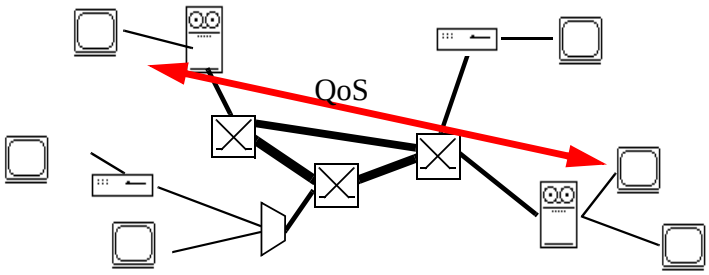
Les mécanismes de contrôle de la qualité de service dans les réseaux informatiques

par Bernard Cousin

(Z:\Polys\QoS\QoS.fm- 5 novembre 2007 18:20)



Plan

- Introduction
- Les différents mécanismes de contrôle dans un routeur
- Le contrôle de conformité du trafic
- Le routage
- Les politiques de gestion des files d’attente
- Les autres techniques de contrôle (contrôle de la congestion)
- Le contrôle d'établissement des connexions
- La notification de congestion
- Le rejet sélectif de paquets
- Conclusion

Bibliographie

- P.Ferguson, Geoff Huston, "Quality of Service", John Wiley & Sons, 1998.

1. Introduction

. Les ressources du réseau sont limitées :
- débit des liens, capacité de stockage des routeurs, etc.

. Les applications soumettent des trafics variés :
- temporellement et quantitativement.

. L'utilisation des ressources doit être optimisée :

=> le multiplexage statistique

Le réseau alloue à chaque connexion un débit inférieur à son débit crête en supposant que la probabilité que toutes les sources transmettent en même temps soit faible (plus le nombre de connexions multiplexées est grand, plus cette probabilité est faible).

=> Congestion

. La congestion des liens : Impossible !
. contrôle d'accès réglé par la politique d'ordonnancement ("scheduling") des paquets par les routeurs.

. La congestion des routeurs : de leurs espaces de stockage ("buffer").

. encombrement => retard des paquets => délai

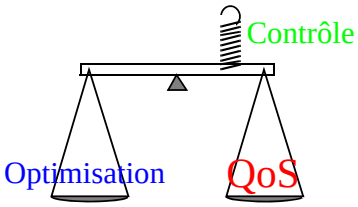
. débordement => pertes des paquets => taux de pertes

1.1. Le contrôle de la qualité du réseau

Besoins contradictoires :
. pour les usagers et leurs applications :
- garantir la qualité du transfert de leurs données (QoS : taux de perte, délai, débit, etc).

. pour les opérateurs et leur réseau :
- optimiser l'utilisation des ressources.

=> Le contrôle gère ce compromis



Propriétés des mécanismes de contrôle :
- flexibilité (s'adaptent à tous les types de trafics)
- efficacité (faible complexité, peu de ressources)
- robustesse (permanence du service en toutes circonstances)

1.2. Difficultés du contrôle

Haut débit :
. les contrôles réactifs sont peu efficaces :
- pendant le délai d'aller et retour une quantité gigantesque de données a le temps d'arriver (de submerger le réseau).
- dépend de la capacité du réseau !
  - LFN (“Long Fat Network”),
  - débit x délai = la capacité.

	débit (Mbit/s)	longueur (km)	capacité (Mbit)
Ethernet	10	2	0,0004
ATM	155	10000	16
X25	0,048	1000	0,0005

Services multiples :
. les applications ont des besoins très variés :
- taux de perte nul, faible, quelconque, etc.
- délai de transmission constant, variable, infini, etc.

Types de trafic multiples :
. constant, périodique, sporadique, continûment variable, variable par palier, quelconque, etc.

1.3. Comment caratériser le débit

. Débit crête ("peak rate") :
- débit maximum atteint,

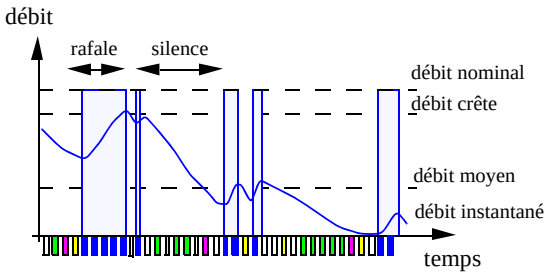
. Débit nominal de la liaison.

. Débit moyen ("mean rate") :
- débit moyen sur un intervalle de temps
- lequel ?

. Débit instantané :

=> débit élémentaire/cellulaire.

. On distingue:
- des périodes d'activité ("burst")
- des périodes de silence.



Le trafic informatique est usuellement "bursty" : par rafale. Cela rend le trafic difficile à gérer.

1.4. Gigue

Gigue : variation du délai de transmission,

. gigue d'insertion (d'empaquetage/dépaquetage) :
- l'instant d'arrivée des données \= d'émission du paquet,
- retard de regroupement des données.

. gigue de multiplexage :
- les paquets (de différents flux) sont multiplexés sur la même liaison,
- à tout moment un seul paquet est émis sur la liaison, les autres attendent !
- politique d'ordonnancement des paquets des différents flux.

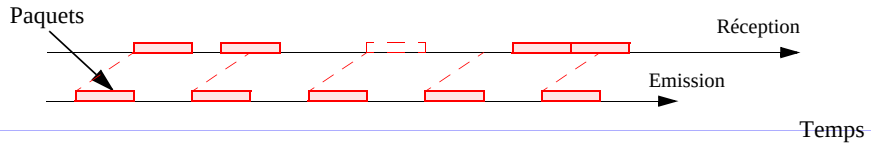
. gigue de charge :
- les délais introduits par le réseau dépendent de sa charge (longueur des files d'attente, durée des traitements, etc).

. gigue de routage (actuellement peu de re-routage dans les réseaux) :
- si la route empruntée par les paquets est modifiée, le délai de transmission est modifié.

La gigue influue sur la forme du trafic :

le trafic reçu n'est plus le trafic émis !

=> difficulté d'établir la conformité du trafic (débit, délai, etc.)

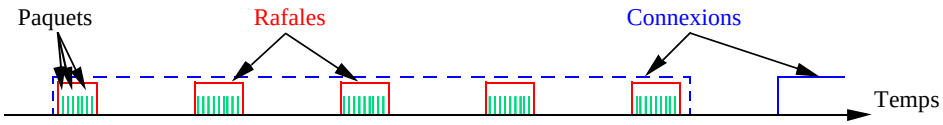


1.5. Niveaux d'analyse du trafic

. Connexion :
- sa nature (variable, en rafale, constante, etc),
- la bande passante requise,
- la qualité de service (QoS : "Quality of Service"):
  . taux d'erreur admissible, délai maximum, variation du délai (gigue), etc.
- échelle de temps : quelques secondes à plusieurs jours.

. Rafale ("burst"):
- fréquence, longueur, intensité (sporadicité : "burstiness"),
- un message => des paquets !
- échelle de temps : la milliseconde.

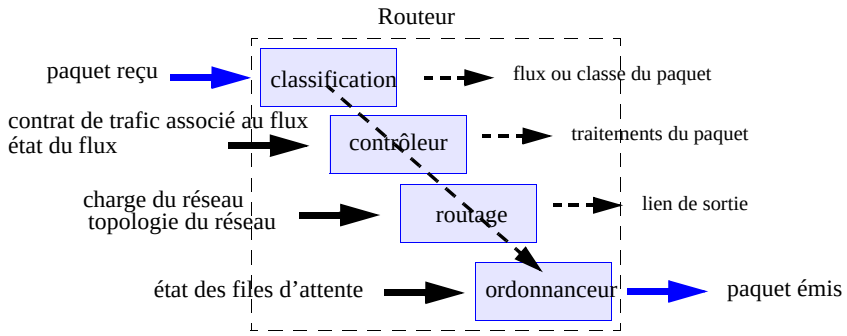
. Paquet :
- échelle de temps : la microseconde.



2. Les différents mécanismes de contrôle dans un routeur

2.1. Introduction

- Classification des paquets
- Contrôle du trafic : mesure et mise en forme du trafic
- Routage
- Ordonnancement



2.2. Classification des paquets

On classe les paquets afin de déterminer les traitements que le routeur devra appliquer aux paquets. Tous les paquets d'un même flux subissent le même traitement.

- Cette détermination peut être effectuée :
  - aux seuls routeurs en frontière du domaine de routage : "ingress router"
  - à chaque routeur
- Cette détermination peut utiliser
  - un seul champ de l'entête du paquet de niveau Réseau
    . exemple : l'adresse de Destination du paquet IP, le label du paquet MPLS
  - un ensemble de champs appartenant à plusieurs niveaux
    . exemple : analyse des champs Protocol et Port Number des l'entêtes IP et TCP pour déterminer avec précision le type de paquet
    => coût / profondeur de l'analyse

• "class-based versus flow-based traffic control"

. le nombre de flux traversant un routeur peut être très grand et la détermination peut être trop coûteuse.

. plusieurs flux peuvent subir exactement le même traitement : on les regroupe dans une même classe

- On définit la notion de classe

. une classe est attribuée à un paquet à l'entrée du domaine d'administration

=> la station émettrice ou le routeur d'accés

. l'appartenance à une classe est identifiée par un marqueur :

=> "Flow label" d'IPv6, "Tag" ou etiquette de MPLS, "CodePoint" d'IPv4 pour DiffServ

2.3. Contrat de trafic

On veut vérifier la conformité du trafic reçu par rapport au contrat de trafic.

L'utilisateur et le gestionnaire du réseau ont négocié un contrat :

- définissant les flux ou les classes de trafic, et les traitements associés
- par exemple : classe "Gold" le trafic issu de la station du directeur,

Classe "Silver" le trafic des videoconférences entre 8h et 20h, classe standard par défaut.

Les paramêtres du contrat sont : débit, délai, taux de perte, gigue, etc.

Exemple ATM :

. type de trafic :

CBR : "constant bit rate," VBR-rt or -nrt : "variable bit rate" ("real time"), ABR : "available bit rate", UBR : "unspecified bit rate".

. descripteur de trafic (débits et QoS) :

. PCR, SCR, MCR : "peak", "sustainable", et "minimum cell rate".

. CLR : "cell loss ratio".

. CTD : "cell transfer delay".

. BT : "burst tolerance", CDV : "cell delay variation".

. pour les 2 niveaux de trafic : CLP=0 et CLP=0+1.

. pour les 2 sens : aller et retour.

3. Le contrôle de conformité du trafic

3.1. Contrôle de trafic

"Usage parameter control" (UPC) ("source policing", "bandwidth enforcement") :
- surveillance des paramètres du contrat,
- durant la phase de transfert des données.
- protection des ressources du réseau contre une inadéquation entre les paramètres du contrat et le comportement réel du trafic :
  => utilisation malveillante,
  => erreurs involontaires.

Idéalement :
- capable de détecter toute situation illégale,
- réaction rapide,
- transparent au trafic conforme,
- simple et efficace.

Existe aussi entre deux réseaux d'opérateurs différents :
. NPC ("network parameter control")

3.2. Techniques

Contrôle de conformité :
- à l'accès du réseau public (UPC),
- entre les réseaux (NPC).

Actions sur les paquets non-conformes :
. destruction
- les paquets non-conformes sont détruits,
- c'est trop tôt !

. marquage ("tagging")
- les paquets non-conformes sont marqués :

Exemples : ATM => bit CLP de la cellule
Frame Relay => bit DE de la trame ("Discard Eligible")
Ou on diminue leur priorité (le paquet est déclassé)

- en cas de congestion, les paquets marqués sont prioritairement détruits.

. réordonnancement temporel (lissage : "traffic shaping")
- les paquets non-conformes sont retardés,
- accumulation dans un tampon du contrôleur.
- grande complexité, mécanisme suceptible d'amplifier la congestion

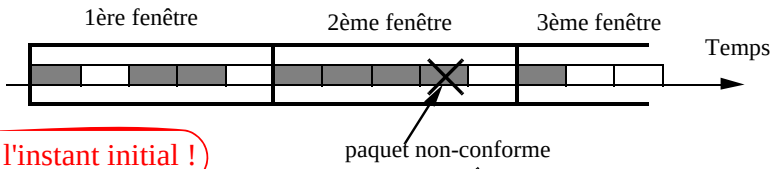
3.3. Contrôle par fenêtre

. La fenêtre définit un intervalle de durée fixe :
- W : la largeur de la fenêtre.

. Le nombre de paquets autorisés par intervalle de temps :
- N, N<W: taille de la rafale.

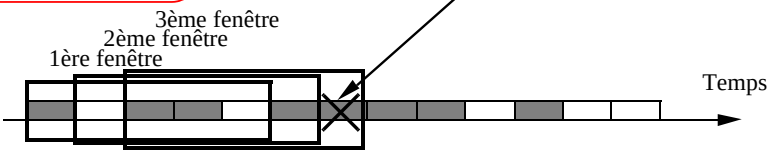
. débit conforme :
- N/W*D (avec D : débit nominal)

"Jumping window" :
W=5, N=3.



Dépend de l'instant initial !

"Moving Window" :



Trop de calcul !

[EWMA] : Exponnentially weighed moving window, [TRJ] : Triggered jumping window

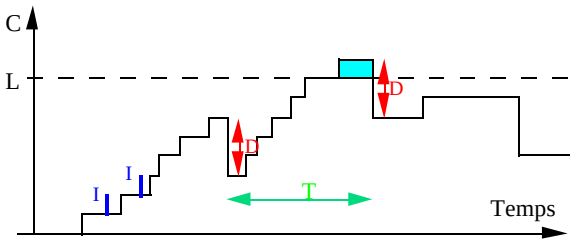
3.4. Leaky Bucket

Seau percé [J.Turner 88]:
- variable C : contenance instantanée du seau
- constante L : capacité maximum du seau
- constante D : taille de la fuite
- constante I : taille du verre
- constante T : période de référence

Débit cible : D/T

Fonctionnement :
- initialement : C=0,
- arrivée d'un paquet(p) : si C< L+I(p) alors C = C+I(p)
  sinon débordement,
- périodiquement (T) : C= C-D (C >= 0).

L = 8, I=1, D=3, T=9



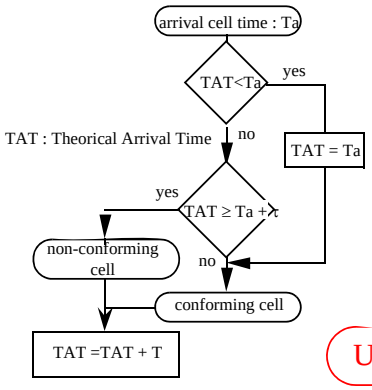
"Token Bucket" : variance du "Leaky bucket" à valeurs entières

3.5. CGRA

"Generic cell rate algorithm" : GCRA(T, τ)
. normalisé l'ATM Forum et l'ITU_T [I.371]
. identique à "Continuous state leaky bucket" (à valeurs réelles)

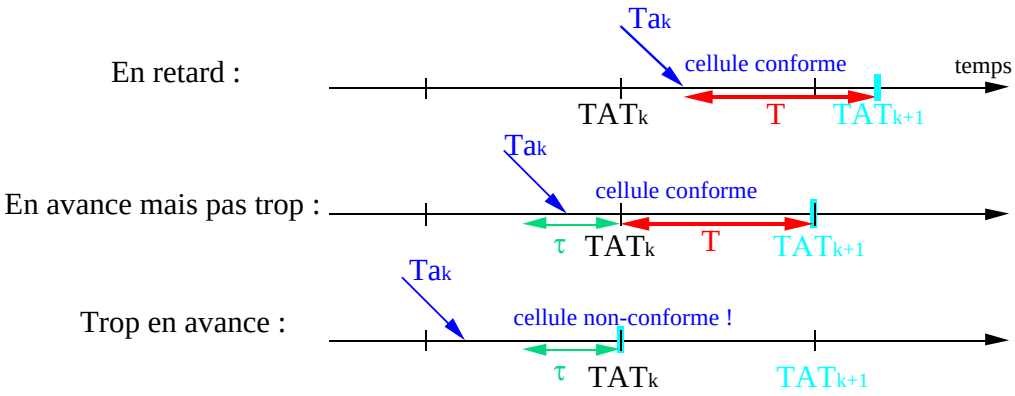
1/T : débit cellulaire contrôlé,

τ : tolérance sur le temps de propagation des cellules.



Un CGRA par seuil et par connexion

3.6. Exemple



Ta : date d'arrivée de la cellule
TAT : date théorique de la cellule
1/T : débit de référence
τ : tolérance

3.7. Conclusion

Ces techniques peuvent être utilisées pour

- mesurer la conformité d'un trafic avec un contrat de débit
- rendre conforme le débit d'un trafic (lisser).

Les paquets non conformes peuvent être

- retardés, c-à-d. que le trafic est lissé (émis à la bonne échéance)
- marqués, c-à-d. associés à une classe inférieure (le paquet sera prioritairement détruit en cas de problème, ultérieurement)
- détruits

Le choix des paramètres est difficile.

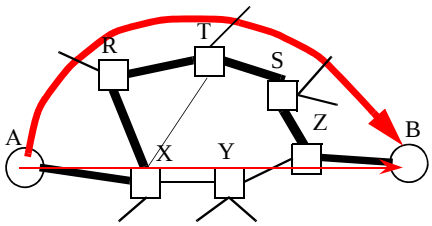
4. Le routage

Des paquets ayant même destination mais appartenant à des flux ayant des caractéristiques différentes, peuvent ne pas suivre le même chemin, et ainsi bénéficier d'une qualité de service différenciée.

- le routage ne se fait pas uniquement en fonction de l'adresse de Destination
  - identification du flux : autres champs ou un champ spécifique

  => la classification qui peut être coûteuse !
- le routage ne se fait pas uniquement en fonction du "hop count"
  - on tient compte de la charge des liens du réseau

  => grande variabilité au cours du temps : manque de précision, délai, instabilité !



Quelques problèmes :

- Les chemins alternatifs sont généralement moins bons (plus longs) et donc augmentent la consommation des ressources.
- Les mécanismes d'équilibrage de charge risquent de désordonner les paquets d'une même connexion :
  - Certains protocoles sont très sensibles au désordonnancement, par ex. TCP !
- Optimisation locale =\= optimisation globale !
- La complexité du calcul sous contraintes des chemins est trop élevé : approximation.

L'exemple d'Internet

- IP et le champ DSCP
- RSVP

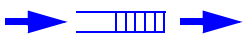
5. Gestion des files d'attente

Lorsque plusieurs paquets doivent être émis simultanément sur le même lien :

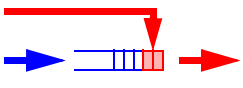
- seul l'un d'entre eux peut être effectivement émis
- les autres doivent être placés en attente

Il existe de nombreuses politiques de gestion des files d'attente

- FIFO
  - l'ordre d'émission est l'ordre d'arrivée
  - avantages et inconvénients
    . simple et rapide car minimale
    . pas de service différencié



- à priorité
  - on définit plusieurs niveaux de priorité
  - les paquets sont placés dans la file en fonction de leur priorité
  - avantages et inconvénients
    . traite différemment les paquets ayant une priorité différente
    . réordonnancement des paquets
    . les paquets de priorité plus faible peuvent ne jamais être acheminés



• "Fair queuing"

- on dispose de plusieurs classes et d'une file par classe,
- à chaque file est allouée une portion égale du débit ("round robin")
- avantages et inconvénients

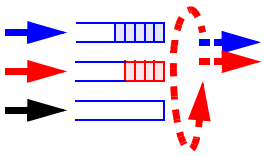
. le trafic en excès d'une classe n'a pas d'impact sur le trafic des autres classes : pas de famine

. la portion non consommée attribuée à une classe peut ne pas être récupérée par une autre classe

. algorithme plus complexe

. la répartition des différents flux entre les différentes classes est soit arbitraire, soit difficile

. la prédiction du service obtenu est difficile (voir instable)



• "Weighted Fair queuing"

- la portion de débit attribuée à une classe dépend de la classe (et peut varier dans le temps)
- avantages et inconvénients

. on peut essayer de prédire le service fourni

. plus grande complexité

. l'allocation précise de la portion de bande passante est difficile

. la portion non consommée attribuée à une classe peut être récupérée par une autre classe (la suivante) : la dernière classe ("best effort"/UBR) qui a un débit réservé nul, récupère toute la bande passante résiduelle

Attention

- en général pas de préemption : un paquet en cours d'émission n'est pas préempté

6. Les techniques de contrôle de congestion

Le contrôle du taux d'erreur.

6.1. Contrôle de flux

Gestion de la disponibilité du récepteur :
. occupation des tampons de stockage.
. capacité de traitement des données.

Protocole Xon/Xoff
. peu précis ou trop contraint !

Sliding window (fenêtre coulissante)
. utilisé par de nombreux protocoles :
HDLC, X25.3, TP, TCP, TPX, SSCOP ("Service Specific Connection-oriented Protocol"), etc.
. numérotation des données (modulo la capacité maximale du champ)
. acquittement,
. largeur de la fenêtre : nombre de données pouvant être émises par anticipation
- implicitement : fenêtre de largeur fixe,
- explicitement : crédit,
. couplé au contrôle d'erreur.

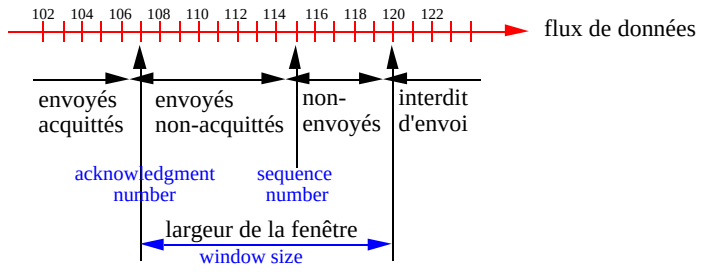
6.2. Sliding Window

Mécanisme permettant à la fois :
- Le contrôle de flux et de congestion.
- Le contrôle des pertes, duplication, déséquencialité.
- La récupération des erreurs par retransmission.
- L'optimisation de l'utilisation de la connexion par l'envoi anticipé de paquets

(avant que les octets des paquets précédents soient acquittés).

Basé sur l'identification des données (octets ou des paquets) :

=> leur numérotation (modulo).



6.3. Méthodes de contrôle de congestion

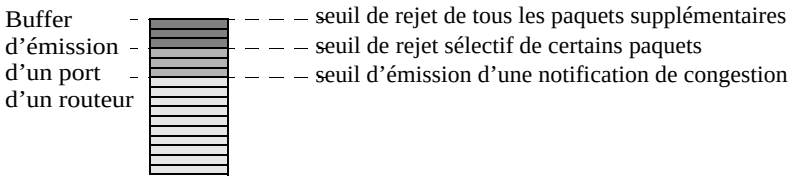
Défini par la recommandation I.371 de l'ITU-T.

Principales méthodes :

. Préventives :
- le contrôle d'admission (d'établissement des connexions),
- le contrôle de trafic.

. Réactives :
- la notification de congestion,
- le rejet sélectif de paquets.

Problème important :
- temps de réaction



7. Le contrôle d'établissement des connexions

7.1. Le CAC

CAC ("Connection admission control") :
- Contrôle les établissements de nouvelles connexions
  . Analyse de la demande (descripteur de trafic)
    => Evaluation de la bande passante équivalente
  . Recherche de chemin optimal,
  . Réservation des ressources et configuration des mécanismes de contrôle des routeurs.

- Optimisation :
  . multiplexage statistique = surallocation
    => faible probabilité d'un grand nombre de rafales simultanées
    => mais pas nulle : perte de paquets !

7.2. Le type de connexion

Connexions permanentes ou établies à la demande

Connexions permanentes ou semi-permanentes :
. lorsque le nombre de noeuds est réduit
. ou lorsque la topologie des communications (et du trafic) est stable dans le temps (ou prévisible)
. par exemple : une application établie sur un petit groupe de stations ou un réseau de coeur ("backbone")

Etablissement automatique ou manuel

Exemple d'ATM

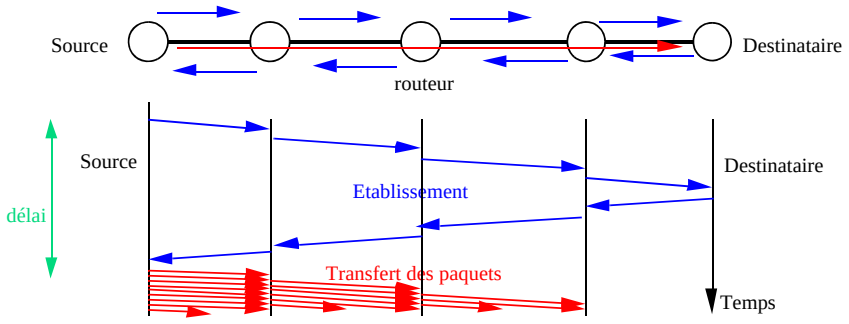
- Utilisation de VPC : conduits virtuels préétablis,
- entre tous les couples de points de trafic importants ($N^2/2$),
- du domaine de la gestion du réseau central.

Exemple identique pour MPLS

- réseau d'infrastructure : "overlay network"

7.3. L'établissement d'une connexion

La durée d'établissement (de négociation) de la connexion est longue :
. délai de propagation de la demande de la source au destinataire aller et retour.
. réservation des ressources au sein de chaque routeur traversé.
. la négociation peut nécessiter des calculs complexes ou plusieurs échanges



Exemple : ATM, TCP, RTP, RSVP, etc.

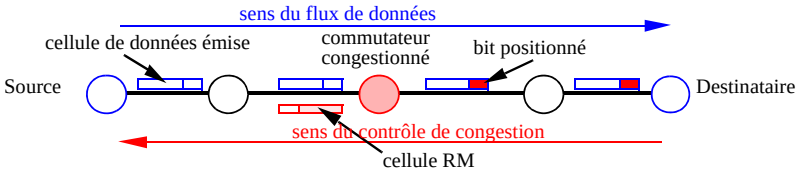
Une solution partielle : FRP ("Fast reservation protocol")
. Les données accompagnent la demande d'établissement de la connexion.
. Si la connexion est refusée, les données sont détruites !

Lorsque l'établissement échoue, il faut tout recommencer sur une nouvelle route !

8. La notification de congestion

8.1. Présentation

. Envoi d'une indication explicite de congestion :
. Par les routeurs :
=> lors d'une perte de paquets
- c'est trop tard !
=> lors d'un dépassement de seuils :
- taux de perte,
- taux d'occupation (des tampons),
- débit, etc.
. Demande de diminution du débit de la source
=> prise en compte optionnelle (IP).
. Exemple d'ATM



8.2. Exemple d'implémentations

Dans ATM

. EFCI : "explicit forward congestion indication"
- dans les cellules de données :
[bit 3 du 4ème octet (bit EFCI du champ PTI) pour les cellules de données (dont le bit 4 est à 0)] .
- utilisation possible des mécanismes de contrôle de congestion des couches supérieures :
. messages spécifiques (destinataire -> source)

. Backward !
- réduction du temps de réaction.
- traitements complexes au sein de chaque commutateur.
- utilisation de cellules spécifiques (RM: "resource management cell")
[code 110 du champ PTI dans l'entête de cellule]
- plus précise : les cellules contiennent plus d'informations (débit explicite, actuel, minimum, longueur des files d'attente, numérotation des cellules RM, etc).

Frame Relay :

• FECN et BECN bits ("Forward/Backward Explicit Congestion Notification")

TCP/IP :

• Source Quench ICMP message/ absence de segments TCP

Quelques problèmes :

Robustesse

. La perte de paquets :
=> de données : perte de précision
=> de notification : perte de détection

. Emission périodique de paquets de contrôle de congestion
(PRCA : "Proportional rate control algorithm")

• Délai de réaction dû à l'acheminement des notifications de congestion

9. Le rejet sélectif de paquets

Lors d'une congestion effective, il faut choisir les paquets à détruire :

• Les paquets des flux les moins importants :

- RUV : "relative usage value".

- Par exemple pour ATM : les connexions de trafic UBR, puis les connexions ayant le CLR le plus élevé

• Les paquets les moins prioritaires parmi le flux

- exemple d'ATM : cellule ayant le Cell Loss Priority bit =1,

• Les paquets ne respectant pas le trafic parmi le flux le moins important

- exemple de Frame Relay : les trames ayant le bit "Discard Eligible"=1.

L'ensemble des paquets appartenant au même message :

. généralement les paquets successifs d'une même connexion sont sémantiquement liés.
. perte d'un paquet => perte de la totalité du message.
. destruction de tous les paquets du message à partir de la détection de la congestion.
. nécessite une marque de fin de message ("drop tail")
. exemple d'ATM :
[bit 2 du 4ème octet (bit ATM_user-to-user.indication du champ PTI) pour les cellules de données (dont le bit 4 est à 0)

Le rejet peut être anticipé

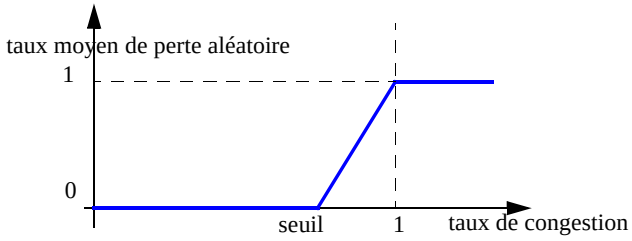
• RED : "Random Early Discard"

- on s'appuie sur les mécanismes de contrôle de congestion du protocole de la couche supérieure : TCP

- à partir d'un certain seuil d'occupation de l'espace de stockage associé à un port de sortie, un certaine proportion de paquets sont choisis aléatoirement pour être détruits

- la proportion croit avec le niveau de congestion

. un nombre croissant de connexions sont touchées : le contrôle de congestion est de plus en plus drastique



10.Conclusion

Tous les mécanismes décrits précédemment ne sont pas tous utilisés ni utiles pour tous les types de trafic :

. Cependant on les retrouve dans beaucoup de types de réseau : ATM, IP, FR, etc.

. Il y a une certaine interdépendance entre ces mécanismes

Il existe d'autres mécanismes :

. Adaptation des paramètres du trafic ("Bandwidth renegociation")

. Notification des paramètres de QoS précède les blocs de données

- par ex. ABT : "ATM block transfer",

. "Credit-based Flow control"

- fenêtre coulissante entre routeurs adjacents,

=>temps de réaction plus court,

=> complexité des routeurs.

Il existe d'autres problèmes :

• Re-routage

- contournement des pannes et des congestions,

• Conception de l'architecture du réseau : localisation et capacité des équipements

• Mesure, analyse et prévision du trafic