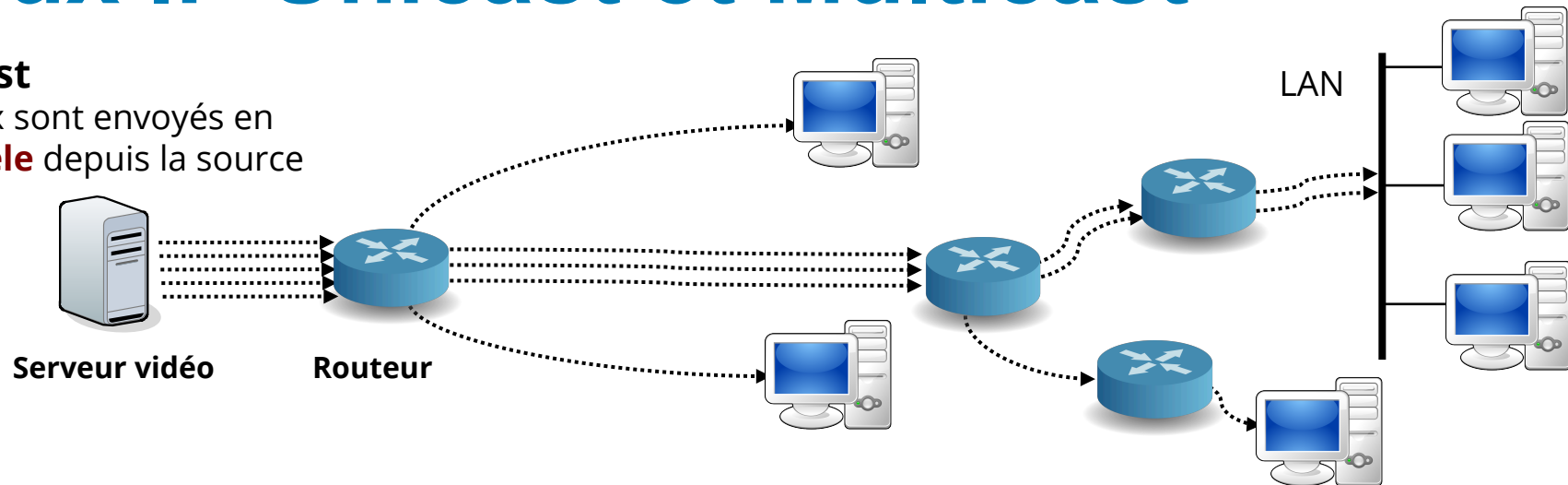
- Le multicast est la transmission simultanée d'un datagramme vers un groupe de stations déterminé au moyen d'une adresse spécifique : **224.0.0.0** à **239.255.255.255**
- Le filtrage des datagrammes multicast se fait à la réception au niveau de la couche 2 au moyen d'adresses MAC spéciales
- L'importance du multicast va croissante. D'abord pour les échanges entre équipements de même type (par exemple tables de routage entre routeurs) ou de même fabricants et plus récemment pour les applications multimédia (diffusion de programmes de télévision par exemple)



Flux IP Unicast et Multicast

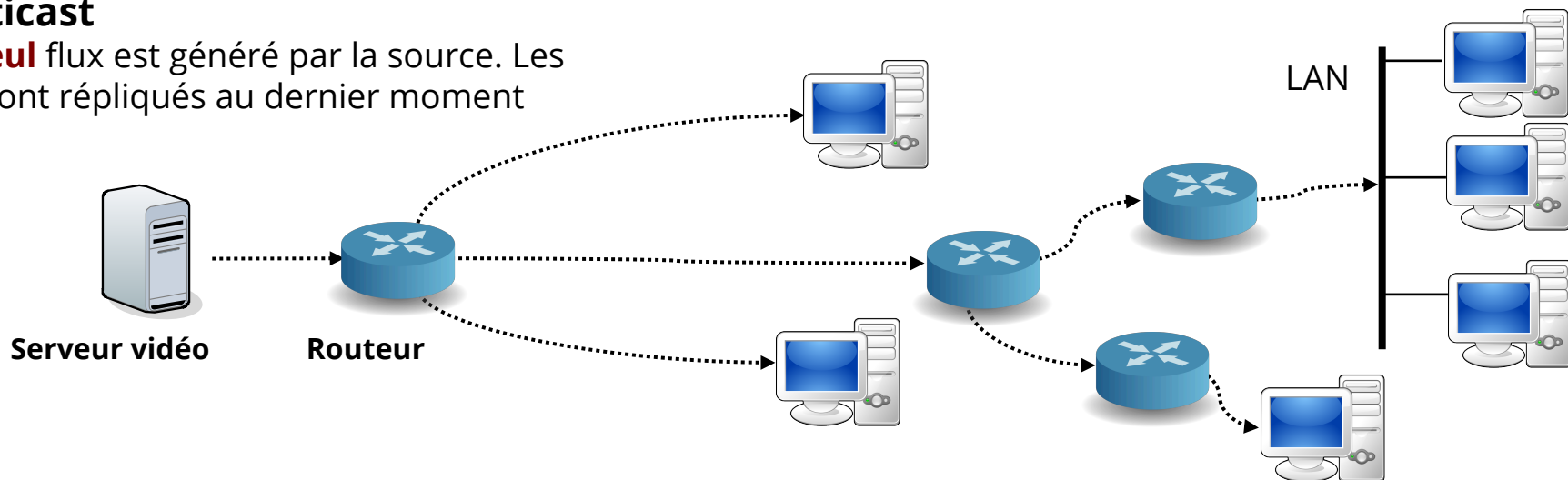
Unicast

Les flux sont envoyés en **parallèle** depuis la source



Multicast

Un **seul** flux est généré par la source. Les flux sont répliqués au dernier moment

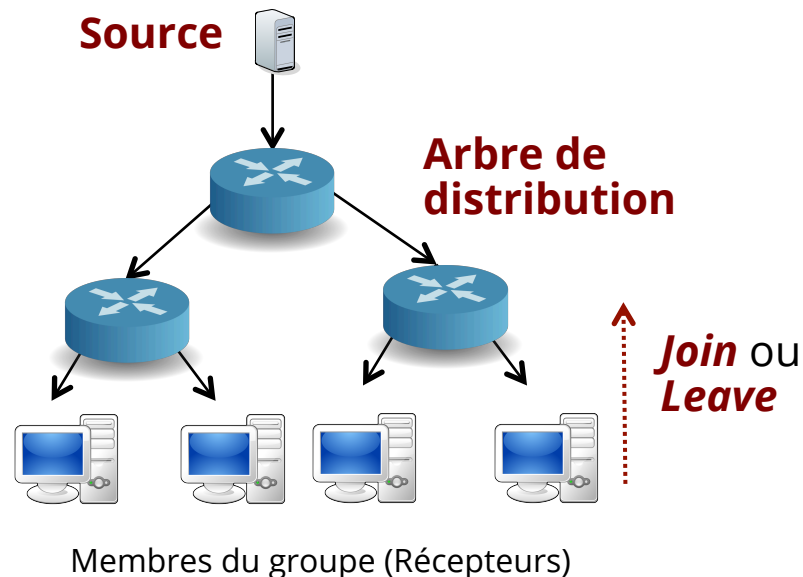


Pourquoi le Multicast ?

- Avantages principaux du Multicast IP
 - Une source peut atteindre un groupe de distribution sans connaître à priori les adresses IP individuelles des stations qui sont membres du groupe
 - Diminution sensible de la bande passante des connexions. C'est par exemple la seule possibilité pour des déploiements à grande échelle de distribution de canaux de télévision (IPTV)
 - Une station peut rejoindre très rapidement un groupe de distribution multicast sans devoir remonter à la source
- Applications typiques
 - Echange d'informations entre équipements de même type (par ex. tables de routage) ou de même fabricant
 - Diffusion Vidéo/Audio (IPTV, Radio IP, Vidéo conférences)
 - Distribution en temps réel de données (*Stock trading, Datacast*)
 - Jeux interactifs
- ... Mais
 - Manque de « *business drivers* » et encore mal supporté sur l'Internet public → Utilisé dans les réseaux « privés » des opérateurs
 - Il n'est pas encore prouvé que le multicast IP puisse supporter des millions de sources et des centaines de millions de récepteurs...
 - Multicast est UDP: *Best effort delivery*, Congestions, Duplications possibles, etc.

Modèle du Multicast IP

Le **Multicast IP** a été proposé par Steve Deering (PhD. 1991)



- Si une source envoie un datagramme multicast à un groupe, tous les membres de ce groupe peuvent le recevoir.
- Les membres d'un groupe peuvent être présents n'importe où dans le réseau IP. Ils s'annoncent ou quittent le(s) groupe(s) auxquels ils participent.
- Le service de multicast IP est du type « *best effort* ». Il n'y a pas de contrôle que toutes les stations du groupe aient effectivement reçu le datagramme.
- La source envoie son flux sans s'inquiéter de sa diffusion. C'est le travail des routeurs d'utiliser un **protocole de routage multicast** pour créer l'arbre de distribution initié par les **récepteurs**.

Le modèle ASM

- **Any Source Multicast** (apparu approx. en 1990): un récepteur rejoint un groupe et reçoit alors potentiellement les flux multicast de toutes les sources qui s'adressent à ce groupe
 - N'importe qui peut devenir membre du groupe, sans autorisation
 - Le même récepteur peut être membre de plusieurs groupes
 - Une source peut envoyer des données vers un groupe multicast, sans en être membre
 - La composition du groupe est dynamique
 - Nul ne connaît la taille du groupe ni l'identité de ses membres.

Inconvénients du modèle ASM

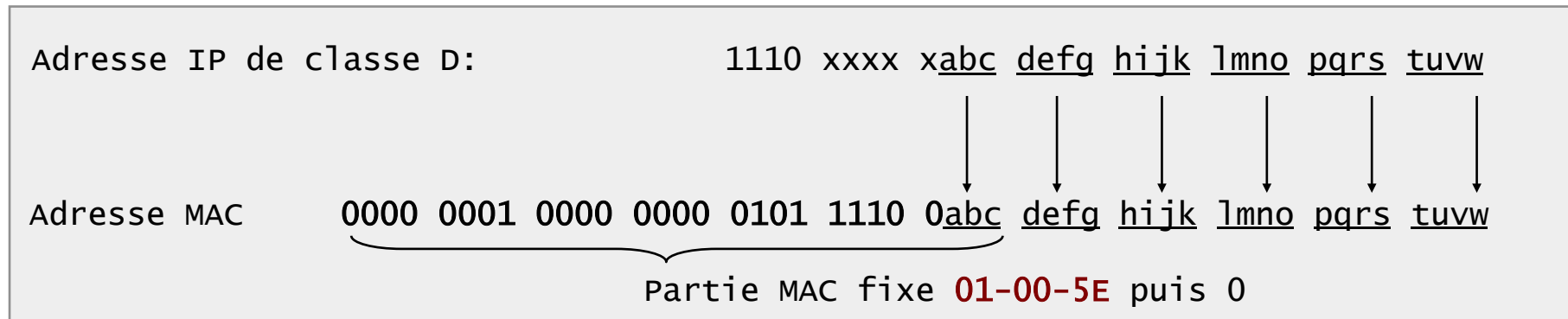
- Le déploiement à grande échelle d'IP Multicast (ASM) a été retardé pour diverses raisons techniques et économiques
- Le mécanisme d'allocation d'adresses est très complexe
- Pas de solution "*scalable*" pour le routage multicast inter-domaine
- Le modèle ouvert n'est pas adapté à des services commerciaux
 - Pas de contrôle strict des émetteurs et des récepteurs multicast
 - Facturation difficile à gérer
- Un modèle de service simplifié était donc nécessaire

Le modèle SSM

- **Source Specific Multicast** (apparu approx. en 2000):
 - Les groupes multicast sont remplacés par des **canaux multicast**
 - Un canal est identifié par la paire d'adresses (S, G) :
 - S – adresse unicast de la source
 - G – adresse multicast du groupe
 - Seulement la source S peut envoyer des données sur le canal (S, G)
 - Les canaux (S₁, G) et (S₂, G) sont différents par définition

Adressage Multicast (1)

- Les groupes multicast peuvent être temporaires ou pré-définis.
- Le service de multicast est du type « *best effort* », il n'y a pas de contrôle que toutes les stations du groupe aient effectivement reçu le datagramme.
- Les adresses IP de multicast sont transposées en adresse MAC de multicast sur le LAN selon le schéma suivant:



L'adresse IP de **classe D** est convertie en une adresse MAC multicast avec les premiers 25 bits fixes (01-00-5E puis 0) et ensuite en recopiant les 23 derniers bits de l'adresse IP de classe D. Comme les 5 bits xxxx x ne sont pas recopiés, 32 adresses multicast IP peuvent être transposées sur la même adresse MAC.

Adressage Multicast (2)

Adresses locales de lien réservées: **224.0.0.0/24**, transmises avec TTL=1:

224.0.0.0	Base Address (Reserved)	[RFC1112,JBP]
224.0.0.1	All Systems on this Subnet	[RFC1112,JBP]
224.0.0.2	All Routers on this Subnet	[JBP]
224.0.0.3	Unassigned	[JBP]
224.0.0.4	DVMRP Routers	[RFC1075,JBP]
224.0.0.5	OSPFIGP OSPFIGP All Routers	[RFC1583,JXM1]
224.0.0.6	OSPFIGP OSPFIGP Designated Routers	[RFC1583,JXM1]
224.0.0.7	ST Routers	[RFC1190,KS14]
224.0.0.9	RIP2 Routers	[GSM11]
224.0.0.11	Mobile-Agents	[Bill Simpson]
224.0.0.12-224.0.0.255	Unassigned	[JBP]

Autres adresses réservées: **224.0.1.0/24**, transmises avec TTL≥1:

224.0.1.0	VMTP Managers Group	[RFC1045,DRC3]
224.0.1.1	NTP Network Time Protocol	[RFC1119,DLM1]

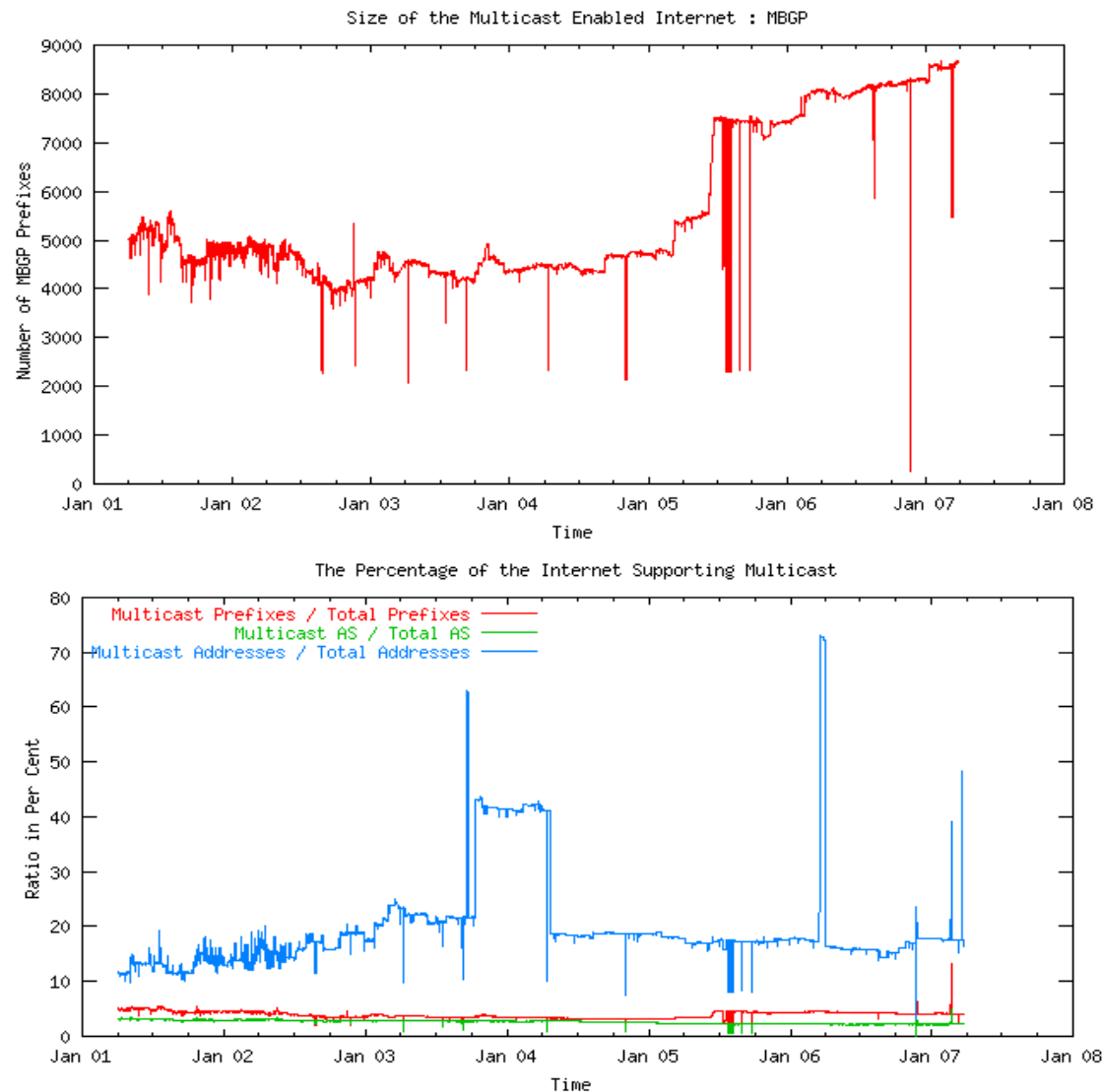
...

Adressage Multicast (3)

- **Global Scope Addresses: 224.2.0.0/16**, utilisée avec les protocoles SDP/SAP (*Session Description/Announcement Protocol*, RFC 2974, *multicast session advertising*). L'allocation d'adresse est dynamique, c'est à dire la génération aléatoire d'une adresse qui n'est actuellement pas annoncée.
- Adresses multicast « **privées** » : **239.0.0.0/8**, similaire au concept d'adresses IPv4 privées (RFC 1918, *administratively scoped addresses*)
- Adresses multicast « **statiques** »: **233.x.y.0/8** (RFC 3180). Permet d'assigner de manière unique et globale un bloc d'adresse multicast /24 pour chaque système autonome. Les 2 octets du numéro d'AS sont encodé dans l'adresse multicast.
Par exemple: AS de Switch = 559 = $2 \cdot 256 + 47 \rightarrow 233.2.47.0/24$ appartient à Switch.
- **Adresses SSM** (*Source Specific Multicast*) 232.0.0.0 à 232.255.255.255 destinées au *Public Internet Broadcast*

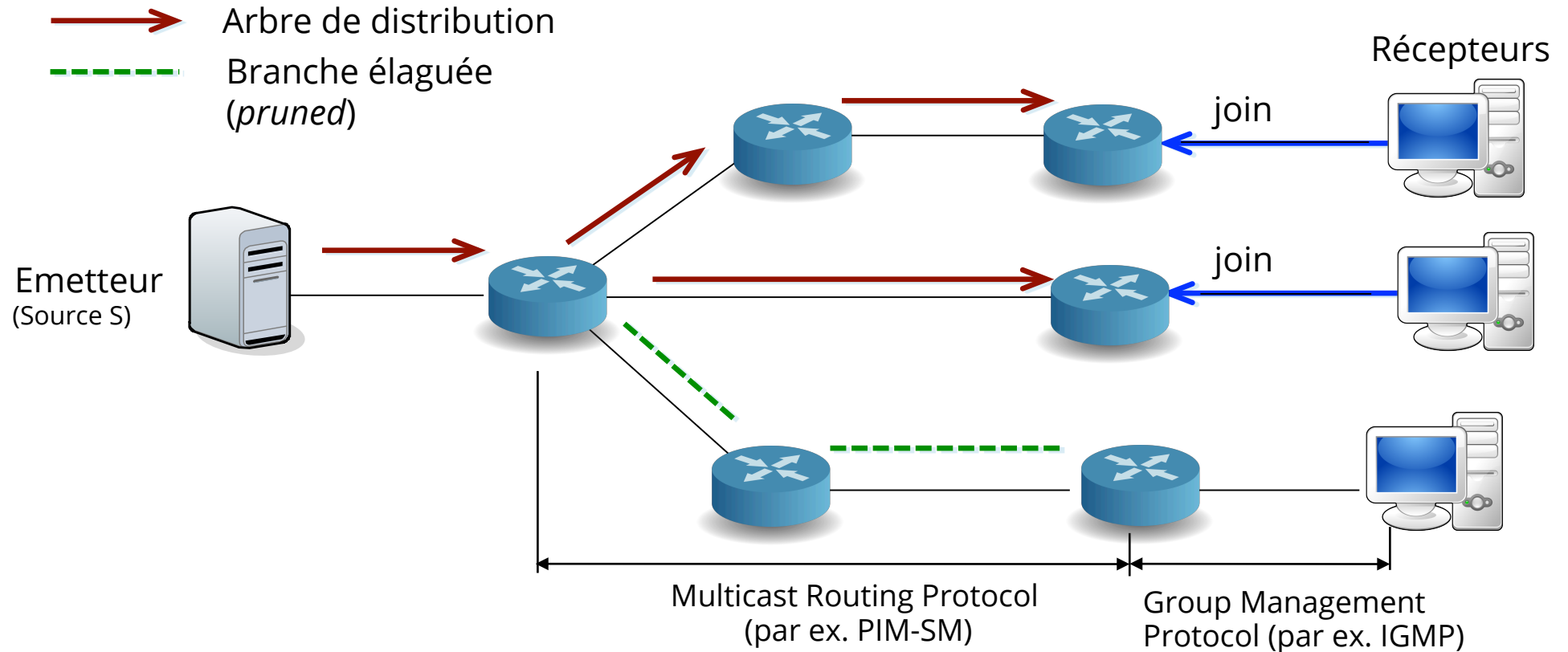
... encore souvent beaucoup de travail „à la main“

Déploiement du multicast sur Internet



<http://www.multicasttech.com/status/>

Protocoles associés au multicast IP



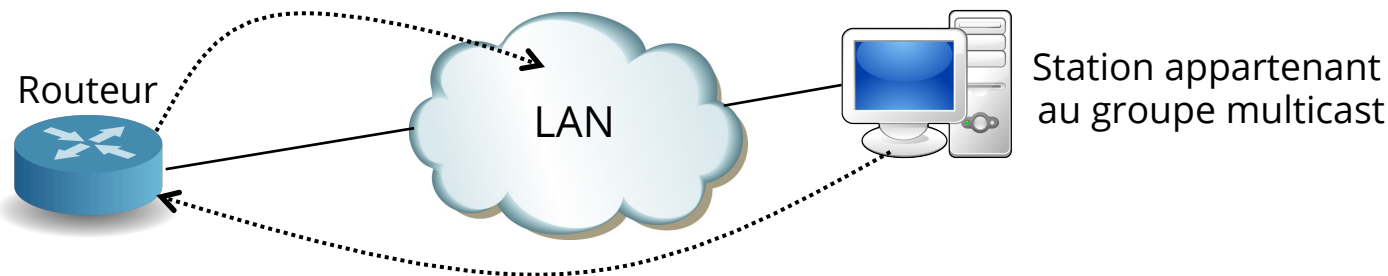
- **Group Management Protocol** – permet aux récepteurs de joindre/quitter un groupe multicast de manière dynamique. Ce dialogue se fait avec le routeur le plus proche.
- **Multicast Routing Protocol** – permet aux routeurs de construire un arbre de distribution entre l'(les) émetteur(s) et le(s) récepteur(s) d'un groupe multicast.

Gestion du multicast IP: le protocole IGMP

Le protocole **IGMP** (*Internet Group Management Protocol*) gère les groupes multicast sur un LAN. En particulier sa tâche est de permettre l'addition de nouveaux membres aux groupes et d'enregistrer leur départ de ces groupes.

- Les stations peuvent annoncer spontanément qu'elles joignent ou quittent un groupe multicast.
- Le routeur multicast peut demander si un membre du groupe est présent. Les stations peuvent répondre à cette requête en s'annonçant au routeur.

Demande d'appartenance (*Query*) entendu par toutes les stations qui appartiennent à un groupe multicast (tous les groupes multicast 224.0.0.1 ou un groupe particulier)

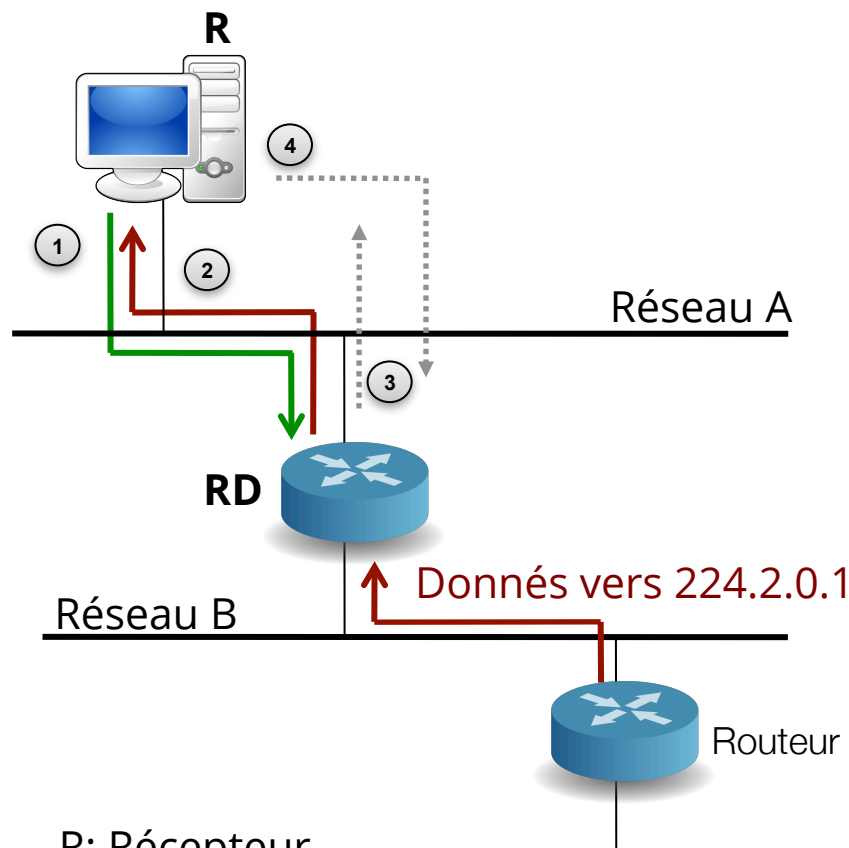


Rapport d'appartenance (*Membership Report*) à un groupe (réponse à une demande ou annonce spontanée) ou message que la station quitte (*Leave Group*) un groupe multicast

Versions d'IGMP

- IGMP est un protocole d'interaction entre
 - le(s) routeur(s) multicast du LAN
 - et les hôtes multicast du LAN
- Il permet à un hôte de s'abonner (désabonner) à un groupe et de dire au routeur :
"envoyez-moi une copie des paquets de cette adresse de groupe"
- Trois versions du protocole IGMP existent:
 - **IGMPv1** (RFC 1112, 1989): comprend uniquement les messages *"Membership Query"* et *"Membership Report"*. Cette version devrait devenir obsolète.
 - **IGMPv2** (RFC 2236, 1997): supporte en plus le message *"Leave Group"*. Le routeur envoie alors une demande « *group specific* » pour savoir si une autre station est écoule le flux multicast. Si ce n'est pas le cas le flux est retiré.
 - **IGMPv3** (RFC 3376/2002, RFC 4604/2006): ajoute la possibilité de s'inscrire à un flux (*Source Specific Multicast, SSM*) particulier (S,G) ou d'exclure des sources de (*,G) (*Source Filtered Multicast, SFM*). Le routeur mémorise les souscriptions et peut ainsi retirer le flux multicast immédiatement après que la dernière station ait quitté la souscription.

IGMPv2 – Joindre un groupe

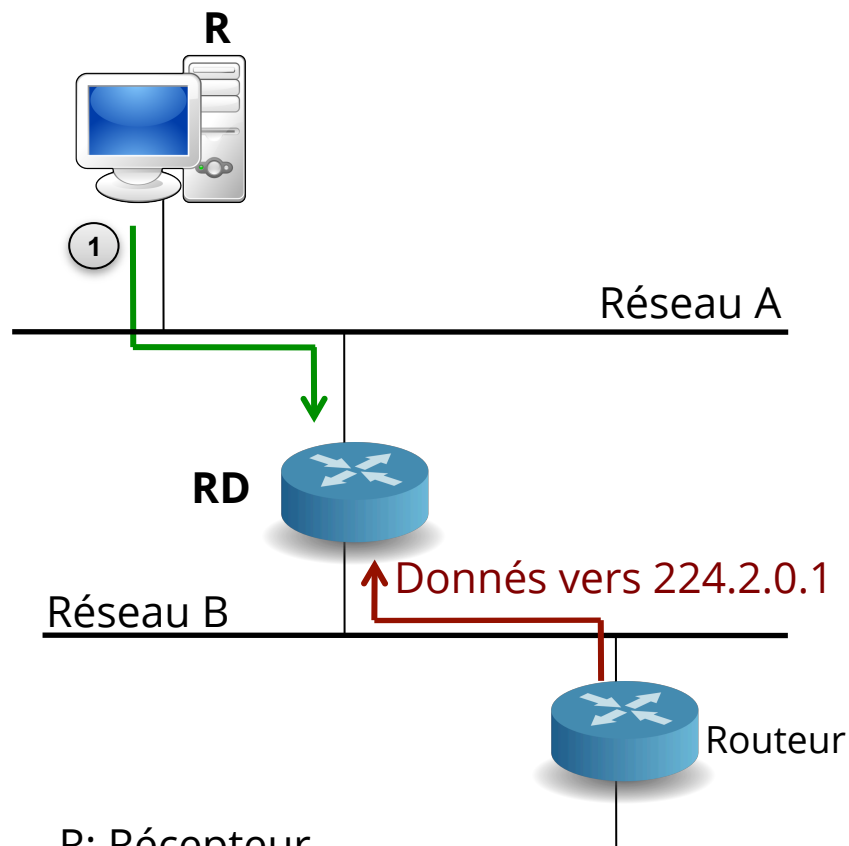


R: Récepteur
RD: Routeur Désigné

Exemple: Le récepteur R veut joindre le groupe multicast 224.2.0.1

1. R envoie un message « **IGMP Membership-Report** » à l'adresse **224.2.0.1**
2. Le RD reçoit ce message et émet les paquets pour l'adresse 224.2.0.1 sur le réseau A (s'il ne le fait pas déjà)
3. Le DR émet périodiquement un message « **IGMP Membership-Query** » à l'adresse **224.0.0.1** (« *all stations* »)
4. R répond par un message « **IGMP Membership-Report** » à l'adresse **224.2.0.1**

IGMPv2 – Quitter un groupe



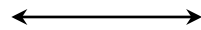
R: Récepteur
RD: Routeur Désigné

Exemple: R veut quitter le groupe 224.2.0.1

1. R envoie un message « **IGMP Leave-Group** » à l'adresse **224.0.0.2**
2. Le RD reçoit ce message et stoppe l'émission des paquets pour l'adresse 224.2.0.1 sur le réseau A, pour autant qu'il n'y ait plus de membre de ce groupe sur le réseau A

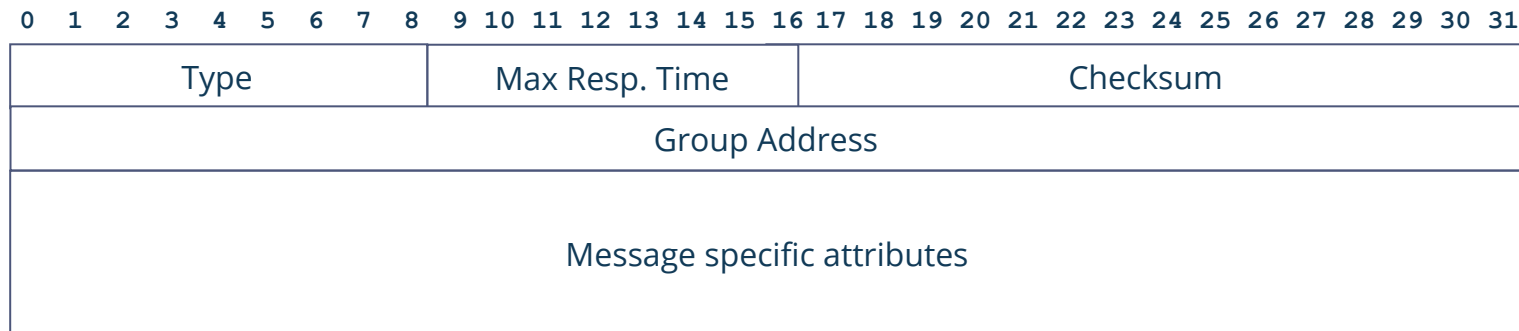
Format des datagrammes IGMP

≥20 octets



IGMP est identifié par le numéro de protocole **0x02** dans le paquet IP.
Les messages IGMP ont le champ TTL de l'entête IP avec la valeur 1.

En-tête commune IGMP



Champs du datagramme IGMP (1)

- **Type** : Il y a trois types de message IGMP pour l'interaction entre la station et le routeur:
 - 0x11 (Membership Query) : avec deux sous-types de messages:
 - General Query, utilisé pour apprendre quel groupe a des membres sur le réseau attaché. Le champ Group Address contient la valeur 0.
 - Group Specific Query: utilisé pour apprendre si un groupe particulier a des membres sur le réseau attaché. Le champ Group Address contient la valeur du groupe questionné.
 - 0x12 Membership Report (pour la compatibilité avec IGMPv1)
 - 0x16 Membership Report (pour IGMPv2)
 - 0x17 Leave Groupe
 - 0x22 Membership Report (pour IGMPv3)
- **Max. Response Time** : Utilisé uniquement avec le message « *Membership Query* » et spécifie le temps maximum autorisé qu'ont les récepteurs avant de répondre avec un « *Membership Report* ». Exprimé en 1/10s. Ce champ est mis à 0 pour tous les autres messages. Il permet aux routeurs de gérer le temps de latence lorsque le dernier membre d'un groupe quitte celui-ci et quand le protocole de routage est notifié qu'il n'y a plus de membres

Champs du datagramme IGMP (2)

- **Checksum** : Complément à 1 de la somme des mots de 16 bits de l'ensemble du payload IGMP
 - **Group Address** : Contient l'adresse à laquelle on s'adresse
 - **Message Specific Attributes** : Varie en fonction du type de message (*Membership Query*, *Membership Report*, ...).
 - Avec IGMPv3, des paramètres complémentaires sont utilisés, tels que par exemple
 - QRV (Querier's Robustness Variable),
 - QQIC (Querier's Query Interval Code),
 - Number of sources,
 - etc.
- Se référer au RFC 3376 pour plus de détails.

Restreindre le trafic en couche 2

Comme les adresses multicast MAC ne sont jamais utilisées comme adresses de sources, les switches (et DSLAMs) diffusent les adresses multicast sur tous les ports par défaut (*broadcast*). Un trafic inutile considérable est ainsi généré. On peut réduire le trafic multicast au moyen de certains protocoles:

- **IGMP snooping**: les switches interprètent les messages IGMP qui les traverse, en particulier pour éviter de diffuser des trames sur un réseau si plus aucun membre du groupe multicast n'est présent. Les switches incluent dans leur table de forwarding les adresses MAC « Multicast » qu'ils auront apprises au travers des messages IGMP. Il s'agit de la solution la plus déployée.
- **CGMP** (Cisco Group Management Protocol). Protocole propriétaire de signalisation entre routeurs et switches.
- **MVRP/MMRP** (IEEE 802.1D). Protocole de niveau 2 (spécifique à Ethernet), déployer dans les switches et les stations. Très peu répandu, spécialement dans les stations.
- **IGMP Proxy**: le switch intercepte le trafic IGMP et se comporte vis-à-vis du routeur comme une seule station multicast. Permet de diminuer le trafic IGMP.

Routage multicast IP

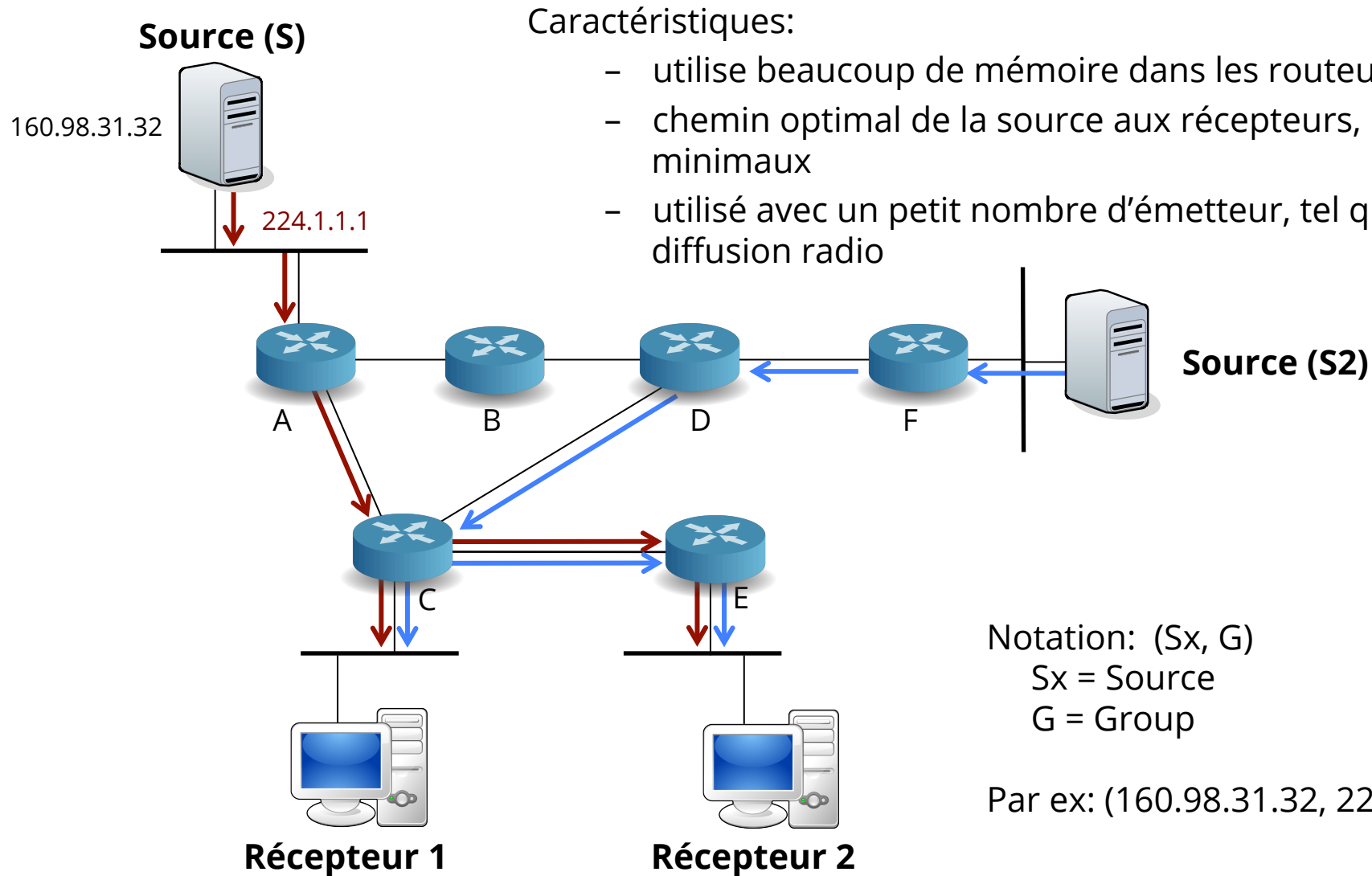
Problème: comment partager l'information des groupes entre les routeurs afin de mieux implémenter le routage pour la distribution des données.

Solution: Une structure avec un arbre de distribution (ou d'acheminement). Ils s'agit d'arbre de distribution spécifique à une source (ou du chemin le plus court, *source tree*) ou d'arbre de distribution partagé (*shared tree*)

Les protocoles de routage multicast sont implémentés en conjonction avec le protocole IGMP et se base sur les protocoles de routage **unicast**.

Un protocole de routage **unicast** est concerné par le chemin vers la **destination**, alors qu'un protocole de routage **multicast** par le chemin vers la **source** (utilise l'algorithme RPF, *Reverse Path Forwarding*)

Arbre de distribution spécifique à une source (Source Specific Tree)



Caractéristiques:

- utilise beaucoup de mémoire dans les routeurs ($G \times S$)
- chemin optimal de la source aux récepteurs, les délais sont minimaux
- utilisé avec un petit nombre d'émetteur, tel que la diffusion radio

Notation: (Sx, G)

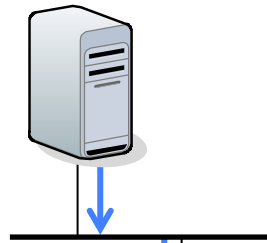
Sx = Source

G = Group

Par ex: (160.98.31.32, 224.1.1.1)

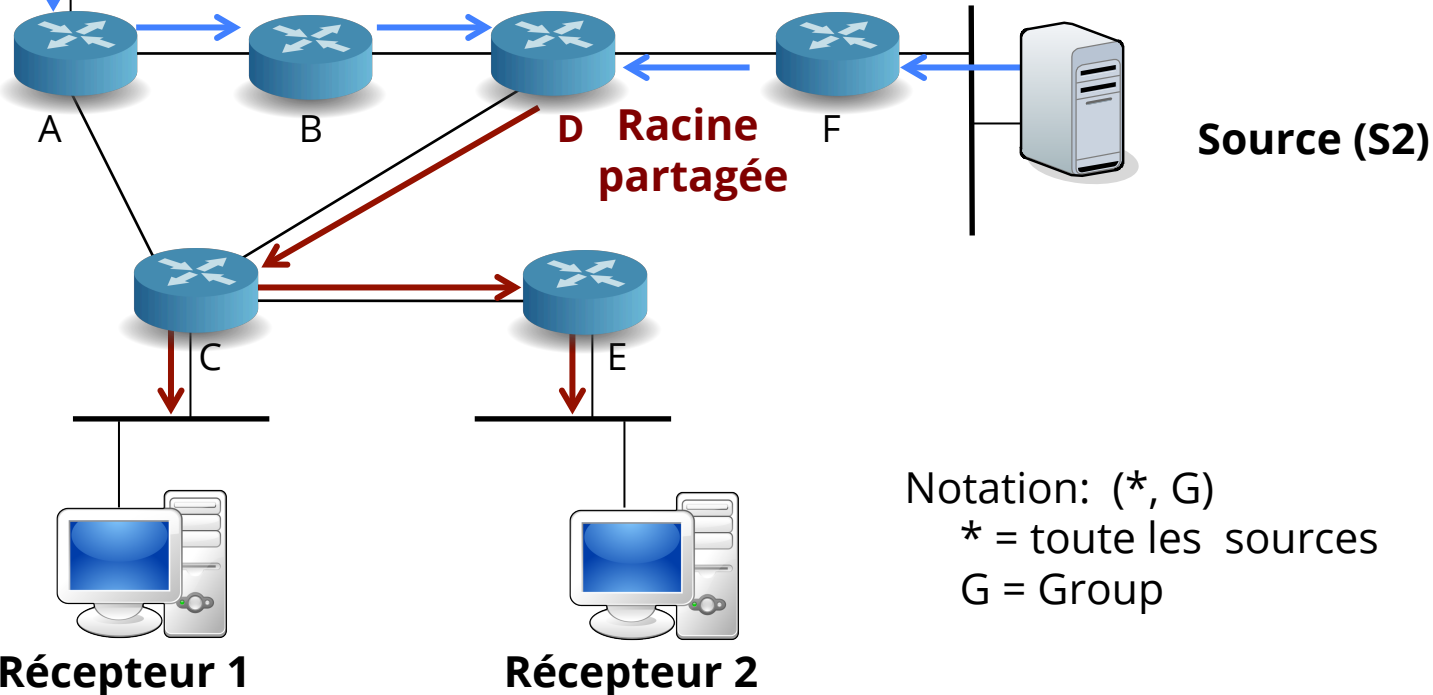
Arbre de distribution partagé (*Shared Tree*)

Source (S1)



Caractéristiques:

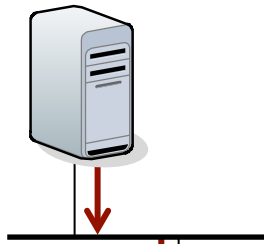
- utilise peu de mémoire dans les routeurs (G)
- chemin sub-optimaux de la source aux récepteurs, introduit des délais supplémentaires (source à la racine)
- peut introduire des transfert dupliqués des données (chemin de la source à la racine et de la racine aux récepteurs)



Notation: (*, G)
* = toute les sources
G = Group

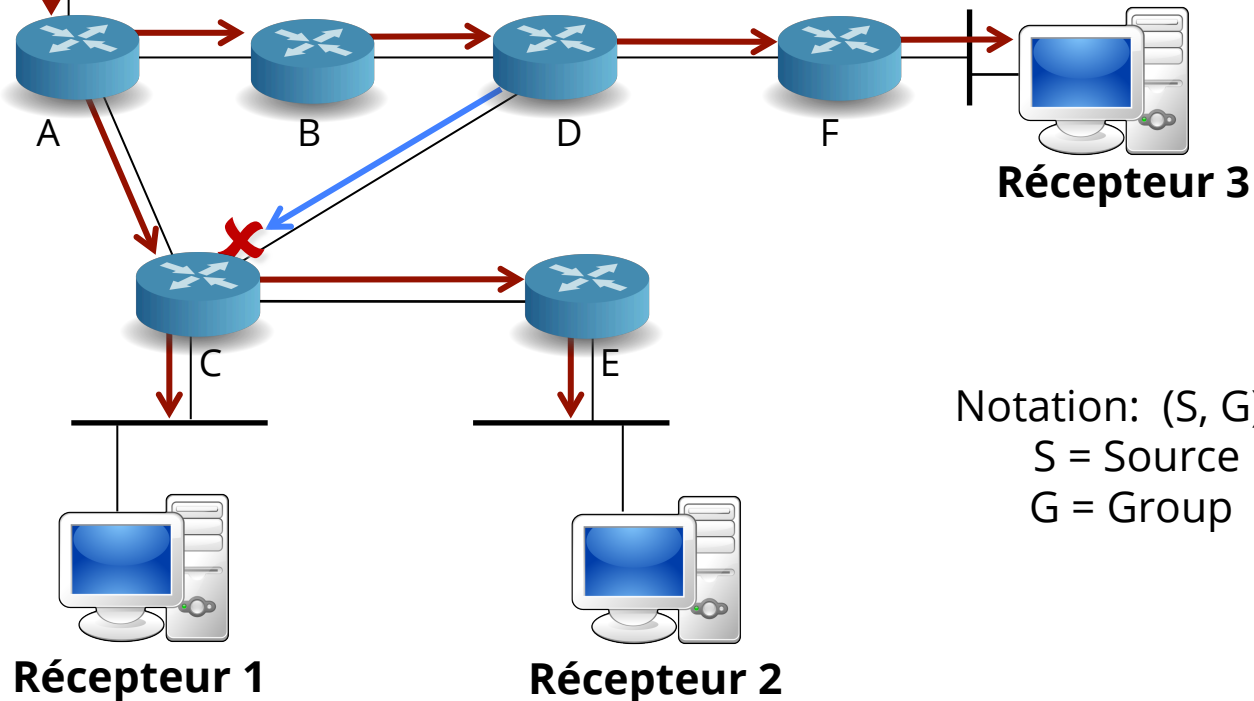
Reverse Path Forwarding (RPF)

Source (S)



RPF est un algorithme permettant de construire un *source specific tree* en contrôlant que les routeurs reçoivent le flux par l'interface qu'ils utiliseraient pour envoyer un paquet vers la source

Ci-dessous le routeur C effectue l'algorithme RPF sur toutes les interfaces sur lesquelles il reçoit un flux multicast. Le chemin provenant du routeur D va être éliminé, car le chemin optimal pour atteindre la source (S) de ce flux passe par le routeur A.



Notation: (S, G)
S = Source
G = Group

Les protocoles de routage multicast (1)

On distingue deux types de protocoles en fonction du mode de transmission des paquets multicast utilisés :

- **Mode dense** (inondation et élagage, « *flood and prune* »)

Pour construire l'arbre de diffusion entre une source et ses destinations, la source diffuse à destination de tous les routeurs voisins son trafic multicast. Ces routeurs envoient eux-mêmes ce trafic à destination de leurs voisins, à l'exception de l'interface par laquelle ils ont reçu les paquets multicast. C'est la phase d'inondation ("**flood**").

Cette méthode permet d'atteindre rapidement tous les points du réseau (domaine multicast). Mais les routeurs ne sont pas nécessairement intéressés à recevoir ce trafic multicast.

Lorsqu'un routeur ne souhaite pas recevoir ce trafic, et qu'il a la certitude qu'aucun de ses voisins avals directs n'est intéressé, il envoie un message « **prune** » (élagage) vers la source. Ce message permet au voisin amont d'arrêter l'émission de ce trafic.

Le **Mode dense** est préférable lorsque les abonnés aux groupes multicast sont nombreux. Il est ainsi mal adapté aux grands réseaux.

DVMRP, PIM-DM et MOSPF sont des exemples de protocoles utilisant le *dense mode*.

DVMRP: *Distance Vector Multicast Routing Protocol*

PIM-DM: *Protocol Independent Multicast – Dense Mode*

MOSPF: *Multicast Open Shortest Path First*

Les protocoles de routage multicast (2)

- **Mode épars** (mode « clairsemé », greffe et élagage, « *sparse mode* »)
Alors qu'en mode dense on parle plutôt d'un système de diffusion (*push*), en mode épars on se trouve dans un système de requête (*pull*).

Ce mode se base sur l'envoi de messages explicites demandant de rejoindre tel ou tel groupe. Ainsi, les paquets de données ne sont transmis qu'aux LANs comportant des hôtes voulant rejoindre le groupe, seules les branches contenant des abonnés reçoivent les paquets.

Le **mode épars** est utilisé avec une **faible population** d'abonnées

PIM-SM et CBT sont des exemples de protocoles utilisant le *dense mode*

PIM-SM: *Protocol Independent Multicast – Sparse Mode*
CBT: *Core Based Trees*

PIM : généralités

PIM (*Protocol Independent Multicast*) est un protocole de routage multicast développé par le groupe de travail IDMR (*Inter Domain Multicast Routing*) de l'IETF et pouvant fonctionner aussi bien en *dense mode* qu'en *sparse mode*. PIM utilise le protocole de routage unicast déjà implémenté dans le domaine (d'où son nom), alors que DVMRP à son propre protocole de routage.

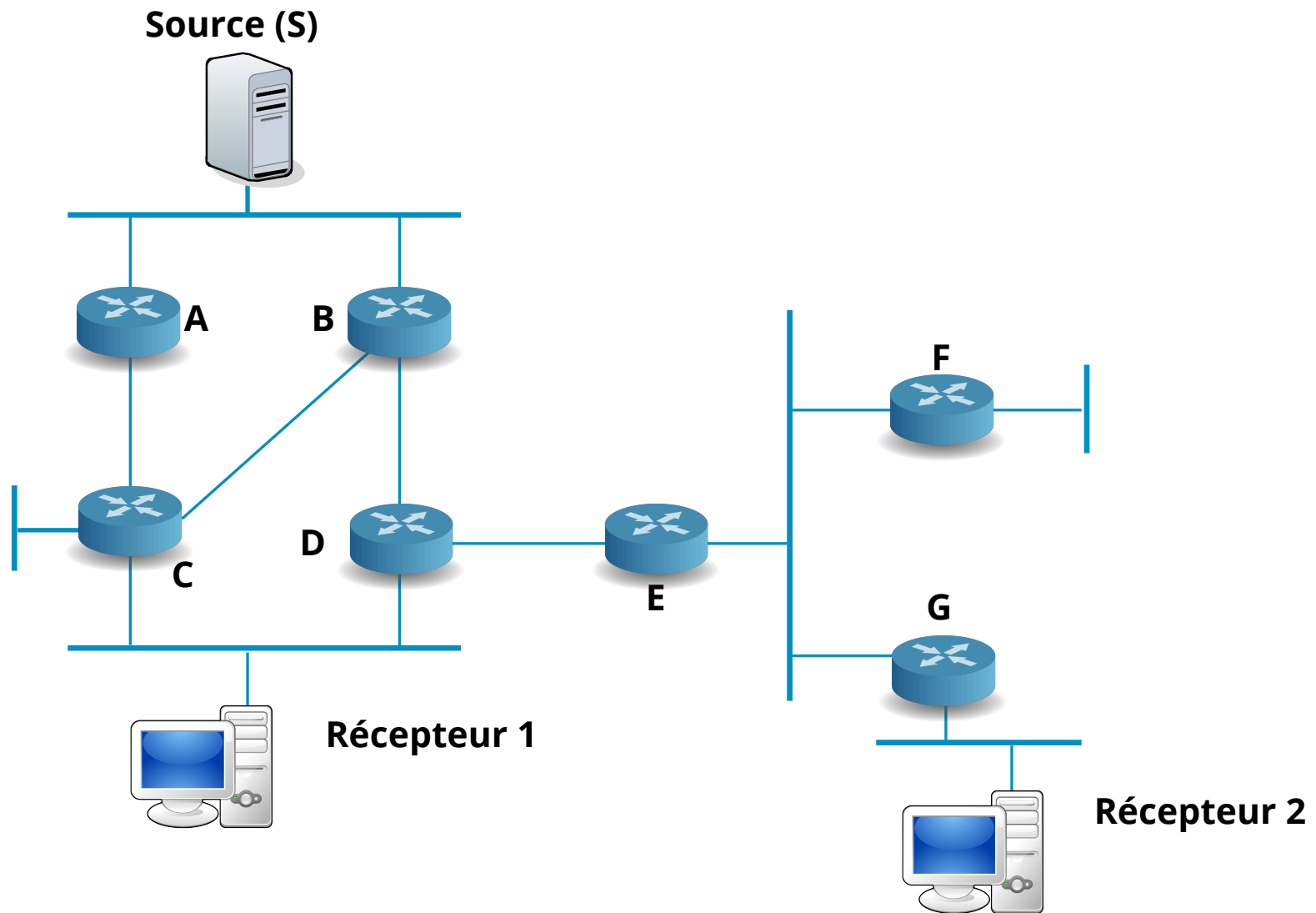
- **PIM Dense Mode** utilise un procédé d'inondation et ne s'adapte pas à un grand réseau.
- **PIM Sparse Mode** est le protocole de routage multicast le plus utilisé actuellement.

Avec IGMPv3 et peu de sources il est avantageux de créer directement un arbre dédié à la source. C'est le **PIM Source Specific Multicast** (PIM-SSM) RFC 3569.

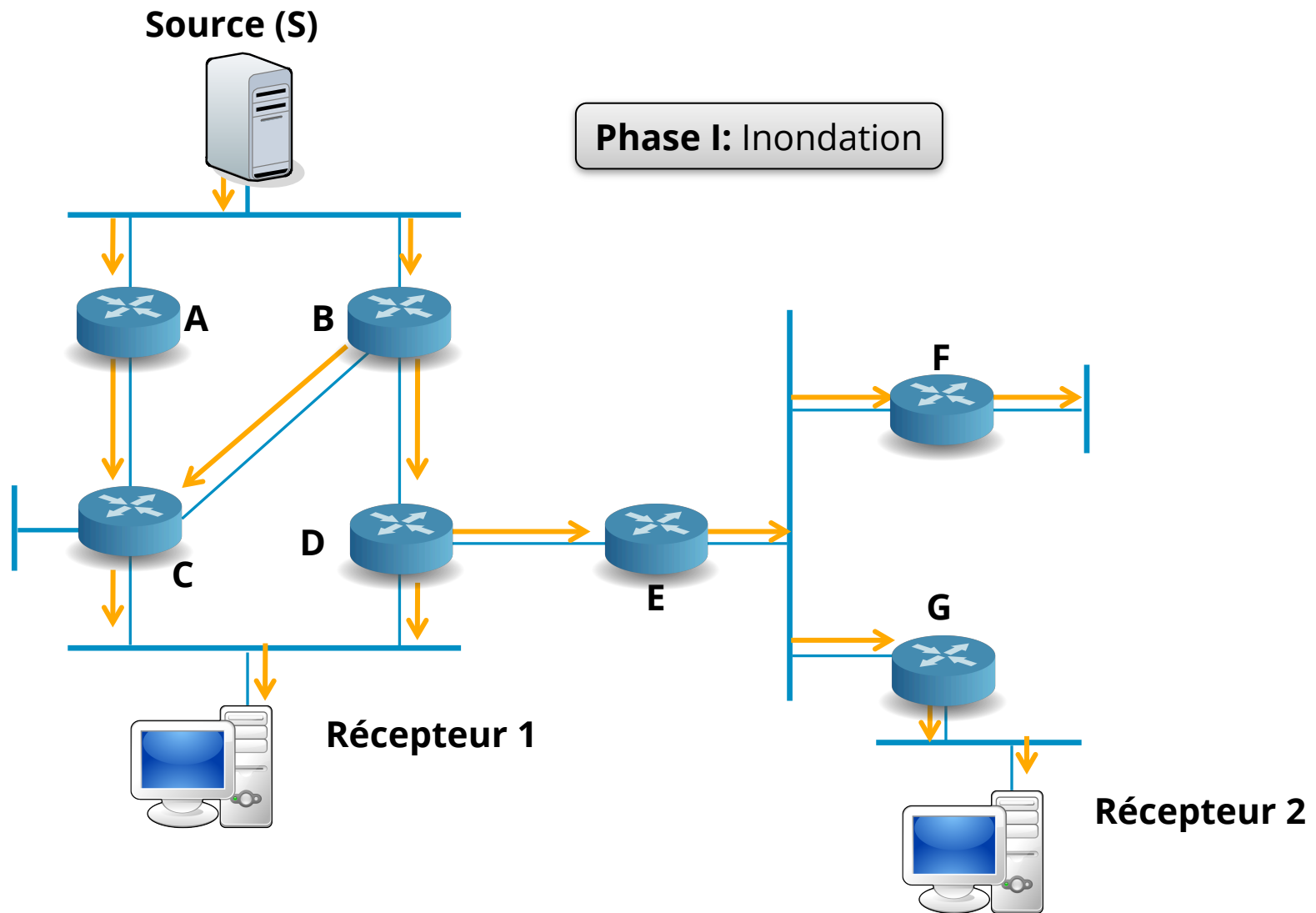
PIM : Dense Mode

- RFC 3973 (janvier 2005)
- Ressemble à DVMRP, mais utilise s'appuie sur le routage unicast en place.
- Mécanismes d'inondation (flooding), d'élagage (pruning) et de greffe (graft).
- Arbres construits par rapport aux sources émettrices avec utilisation de l'algorithme RPF pour effectuer l' élagage vers ses voisins
- Utilisation de déclaration (assert) pour élire un routeur (router désigné, RD) sur un LAN à plusieurs routeurs

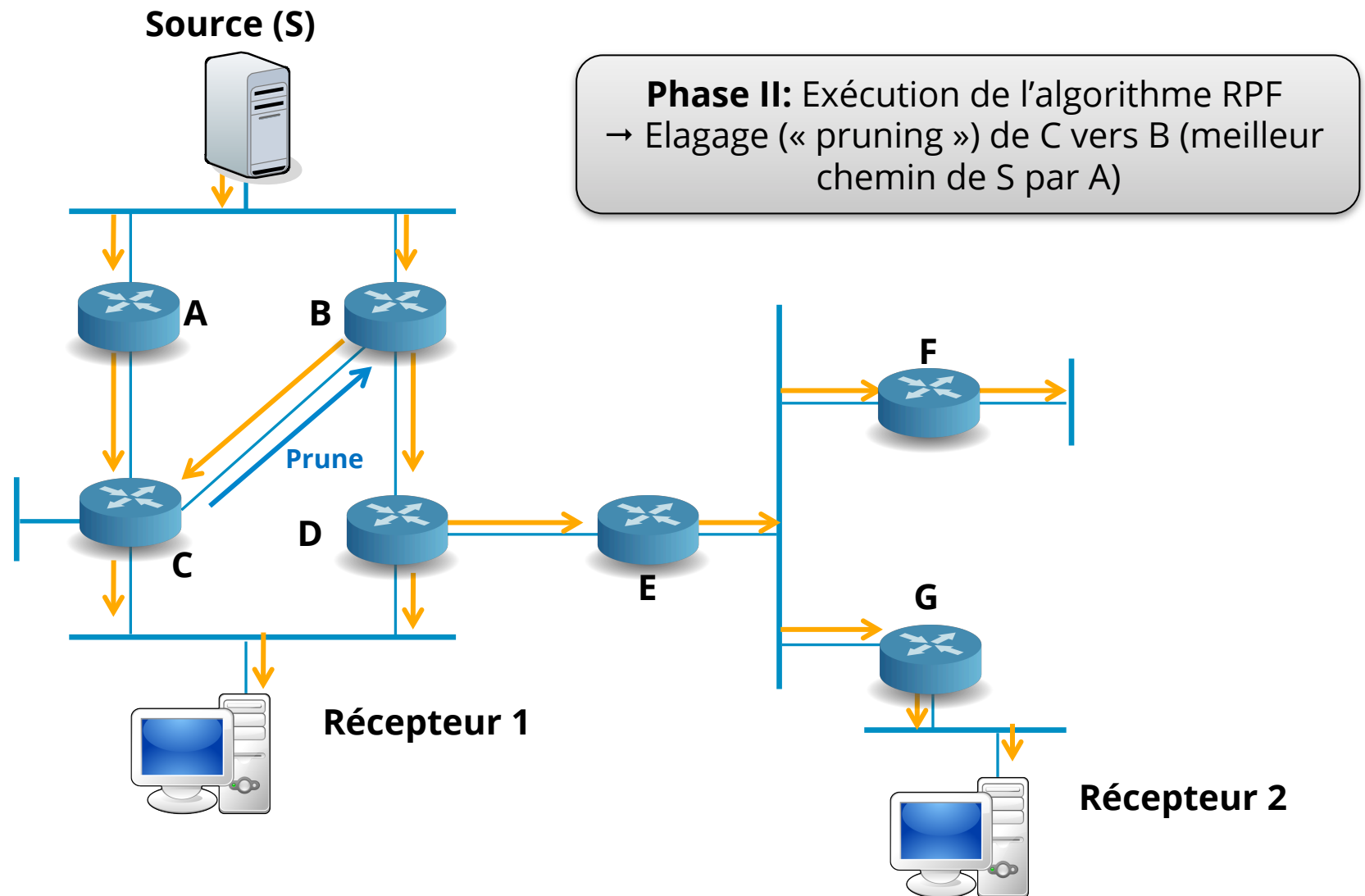
PIM Dense-Mode : exemple (1)



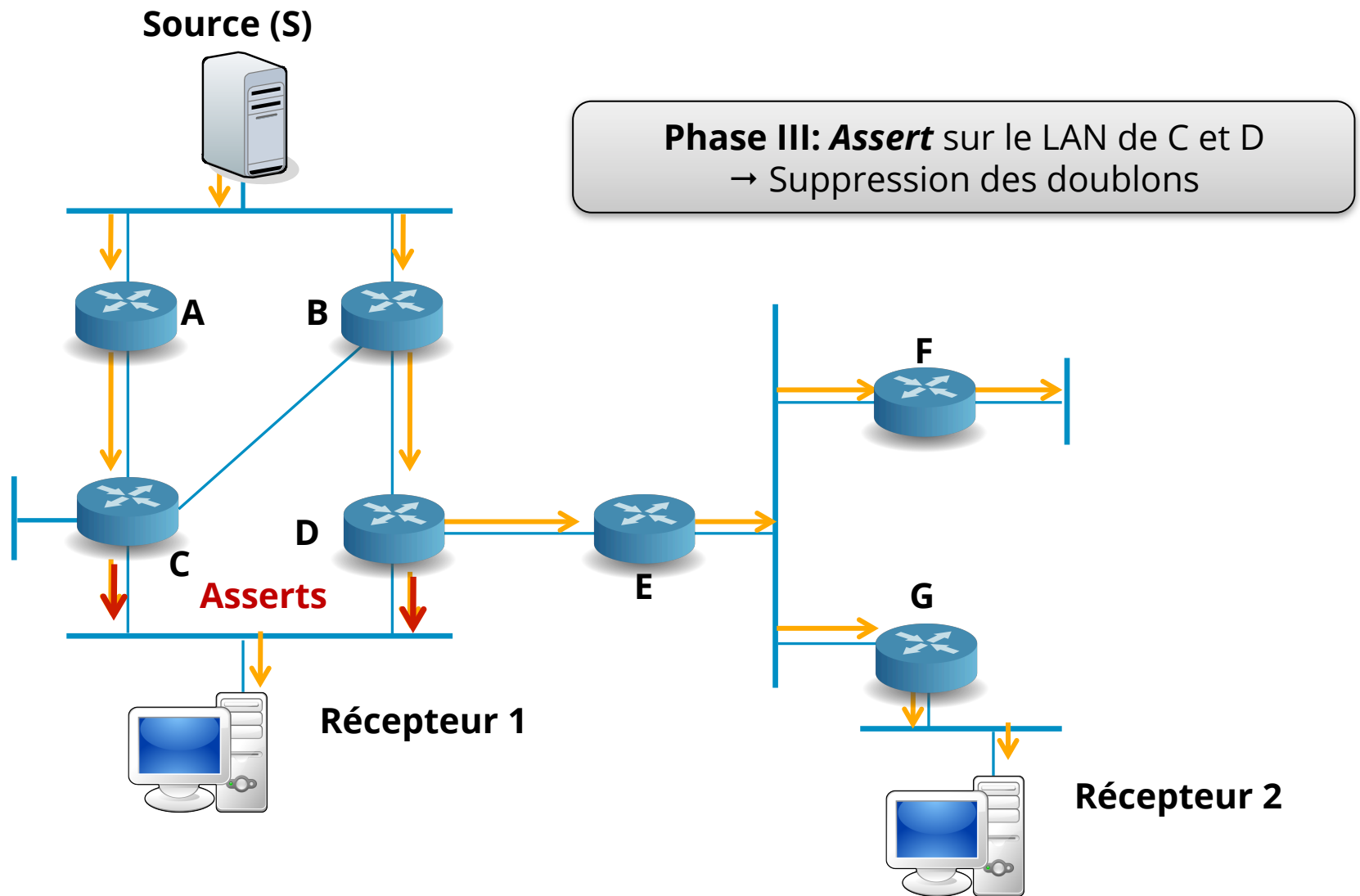
PIM Dense-Mode : exemple (2)



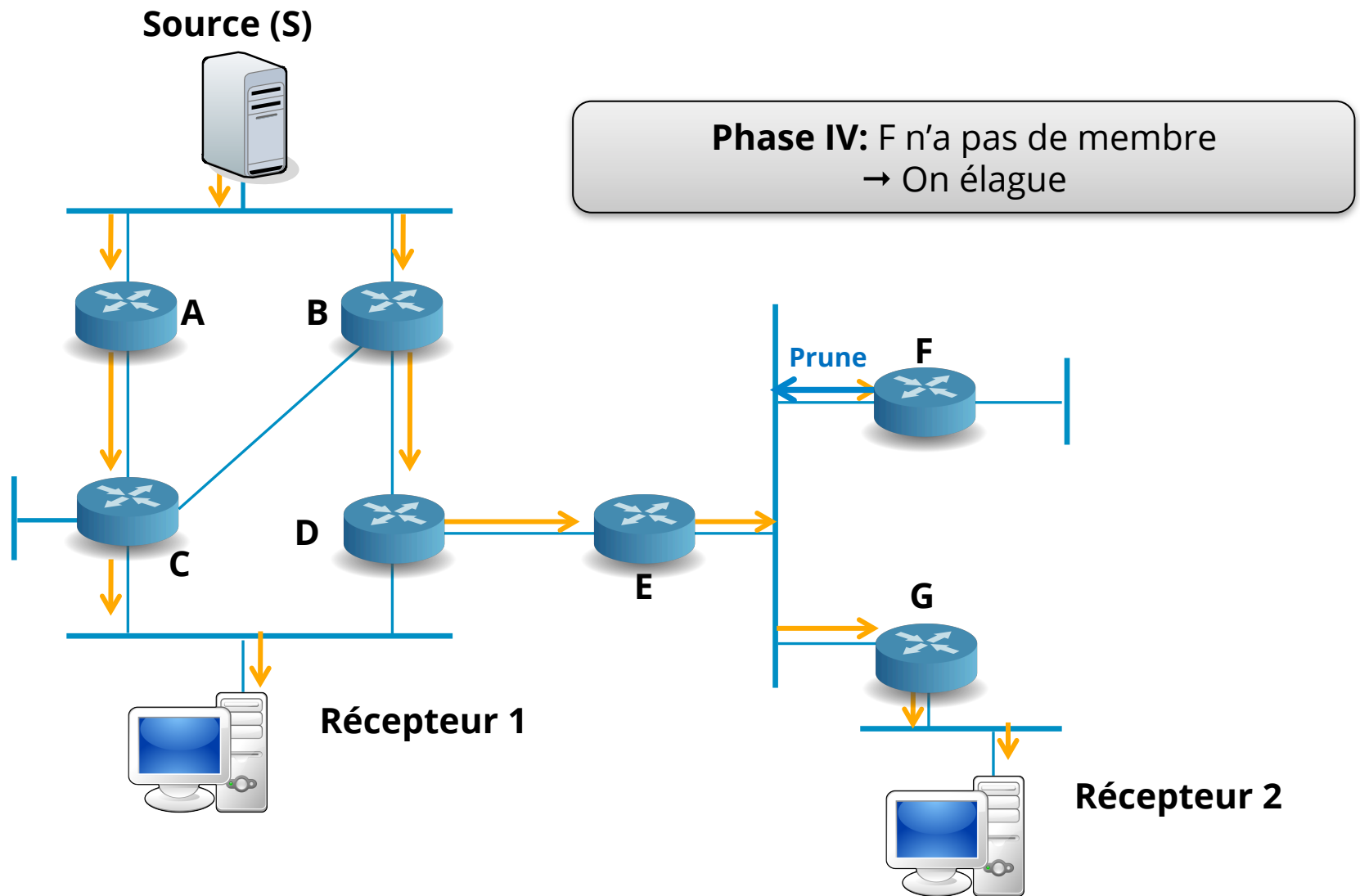
PIM Dense-Mode : exemple (3)



PIM Dense-Mode : exemple (4)



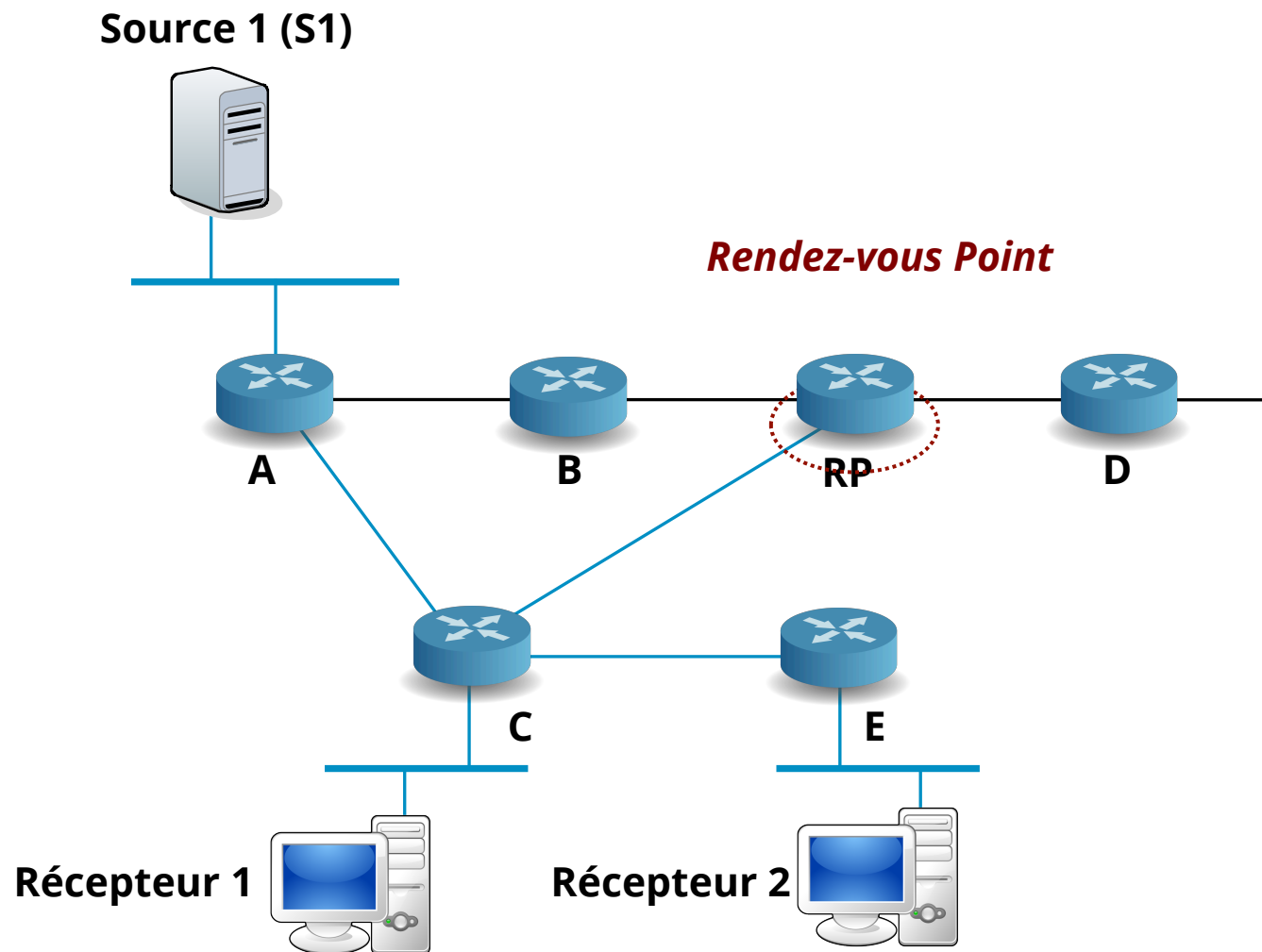
PIM Dense-Mode : exemple (5)



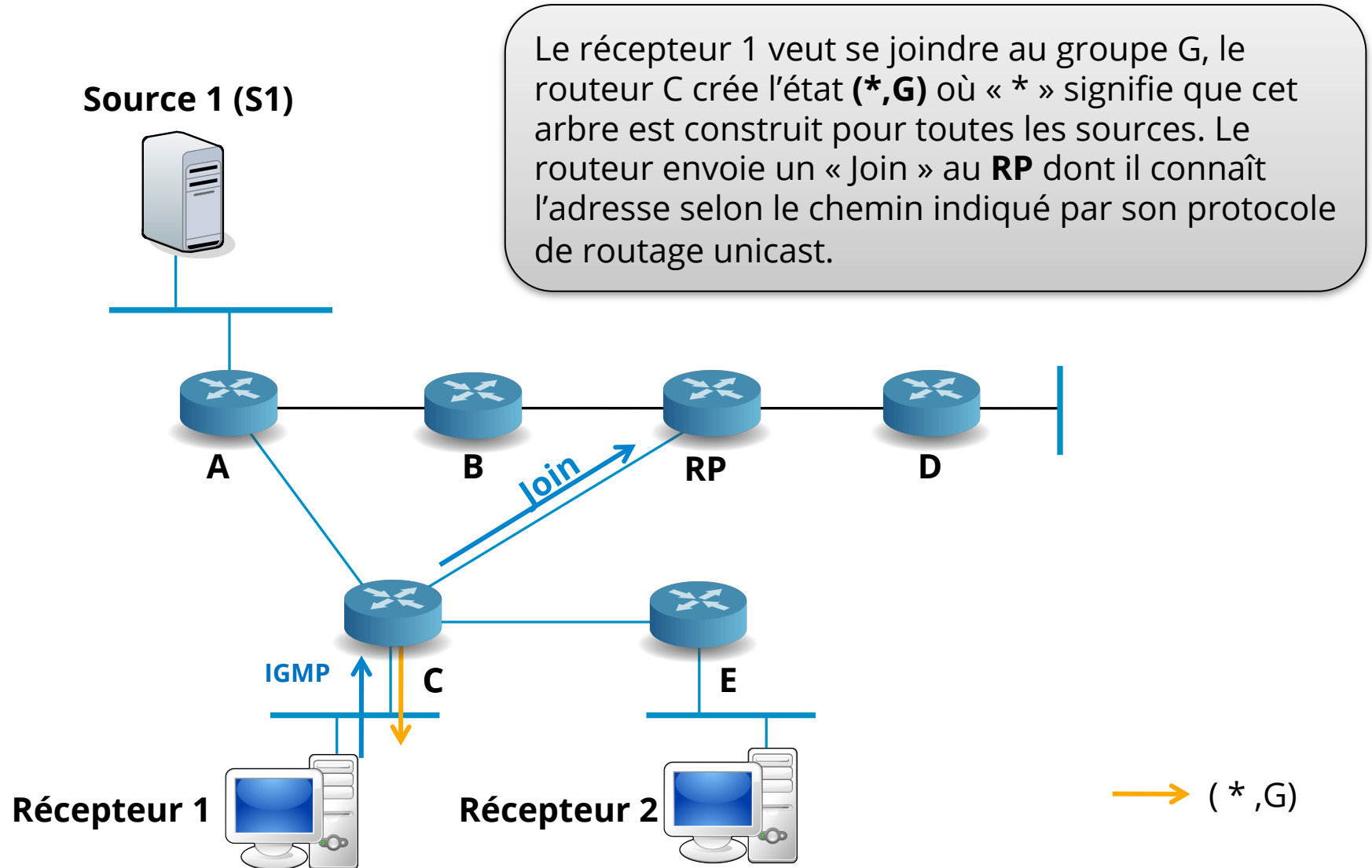
PIM : Sparse mode

- RFC 4601 (Août 2006)
- Mode de souscription explicite (*Join*) des routeurs pour recevoir un flux multicast:
 - la source s'enregistre auprès d'un Point de Rendez-vous (RP, *Rendez-vous Point*) qui est le routeur à la racine de l'arbre de distribution
 - Le RP est soit configuré statiquement soit élu sur la base d'annonces et de sélection de l'adresse IP la plus haute
 - Tous les routeurs du domaine connaissent le RP du groupe multicast.
 - Les routeurs envoient directement des messages « *Join* » et « *Prune* » au RP.
- Le flux multicast parcourt un arbre partagé
 - les routeurs feuilles peuvent se joindre à l'arbre
 - les flux ne vont que là où c'est utile
- Dès que le débit du flux multicast dépasse une certaine valeur, le routeur le plus proche de la destination (pour une destination donnée) essaye de rejoindre un arbre de distribution dédié (*source specific tree*) en envoyant un « *Join* » en direction de la source.

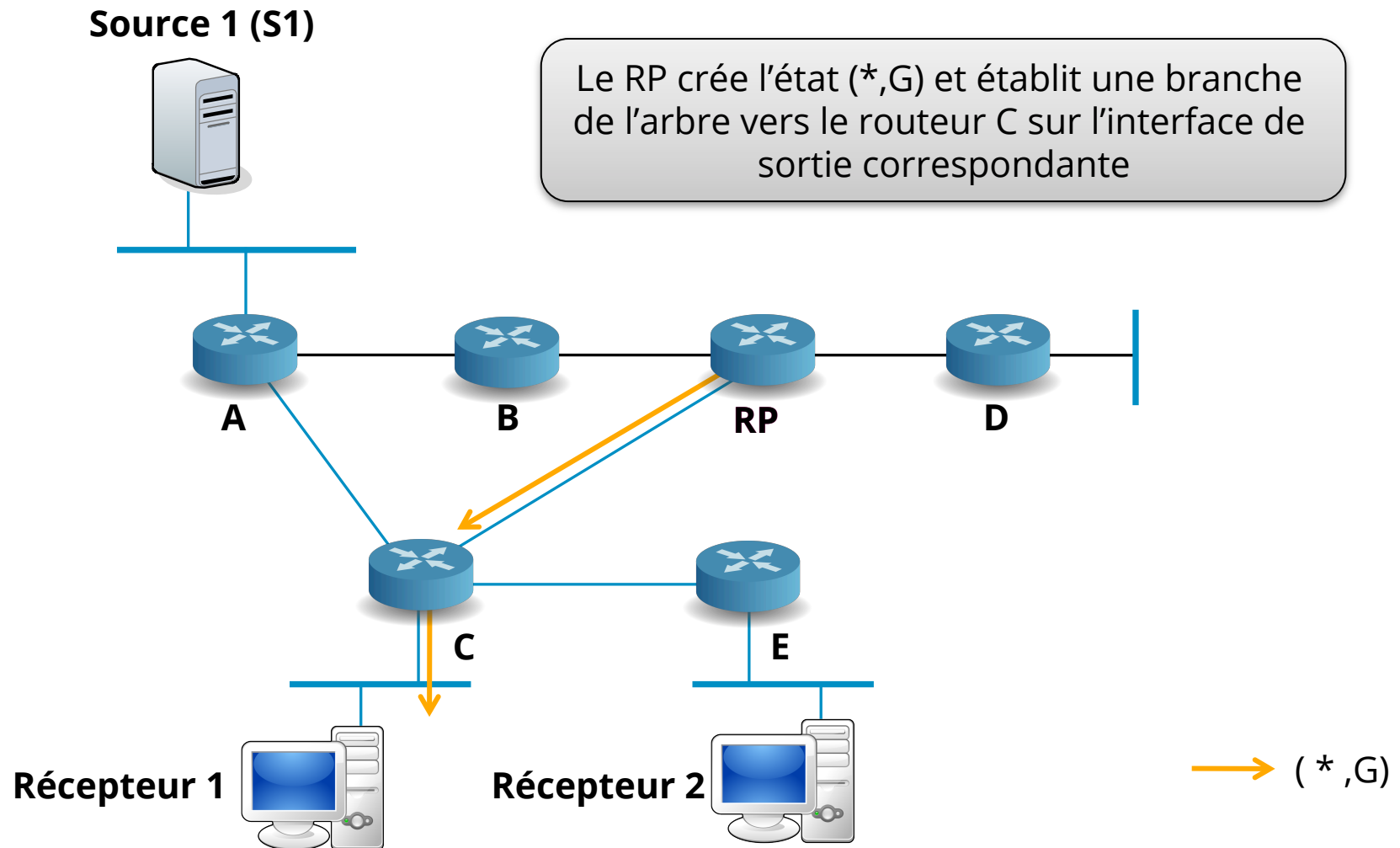
PIM Sparse Mode: Exemple (1)



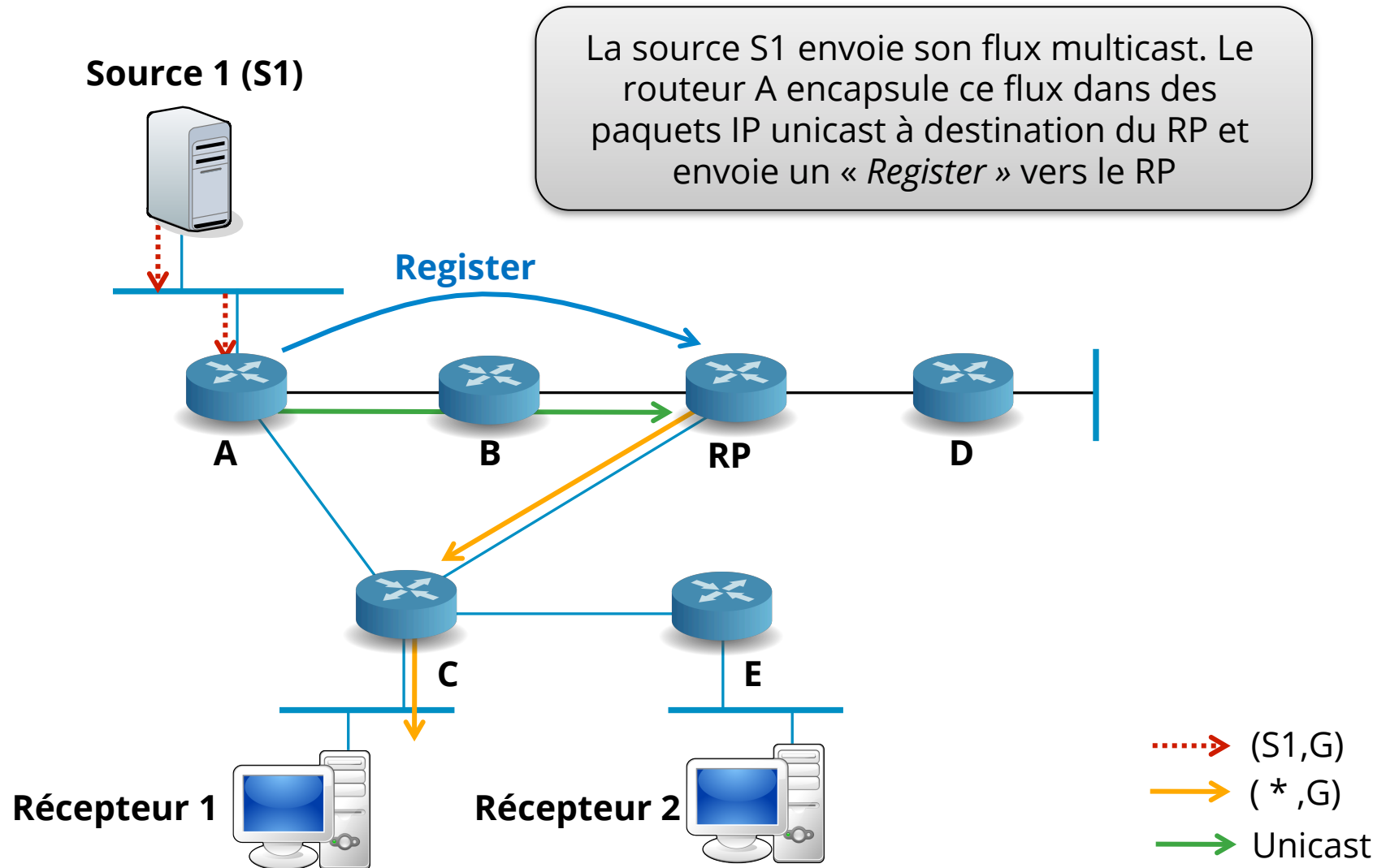
PIM Sparse Mode: Exemple (2)



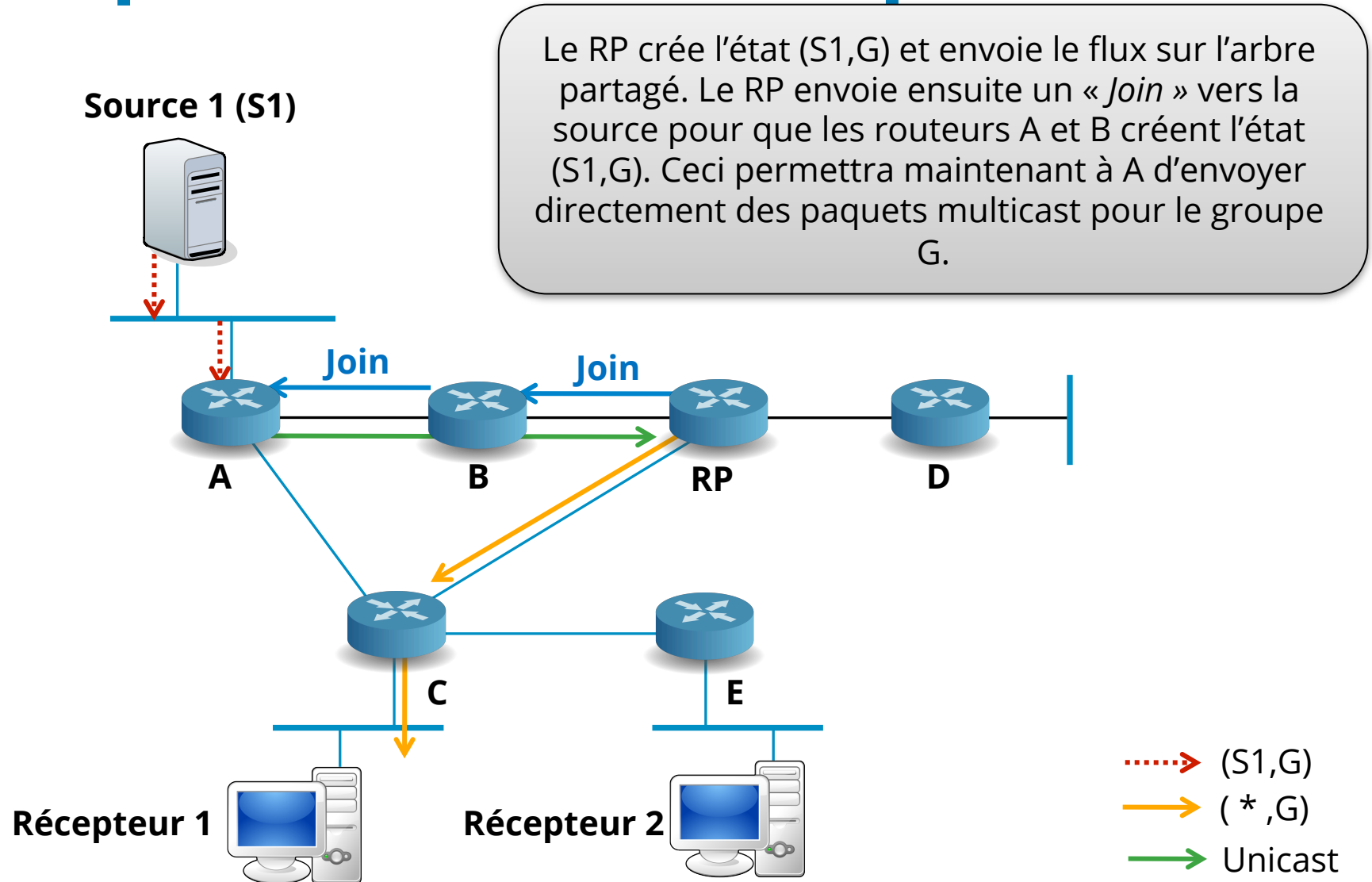
PIM Sparse Mode: Exemple (3)



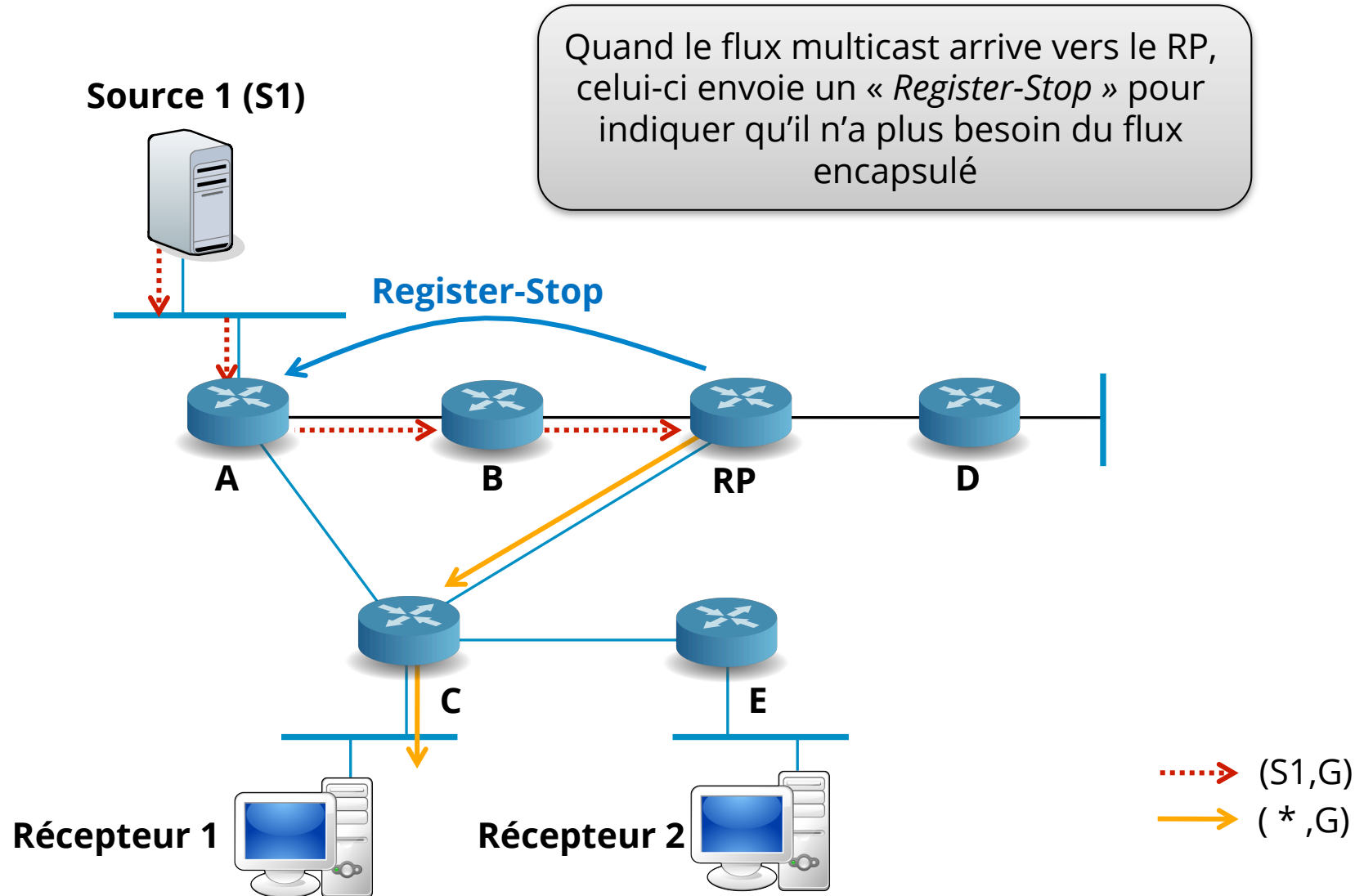
PIM Sparse Mode: Exemple (4)



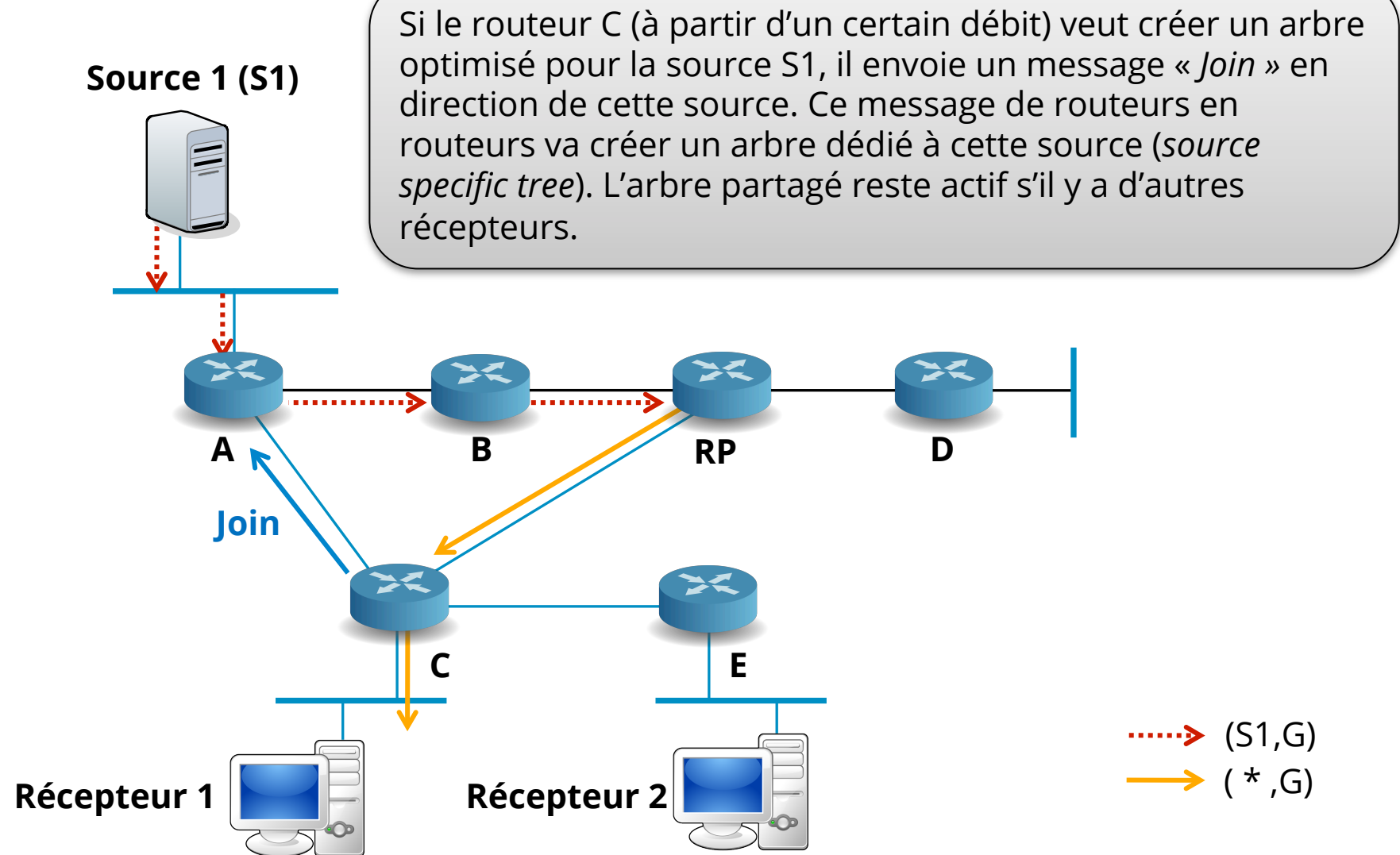
PIM Sparse Mode: Exemple (5)



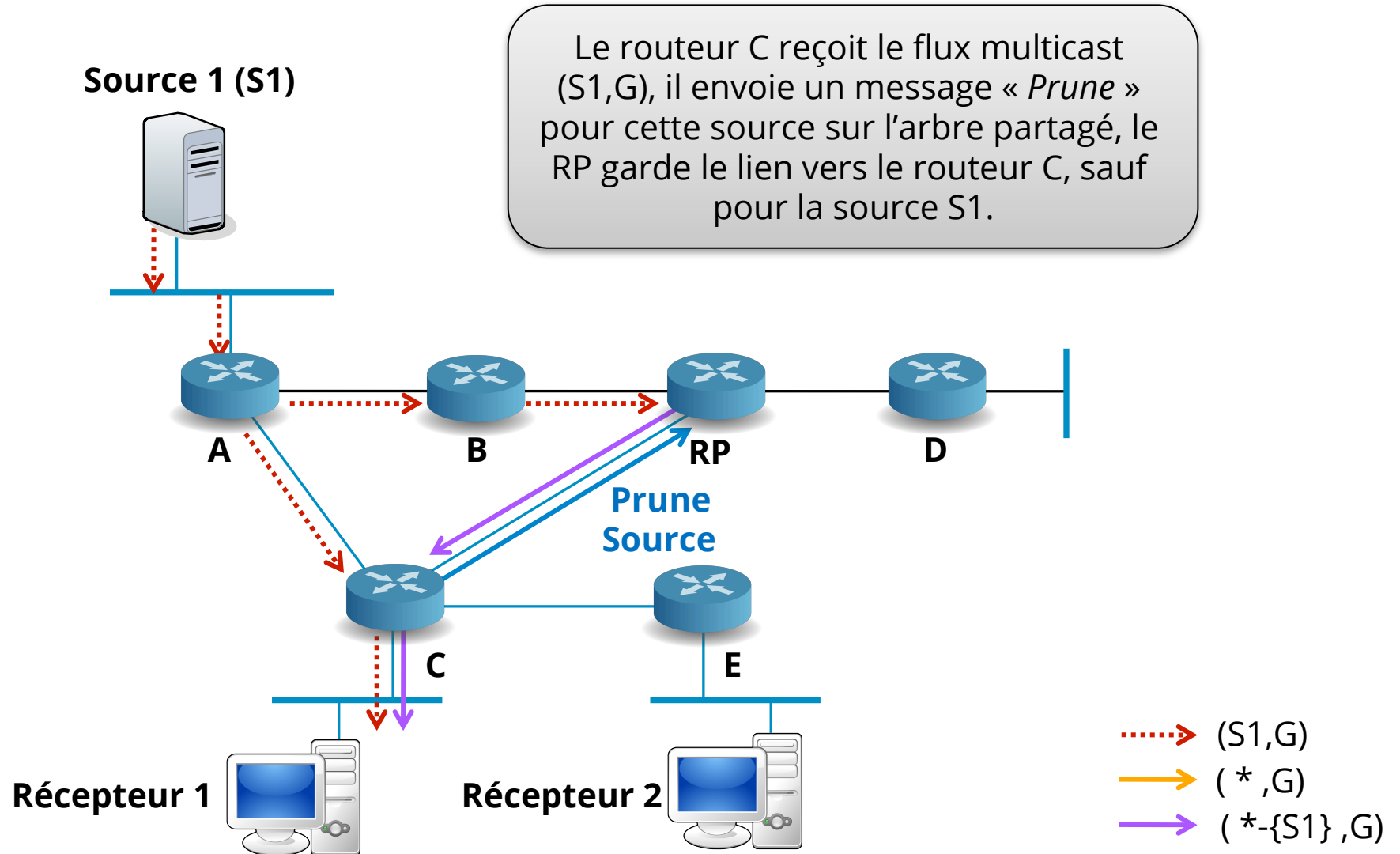
PIM Sparse Mode: Exemple (6)



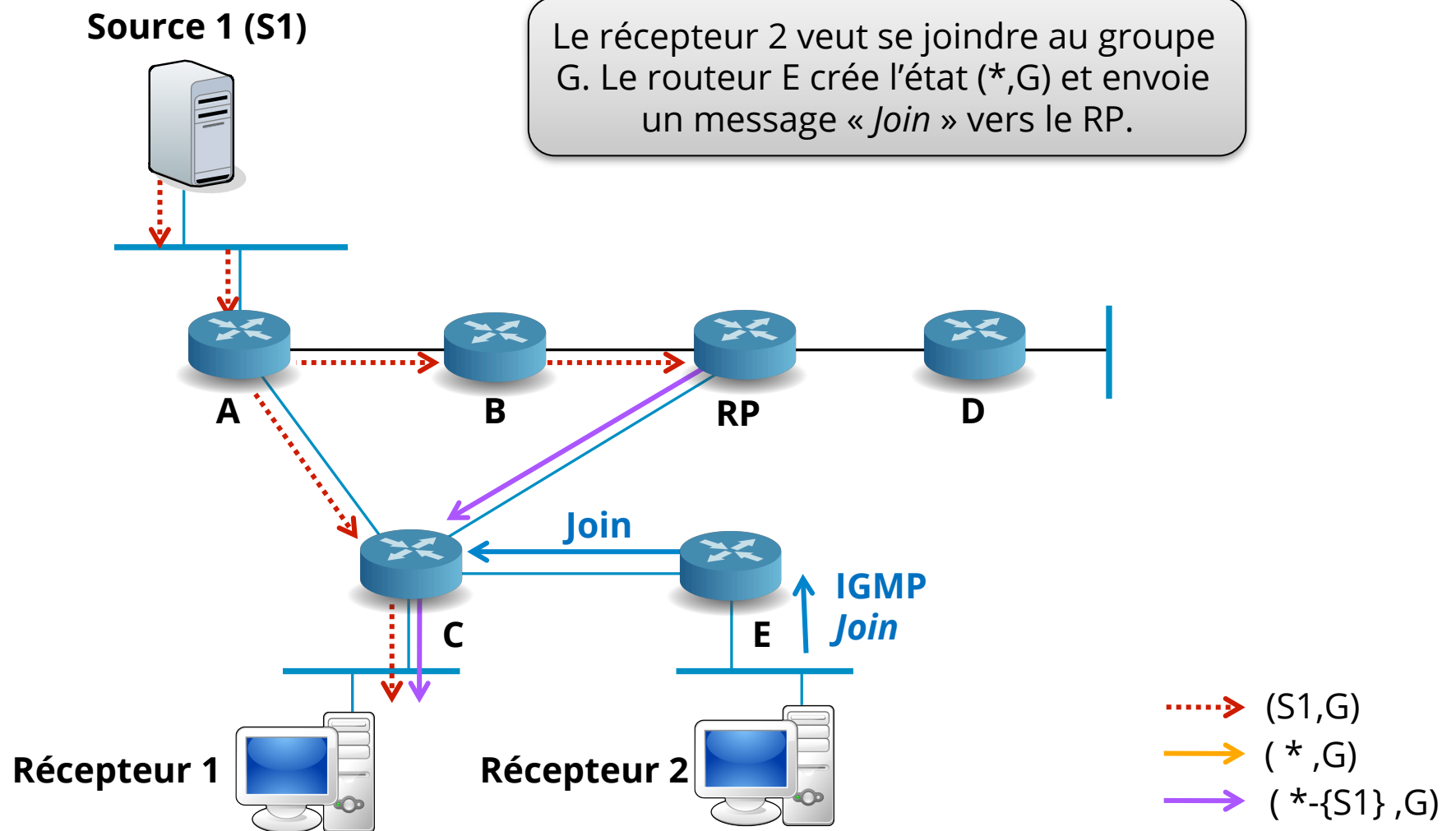
PIM Sparse Mode: Example (7)



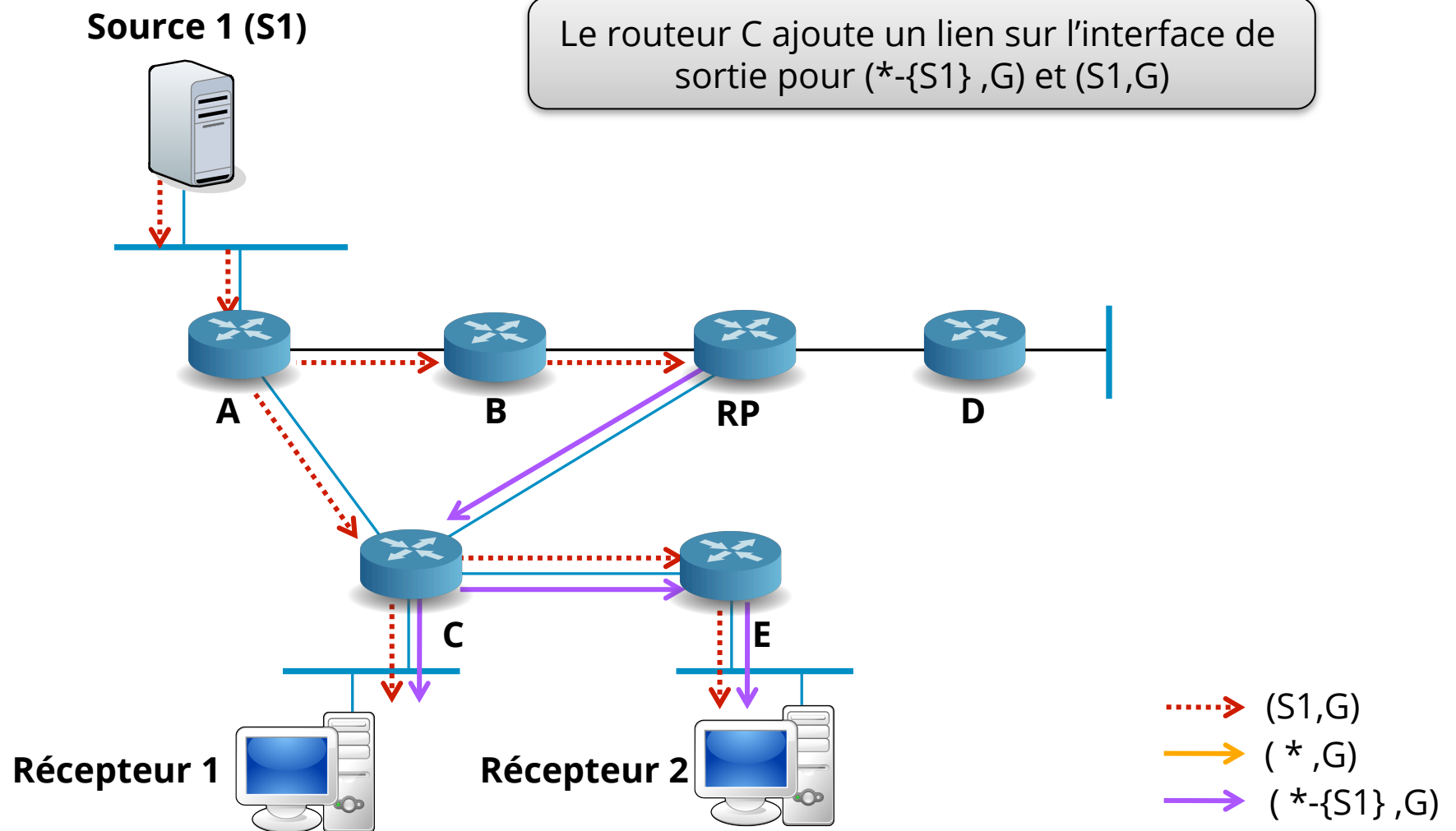
PIM Sparse Mode: Exemple (8)



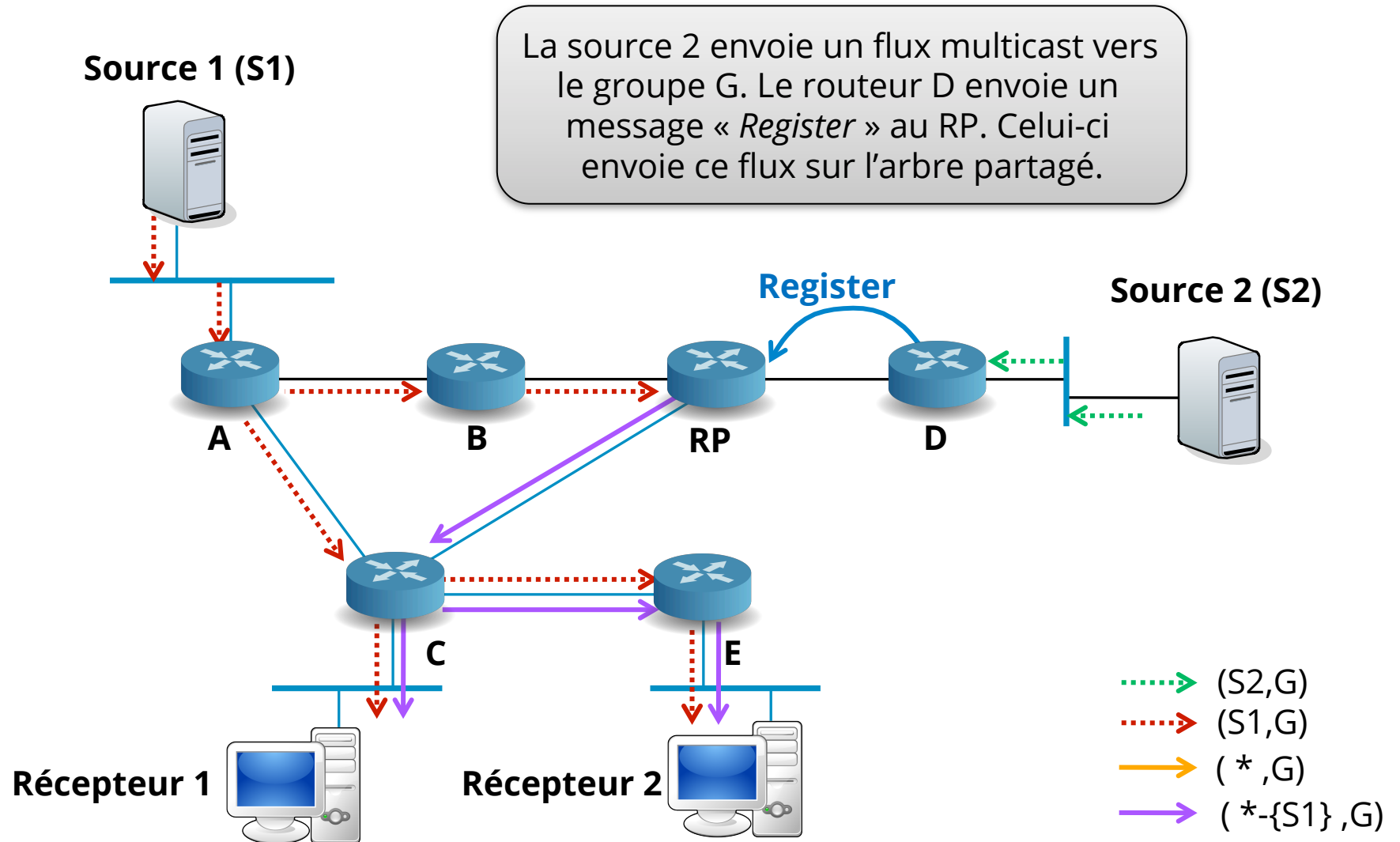
PIM Sparse Mode: Exemple (9)



PIM Sparse Mode: Exemple (10)

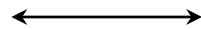


PIM Sparse Mode: Exemple (11)



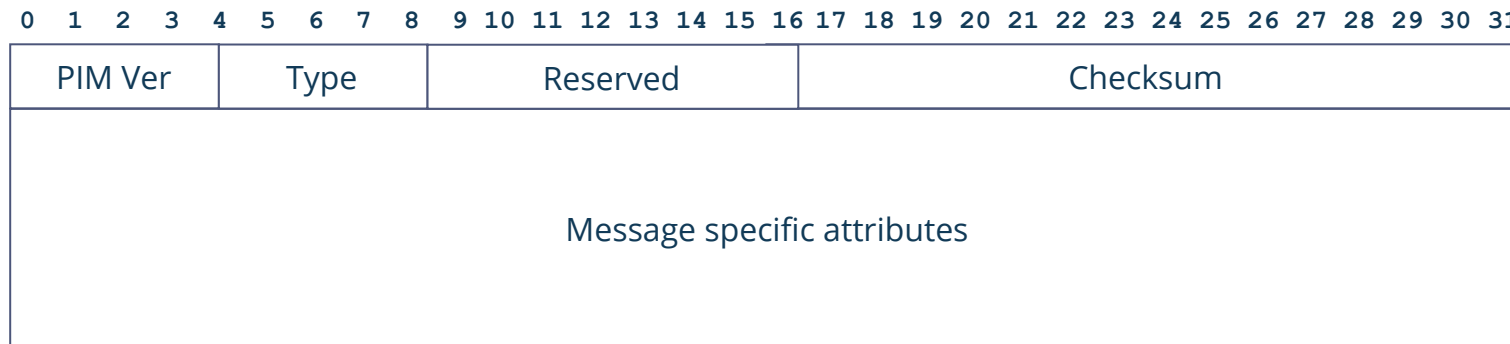
Format des datagrammes PIM

≥20 octets



PIM est identifié par le numéro de protocole **103** dans le paquet IP.
Les messages PIM sont soit des unicasts ou des multicasts (avec TTL=1)

En-tête commun PIM



Champs du datagramme PIM (1)

- **PIM Ver** : La version actuelle est la 2
- **Type** : Type des messages PIM, les différents messages sont:

Message Type	Destination

0 = Hello	Multicast to ALL-PIM-ROUTERS
1 = Register	Unicast to RP
2 = Register-Stop	Unicast to source of Register packet
3 = Join/Prune	Multicast to ALL-PIM-ROUTERS
4 = Bootstrap	Multicast to ALL-PIM-ROUTERS
5 = Assert	Multicast to ALL-PIM-ROUTERS
6 = Graft (used in PIM-DM only)	Unicast to RPF'(S)
7 = Graft-Ack (used in PIM-DM only)	Unicast to source of Graft packet
8 = Candidate-RP-Advertisement	Unicast to Domain's BSR

- **Checksum** : Complément à 1 de la somme des mots de 16 bits de l'ensemble du payload PIM.

Pour le format des différents types de messages, se référer à la RFC 4601

Exercices

Multicast IP

- 400.48 à 400.50