à l'Université Pierre et Marie Curie, le 5 octobre 2004

M2 Informatique Réseaux

Multimédia et Qualité de Service

Cours 2 : Le multicast

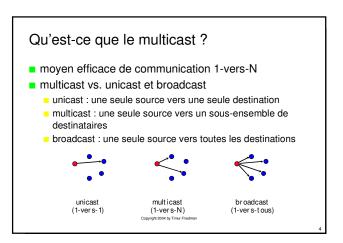
Timur FRIEDMAN

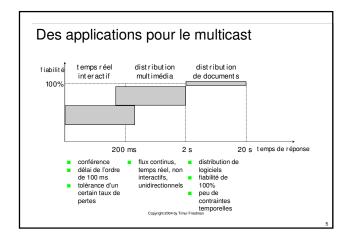
à partir des transparents de Kim THAI, avec modifications

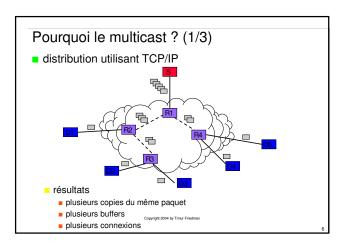
Plan Introduction Le multicast au niveau réseau Le multicast au niveau transport Perspectives de recherche

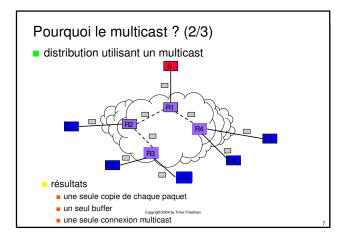
Plan

Introduction
Définition
Notion de groupe
Problématique
Le multicast au niveau réseau
Le multicast au niveau transport
Perspectives de recherche





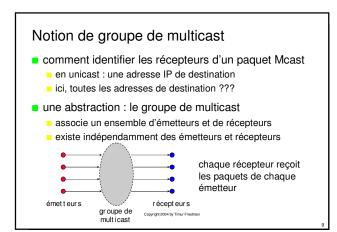




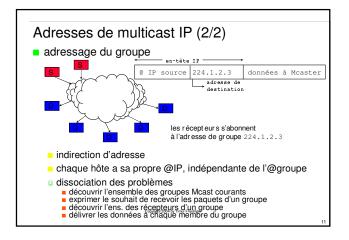
Pourquoi le multicast ? (3/3)

- utilise la bande passante de façon efficace
- prévient la congestion du réseau
- minimise la charge des serveurs
- fournit l'information à davantage d'utilisateurs simultanément
- touche un nombre quelconque de personnes en une seule fois
- etc.

Copyright 2004 by Timur Friedman



Adresses de multicast IP (1/2) un groupe de multicast : une adresse de classe D A 0 réseau station B 10 réseau station réseau C 110 station D 1110 adresse multicast 11100000 00000000 00000000 00000000 de 224.0.0.0 à 239.255.255.255 11101111 11111111 11111111 11111111 224.0.0.0 : non utilisée 224.0.0.1 : représente l'ens. des stations du sous-réseau considéré il n'y a pas d'adresse pour l'ens. des machines de l'Internet



Multicast : les problèmes des questions... quand et comment un groupe naît-il et prend-il fin ? quand et comment l'@ groupe est-elle choisie ? comment de nouvelles stations se joignent-elles à un groupe ? y-a-t-il des conditions pour l'appartenance à un groupe ? comment les routeurs interopèrent-ils pour délivrer les paquets ? des choix... un récepteur doit pouvoir joindre ou quitter un groupe en cours de transmission un récepteur doit pouvoir joindre ou quitter un groupe sans le signaler explicitement aux émetteurs des constats.... les hôtes récepteurs sont souvent connectés à des réseaux locaux...

Plan

- Introduction
- Le multicast au niveau réseau
 - Le multicast sur un LAN
 - Le protocole IGMP
 - Le modèle de service
 - Les algorithmes de routage multicast
 - Les protocoles de routage multicast
- Le multicast au niveau transport
- Perspectives de recherche

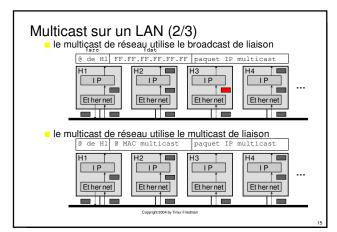
Copyright 2004 by Timur Friedma

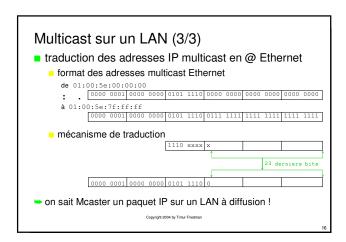
Multicast sur un LAN (1/3)

- l'existant (cas d'Ethernet)
 - Ethernet repose sur un support à diffusion
 - chaque station a une carte réseau avec une @ matérielle spécifique
 - il existe une adresse de diffusion (FF:FF:FF:FF:FF)
- que faire si l'on souhaite joindre uniquement un sousensemble de stations ?
 - <u>ex</u>: H1 souhaite envoyer un paquet Mcast à H2 et H4 qui sont sur le même réseau que lui
 - deux possibilités...

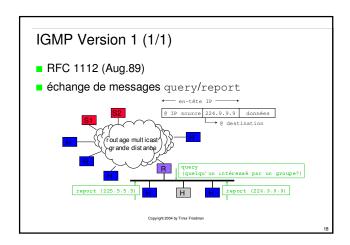
Copyright 2004 by Timur Friedman

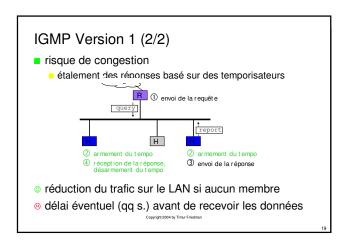
14

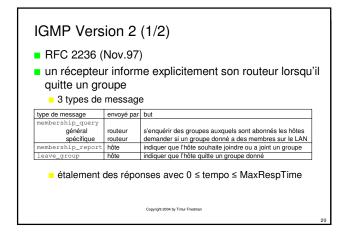


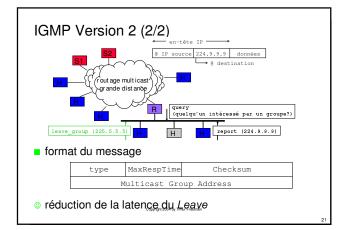


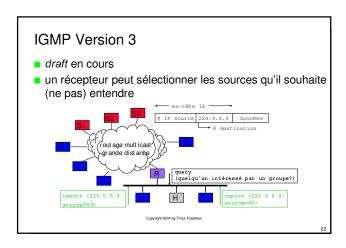
IGMP: Qu'est-ce que c'est? comment un routeur détermine-t-il si son LAN possède des récepteurs pour un groupe donné? Internet Group Management Protocol permet à un hôte d'indiquer à son routeur local qu'il souhaite joindre un groupe est utilisé sur les LAN à diffusion











Le modèle de service du multicast (1/5) issu des travaux de Steve Deering caractéristiques de la transmission multicast IP: transmission d'un paquet IP à un groupe d'hôtes identifié par une seule adresse de destination transmission best effort caractéristiques du groupe appartenance dynamique pas de restriction quant à la localisation et au # de membres un hôte peut être simultanément membre de plusieurs groupes un hôte n'a pas besoin de faire partie d'un groupe pour être source groupe permanent/transitoire l'opération de Join est receiver-driven

Le modèle de service de multicast (2/5) l'émetteur ne contrôle pas qui joint le groupe il n'y a pas de contrôle sur qui envoie au groupe les paquets issus de plusieurs sources peuvent être reçus entrelacés les paquets issus de plusieurs sources peuvent être reçus entrelacés les groupes différents peuvent choisir la même rôle des routeurs Mcast locaux co-résidents ou séparés des routeurs classiques un routeur local qui reçoit un paquet Mcast d'un de ses hôtes, avec un TTL-1, le fait suivre vers tous les sous-réseaux connectant des membres récepteurs sur les sous-réseaux destinataires, le routeur local termine la transmission en Mcastant le paquet en local

Le modèle de service de multicast (3/5) • le RFC 1112 spécifie les extensions à apporter à un hôte IP pour supporter le Mcast • 3 niveaux de conformité • 0 : l'hôte ne supporte pas le Mcast • 1 : l'hôte peut émettre à destination d'un groupe • 2 : l'hôte supporte le Mcast en émission et réception • modèle d'implémentation IP d'un hôte | modules de protocoles de niveau sup. | |

I CMP I GMP

traduction d'8

----- interface de service LAN -----

module LAN

Le modèle de service de multicast (4/5) • les extensions pour l'envoi Mcast • interface de service IP • utilisation de SendIP • @ dest = @ de groupe • le niveau supérieur doit pouvoir spécifier un TTL • module IP • usi IP-dest est sur le même réseau local ou si IP-dest est une ê de groupe alors envoyer le paquet en local à IP-dest sinon envoyer le paquet en local à IP-dest interface de service LAN • module LAN • mécanisme de traduction des @IP Mcast en @MAC Mcast

Copyright 2004 by Timur Friedman

Le modèle de service du multicast (5/5) les extensions pour la réception Mcast interface de service IP utilisation de ReceiveIP ajout de JoinHostGroup (group-address, interface) ajout de LeaveHostGroup (group-address, interface) module IP maintien de la liste des groupes dont l'hôte est membre pour chacune des interfaces (mise à jour avec les Join et Leave) intégration de IGMP et adhésion à 224.0.0.1 interface de service LAN ajout de JoinLocalGroup (group-address) ajout de LeaveLocalGroup (group-address) module LAN mécanismes de filtrage par la carte souhaités

Plan Introduction Le multicast au niveau réseau Le multicast sur un LAN Le protocole IGMP Le modèle de service Les algorithmes de routage multicast Shortest Path Tree Minimum Cost Tree Constrained Tree Les protocoles de routage multicast Le multicast au niveau transport Perspectives de recherche

Les algorithmes de routage multicast

- <u>objectif</u>: calculer un arbre de liens connectant tous les routeurs ayant des hôtes appartenant au groupe
- buts:
 - minimiser la distance entre la source et chaque récepteur
 - minimiser l'utilisation de liens dans le réseau
 - ★ ces buts ne sont pas compatibles

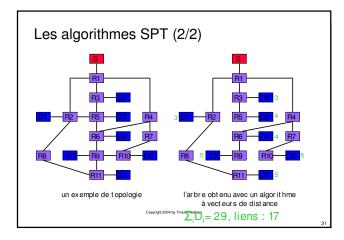
Copyright 2004 by Timur Friedman

Les algorithmes SPT (1/2)

- but : calculer un arbre
 - ayant la source S pour racine
 - couvrant tous les récepteurs D_i du groupe
 - tel que la distance entre S et Di soit minimum
- algorithmes de base
 - Bellmann-Ford : à vecteurs de distance
 - Dijkstra : à états des liens
- o un arbre par émetteur

Copyright 2004 by Timur Friedma

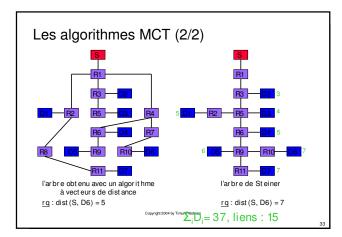
3



Les algorithmes MCT (1/2)

- but : minimiser le coût total de l'arbre
- 2 familles
 - les algorithmes Minimum Spanning Tree
 - contrainte : l'arbre ne doit toucher aucun nœud qui ne soit pas membre du groupe (pas réaliste car les routeurs ne sont pas de membres)
 - ex : algorithme de Prim
 - les algorithmes Minimum Steiner Tree
 - la contrainte est levée
 - o problème NP-complet
 - o ils supposent de connaître toutes les liaisons du réseau
 - ils sont monolithiques
 - o ils n'exploitent pas les informations déjà disponibles de routage unicast

ouright 2004 by Timur Erladman



Les algorithmes CT

- <u>but</u>: minimiser simultanément la dist(S, D_i) et le coût total de l'arbre
- principe
 - associer à chaque lien 2 métriques (distance/délai et coût)
 - rechercher l'arbre à coût minimum tel que dist(S, Di) ≤ Δ

Copyright 2004 by Timur Friedman

34

Qu'appelle-t-on IP multicast?

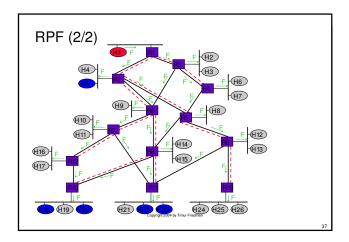
- mécanisme utilisé dans l'Internet pour construire un arbre de routage multicast efficace et sans boucles
- IGMP + protocole de routage Mcast

Copyright 2004 by Timur Friedman

RPF (1/2)

- Reverse Path Forwarding (Source-based Routing)
- l'une des premières techniques utilisées
- <u>but</u>: construire un arbre ayant S comme racine et minimisant dist(S, D_i)
- principe: utiliser l'inondation (flooding) avec
 - si un paquet est reçu par l'if utilisée par le routeur pour joindre S alors le paquet est retransmis sur les autres if sinon le paquet est rejeté
- o mécanisme simple
 - les informations utilisées sont celles du routage unicast
 - R_i n'a pas à connaître les arbres recouvrants
 - pas de mécanisme particulier pour arrêter l'inondation

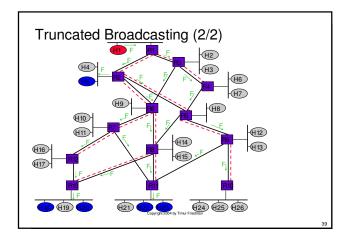
3



Truncated Broadcasting (1/2)

- <u>but</u> : réduire le trafic sur les LAN feuilles
- <u>idée</u>: utiliser les informations d'appartenance fournies par IGMP pour déterminer s'il faut ou non Mcaster un paquet sur un LAN feuille
 - forme d'élagage (pruning) des feuilles
 - pas de réduction de trafic au cœur du réseau

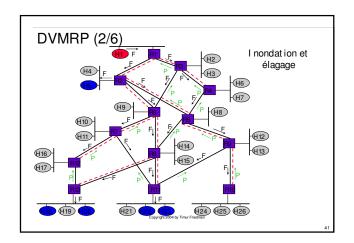
Copyright 2004 by Timur Friedman

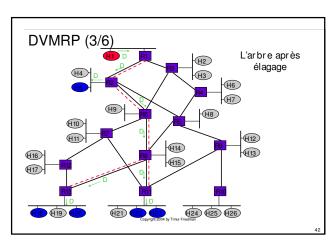


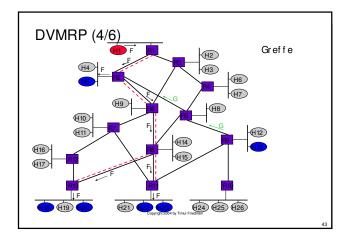
DVMRP (1/6)

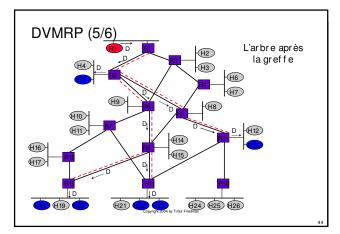
- RFC 1075 (Nov.88), draft Version 3 en cours
- Distance Vector Multicast Routing Protocol
- but : réduire le trafic au cœur du réseau
- <u>principe</u>: inondation et élagage (flooding and pruning)
 - s'il n'a pas de membre sur son LAN, un routeur feuille envoie un message prune à ses voisins
 - un routeur feuille peut envoyer un prune sur toutes ses if, sauf celle correspondant à son SP avec la source (i.e. l'if RPF)
 - quand un routeur intermédiaire reçoit un prune sur chacune de ses if, sauf l'if RPF, il remonte le prune en amont
 - quand un routeur envoie un prune, il mémorise la paire (Source, Groupe) pour laquelle le prune a été envoyé

Copyright 2004 by Timur Friedman









DVMRP (6/6)

- problèmes communs aux protocoles à vecteurs de distance (e.g. temps de convergence)
- processus périodique d'inondation et d'élagage pour chaque source
- mémorisation des enregistrements prune (Source, Groupe)

Copyright 2004 by Timur Friedman

MOSPF

- RFC 1584 (March 94)
- Multicast Open Shortest Path First
- principe
 - opère dans un AS qui utilise OSPF pour l'unicast
 - étend OSPF en ajoutant les informations d'appartenance aux informations d'états des liens qui sont diffusées par OSPF
- o problème : scalability avec la taille du réseau
 - mémorisation d'un enregistrement par groupe et par lien du réseau
 - un arbre par source

Copyright 2004 by Timur Friedman

4

CBT (1/3)

- RFC 2189 et 2201 (Sept.97)
- Core Based Tree (Group-shared Tree)
- but : éviter les inconvénients de DVMRP et MOSPF
 - résistance au facteur d'échelle : un seul arbre pour le groupe
 - efficacité (éviter les inondations) : messages de Join et de Leave explicites
- <u>principe</u>: construire un arbre partagé, bidirectionnel, avec un cœur unique

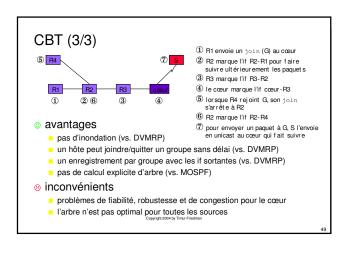
Copyright 2004 by Timur Friedma

CBT (2/3)

- construction de l'arbre
 - un routeur local qui a un nouveau membre pour un groupe envoie un message join-request vers le cœur en unicast
 - le cœur ou le premier routeur sur le chemin faisant déjà partie de l'arbre répond par un join-ack
 - chaque routeur ayant vu passer le join-request marque l'if sur laquelle il l'a reçu
- maintien de l'arbre
 - chaque routeur envoie périodiquement des echo-request à son routeur amont
 - le routeur amont répond par des echo-reply
 - si un routeur aval n'obtient pas de réponse au bout de N essais, il détache son sous-arbre en envoyant un flush-tree

Copyright 2004 by Timur Friedma

4



PIM (1/3)

- RFC 2362 (June 98)
- Protocol Independent Multicast
- idée : distinction explicite de 2 scénarios de distribution
- le mode dense

ou épars

- les membres sont géographiquement concentrés dans une
- idée: RPF avec flood-and-prune, similaire à DVMRP est alors raisonnable

PIM (2/3)

le mode épars

- les membres sont géographiquement éparpillés
- but : un routeur ne doit pas avoir à travailler, à moins de rejoindre un arbre
- principe: approche center-based, similaire à CBT
 - sauf :
 - pas d'acquittement en réponse au join
 - le join est envoyé périodiquement pour «rafraîchir» l'arbre
 - le point de RDV informe une source active d'arrêter d'émettre lorqu'il n'y a plus de routeurs dans l'arbre
 - changement de mode possible : de l'arbre partagé vers l'arbre par
 - les points de RDV émettent périodiquement en aval pour indiquer leur activité

PIM (3/3) R1 sait que son SP avec S passe par son if R1-R4 • or, R1 reçoit les paquets Mcast sur son if R1-R2 • R1 envoie un join à R4 R1 envoie ensuit e un prune au cœur • le cœur arrête le transfert Mcast sur cœur-R2 chemin avec CBT ⊚ le changement de mode per met de décharger le cœur o en cas de panne du cœur, les hôt es ayant commut é de mode continuent de recevoir @ PI M ne dit pas comment un rout eur détermine le point de RDV d'un groupe @ PI M ne dit pas comment déterminer si un groupe est dense

Copyright 2004 by Timur Friedman

Le Mbone (1/3)

problème

- pour mettre en œuvre le Mcast sur l'Internet, il faut que tous les routeurs aient des fonctions de Mcast et que les routeurs locaux supportent IGMP
- a plupart des routeurs de l'Internet ne supportent pas le Mcast !!!

idée

- bâtir des sous-réseaux capables de Mcast à la périphérie de l'Internet
- les interconnecter par des tunnels, les extrémités des tunnels sont des stations avec mrouted et un support de l'OS pour le Mcast

Multicast Backbone of the Internet

- réseau virtuel de recouvrement, solution transitoire
- premier tunnel en 88 entre BBN et Stanford
- des milliers de sous-réseaux aujourd'hui
- utilisé pour diffuser des sessions IETF ou des conf. IEEE/ACM

Le Mbone (2/3) principe : le tunneling encapsulation des paquets Mcast transmis sur le Mbone dans des paquets IP classiques ■ l'extrémité du tunnel réceptrice détecte qu'elle a un paquet IP encapsulé dans un paquet IP (protocol =4) après désencapsulation, elle fait suivre le paquet Mcast soit en local sur son sous-réseau, s'il a des hôtes membres soit au prochain routeur Mcast, après ré-encapsulation en-tête IP unicast le paquet Mcast d'origine encapsulation désencapsulation t unnel M2

Le Mbone (3/3)

- trafic
 - les conférences génèrent typiquement 100-300 kbits/s (limité à 500 kbit/s)
 - pas de mécanisme de «police» mais une déontologie de l'utilisateur
- applications
 - annuaires de session (sd, sdr)
 - conférences audio (vat, nevot, rat)
 - conférences vidéo (nv, ivs, vic, nevit)
 - tableau blanc (wb)
 - éditeur de textes (nte)
 - jeux distribués interactifs (MiMaze)

Copyright 2004 by Timur Friedman

Plan

- Introduction
- Le multicast au niveau réseau
- Le multicast au niveau transport
 - fiabilité
 - SRM
 - RMTP
- Perspectives de recherche

Copyright 2004 by Timur Friedman

Fiabilité

peut-on étendre l'approche utilisée en unicast (ACK) ?

- chaque destination doit envoyer un ACK pour chaque (groupe de) message(s)
- un msg est retransmis jusqu'à réception d'un acquittement de chaque
- congestion du réseau
- 8 implosion de la source

idée : utiliser des NAK

- le contrôle est déplacé de l'émetteur vers les récepteurs
- la source émet sans se préoccuper des ACK
- les récepteurs détectent les pertes sur «trous» de N° de séquence

de nombreux protocoles ont été proposés

- atomicité : soit 0 soit tous les récepteurs ont reçu le msg
- terminaison : le résultat d'une transm. est connu en un temps fini
- SRM, RMTP, RAMP, RMP, etc.

SRM (1/2)

Scalable Reliable Multicast

- offre une transmission fiable, sans séquencement, «scalable» (car receiverbased + reprise en local)
- 2 composants
 - un composant indépendant de l'application : offre les mécanismes pour demander et récupérer les segments de données manquants
 - un composant dépendant de l'application: est responsable du nommage des segments de façon à ce qu'ils soient identifiés de manière unique par tout le groupe et de l'ordonnancement

idée : un segment manquant n'est pas forcément retransmis par la source

- sur détection d'une perte, la demande de retransm. est Mcastée
- le récepteur le plus proche du demandeur Mcaste la retransmission

Copyright 2004 by Timur Friedman

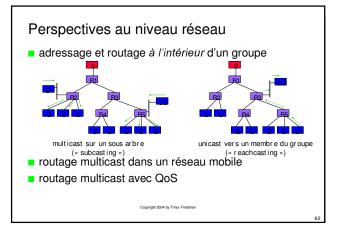
SRM (2/2) ■ <u>but</u> : minimiser le trafic hyp : D5, D6 et D7 ont un seul membre demande la retransm. un msg manquant un seul membre retransmet le msg manguant principe \$3/D3 envoi des requêtes : slotting + damping ■ Request Timer (D4) envoi des retransmissions ■ Repair Timer difficulté D7 dimensionnement des Timers nouveaux msgs estimation du RTT pour chaque req. de retransm. paire (D_i, D_i)

RMTP Reliable Multicast Transport Protocol offre une transmission point à multipoint, fiable, avec maintien de séquence idées notion de hiérarchie réduire l'implosion à la source réduire les temps de réponse notion de reprise en local principe les récepteurs sont groupés dans des régions locales il y a un DR (Designated Receiver) par région, chargé d'agréger les msg de status difficulté : construire l'arbre logique

Plan

- Introduction
- Le multicast au niveau réseau
- Le multicast au niveau transport
- Perspectives de recherche
 - niveau réseau
 - niveau transport
 - niveau application

Copyright 2004 by Timur Friedman



Perspectives au niveau transport

- contrôle de flux/congestion
- fiabilité assistée par les routeurs
- auto-configuration des membres du groupe

Perspectives au niveau application

- allocation des adresses multicast
- nommage d'objets partagés

Copyright 2004 by Timur Friedman

Bibliographie

- [RFC 1112] S. Deering, «Host Extensionsfor IP Multicasting», August 1989.
- [RFC 1075] D. Waitzman, S. Deering, C. Partridge, «Distance Vector Multicast Routing Protocol», November 1988.
- [RFC 1584] J. Moy, «Multicast Extensions to OSPF», March 1994.
- [RFC 1884] J. Moy, «Multicast Extensions to OSPF», March 1994.

 IRFC 21891, Ballardie, «Core Base Trees (CBT Version 2) Multicast Routing: Protocol

 Specification», September 1997.

 [RFC 2201] A. Ballardie, «Core Base Trees (CBT Version 2) Multicast Architecture», September 1997.

 [RFC 2236] R. Fenner, «Internet Group Management Protocol, Version 2», November 1997.
- RFC 2362 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, June 1998.
- June 1998.

 C. Diot, W. Dabbous, J. Crowcroft, «Multipoint Communication: A Survey of Protocols, Functions and Mechanisms», IEEE J SAC, Vol.15, N°3, April 1997.

 S. Floyd, V. Jacobson, S. McCanne, C.G. Liu, L. Zhang, «A Reliable Multicast Framework for Lightweight Sessions and Applications Level Framing», Proc. of ACM SIGCOMM 95, October 1995.

 S. Paul, K.K. Sabnani, J.C. Lin, S. Bhattacharyya, «Reliable Multicast Transport Protocol (RMTP)», IEEE JSAC, Vol.15, N°3, April 1997.

- S. Paul, «Multicasting on the Internet and its Applications», Kluwer Academic Publishers, 1998.