

Brève introduction à La Qualité de Service sur Internet

Chaput Emmanuel

2011-2012

Notes :

- 1 Introduction
- 2 Un exemple : la modélisation du débit
- 3 La qualité de service dans les réseaux circuit
- 4 La qualité de service dans les réseaux paquet
- 5 La qualité de service sur Internet

Notes :

Introduction

Introduction

- 1 Introduction
 - De quoi parle-t-on ?
 - Les difficultés

Notes :

De quoi parle-t-on ?

- 1 Introduction
 - De quoi parle-t-on ?
 - Les difficultés

Notes :

Remarque préliminaire

La qualité d'un service ne se mesure pas à ses seules performances. D'autres facteurs tout aussi importants (parfois plus, selon le contexte) doivent être pris en considération

- La sécurité
- La disponibilité
- La sûreté/la fiabilité
- ...

Ce cours n'a pas la prétention d'aborder ces sujets. Chacun d'entre eux mérite des cours et des livres complets.

Objectif de ce cours

Nous nous intéresserons ici aux outils permettant à un réseau (qualifié de prestataire/opérateur) de garantir un certain niveau de performance à une application (qualifiée de client ou utilisateur).

Notes :

Qualité de service : définition

Pas de définition universelle

Une tentative de définition

"A set of quality requirements on the collective behavior of one or more object."

Une autre

"The set of those quantitative and qualitative characteristics of a distributed multimedia system, which are necessary in order to achieve the required functionality of an application." [?]

Une dernière pour la route

Un ensemble de paramètres permettant d'évaluer les performances d'un service offert par un système et la satisfaction qu'en tirera l'utilisateur.

Notes :

Les difficultés

1 Introduction

- De quoi parle-t-on ?
- Les difficultés

Notes :

Les difficultés

Premier niveau de difficultés, la *spécification* des contraintes/besoins.

- Comment caractériser le trafic engendré par une application ?
 - Incidence sur la charge du réseau, et donc sur les moyens à mettre en œuvre
- Comment caractériser les besoins en qualité de service ?
 - Besoin de définir des métriques modélisant les besoins applicatifs
 - Paramètres pertinents et pouvant être définis sans ambiguïté
 - Traduction de ces paramètres entre les différentes entités
- Comment caractériser la qualité du service rendu par le prestataire ?
- Comment négocier la qualité de service ?
 - Notion de contrat entre un prestataire et un client
 - *Service Level Agreement* (SLA)

Notes :

Les difficultés

Deuxième niveau de difficultés, la *mise en œuvre* de la qualité de service.

- Comment assurer le respect du contrat ?
 - Mécanismes d'acceptation (refus) d'un contrat (qui ne pourra être assumé)
 - *Call/Connexion Admission Control* (CAC)
 - Mécanismes de rejet éventuel du trafic du client qui ne respecte pas sa part du contrat
 - *Policing, shaping*
- Comment mettre en œuvre la qualité de service ?
 - Implantation de mécanismes dédiés
 - Positionnement des ces mécanismes dans les différentes entités impliquées
 - Coordination de ces entités entre elles

Notes :

Les caractéristiques et besoins applicatifs

Historiquement

- Téléphonie
 - Temps de traversée faible et constant
 - Débit constant
 - Erreurs acceptables
 - Utilisation sporadique
- Télévision
 - Temps de traversée quelconque mais constant
 - Diffusion à débit (constant) élevé
 - Erreurs acceptables
 - Transmission permanente
- Informatique
 - Temps de traversée "quelconque"
 - Débits variables
 - Erreurs inacceptables
 - Utilisation sporadique

Notes :

Les caractéristiques et besoins applicatifs

Convergence des services (voix, télévision, accès Internet sur le même équipement)

- Comment traiter les flots de données ?
 - Traitement individuel
 - Réaliste à l'échelle de l'Internet ?
 - Traitement par groupe
 - Quelle efficacité pour chaque flot ?
- Comment modéliser le trafic ?
 - Certains trafics sont très difficiles à modéliser
 - Le trafic global est trop complexe
 - Métrologie
- Comment dimensionner le réseau ?
 - Plus complexe encore que les réseaux dédiés
 - Voir par exemple ATM
 - Approche plus réactive
 - Ingénierie de trafic

Notes :

Quelles métriques ?

- Débit
 - Débit utilisable pour un trafic
 - Quelle mesure (échelle de temps) ?
 - Débit moyen, burst, ...
- Taux de perte
 - Nombre moyen max de pertes
 - Quelle distribution ?
- Gigue
 - Variation du délai
 - Important pour les applications "temps-réel"
- ...
 - Satisfaction de l'utilisateur
 - Paramètre difficile à traduire entre couches

Notes :

Quels outils ?

- Préventifs
 - Acceptation/refus de connexion (CAC)
 - Réserve de ressources
- Réactifs
 - Changements de routes
 - Adaptation des applications
- Curatifs
 - Altération de la qualité
 - Rejet/déclassement de flux moins prioritaires

Notes :

Un exemple : la modélisation du débit

- ② Un exemple : la modélisation du débit
 - Le problème
 - Un exemple de mise en œuvre : le token bucket
 - Un autre exemple de mise en œuvre : le GCRA
 - Bursts instantanés ?

Notes :

Le problème

- ② Un exemple : la modélisation du débit
 - Le problème
 - Un exemple de mise en œuvre : le token bucket
 - Un autre exemple de mise en œuvre : le GCRA
 - Bursts instantanés ?

Notes :

Le problème

- Qu'est-ce qu'un débit ?
 - Un volume d'information par unité de temps
- Comment le définir dans le cadre d'un contrat de QoS ?
 - Quelle échelle de temps ?
 - Quelle variation acceptable ?

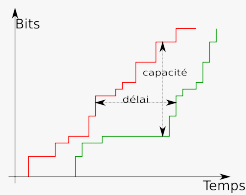
Un exemple : 64 Kbit/s

- Pour un service téléphonique
 - Un échantillon de 8 bits toutes les 125 μs
 - Pendant toute la durée de la communication et "à chaque instant"
- Pour un service de transfert de fichier
 - Un fichier de 1 Mo transféré en 128 secondes
 - Durée moyenne et rythme éventuellement variable

Notes :

Définition microscopique

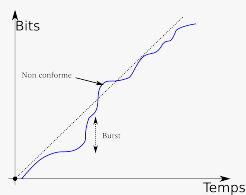
- Observons ce qui se passe sur un équipement (un routeur) à l'échelle temporelle d'un paquet
 - Chaque instant d'arrivée
 - Chaque instant de départ
- La "pente" de chaque courbe est un débit
 - Localement "infini" (égale au débit du support)
- Notion de *courbe de trafic*
 - $a(t)$ est le nombre de bits cumulés à l'instant t



Notes :

A une échelle "mésoscopique"

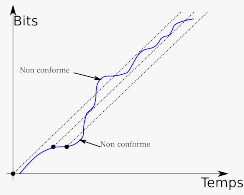
- Contraindre un trafic à respecter un débit
 - Maintenir le trafic $a(t)$ sous la droite (débit constant) $y = r.t$
- Simple à implanter
- Efficace en moyenne sur du long terme
- Et à court terme ?
 - N'empêche pas les salves (*burst*)



Notes :

Translation

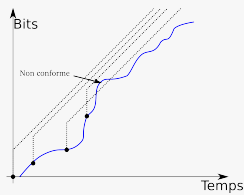
- Idée
 - Repartir de (0, 0) à chaque instant significatif
 - La demi-droite du débit "glisse" le long de la courbe
- On oublie le passé
 - $\forall t_0 \geq 0, \forall t \geq t_0, a(t) < a(t_0) + r \cdot (t - t_0)$
- Notion d' *enveloppe de trafic*



Notes :

Introduction des bursts

- Trop contraignant (bursts impossibles)
 - Ne permet pas de "rattraper le retard"
 - Le débit ne peut que décroître
- On autorise des bursts d'amplitude b
 - $\forall t_0 \geq 0, \forall t \geq t_0, a(t) < a(t_0) + b + r \cdot (t - t_0)$
- Comment implanter ?



Notes :

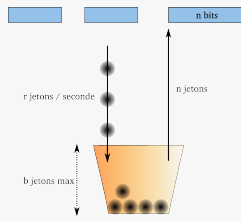
Un exemple de mise en œuvre : le token bucket

- ② Un exemple : la modélisation du débit
 - Le problème
 - Un exemple de mise en œuvre : le token bucket
 - Un autre exemple de mise en œuvre : le GCRA
 - Bursts instantanés ?

Notes :

Un exemple de mise en œuvre : le token bucket

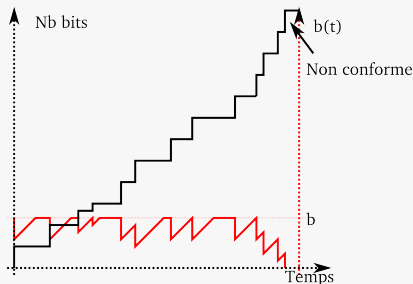
- Principe du seau à jetons
 - Un seau de capacité b tokens
 - est rempli au rythme permanent de r token par seconde
- Un paquet de n bits est conforme si
 - Le seau contient au moins n tokens
 - Les n token sont alors supprimés du seau
- Sinon, selon la stratégie et les ressources, le paquet est
 - Détruit
 - Mis en attente
 - Marqué



Notes :

Un exemple de mise en œuvre : le token bucket

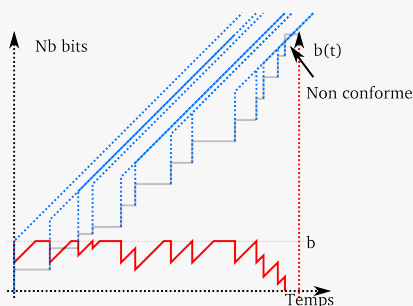
- Exemple d'application du *token bucket* à une courbe de trafic
- À chaque paquet, on décroît $b(t)$ de la taille du paquet
- $b(t)$ croît linéairement avec t (coefficient r)
- $\forall t, b(t) \leq b$



Notes :

Un exemple de mise en œuvre : le token bucket

- Superposons en chaque temps t_0 une courbe de la forme $a(t_0) + b(t) + c.(t - t_0)$
- Attention, ici, on utilise $b(t)$ pour un effet mémoire
- Ainsi le *token bucket* déclare le dernier paquet non conforme



Notes :

Un autre exemple de mise en œuvre : le GCRA

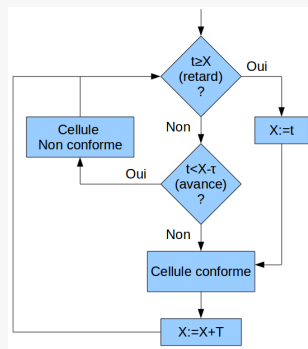
2 Un exemple : la modélisation du débit

- Le problème
- Un exemple de mise en œuvre : le token bucket
- Un autre exemple de mise en œuvre : le GCRA
- Bursts instantanés ?

Notes :

L'algorithme GCRA

- Algorithme $GCRA(T, \tau)$
- T est la durée entre deux cellules ATM
- τ est une tolérance sur T
- Algorithme appliqué à chaque cellule
 - t date d'arrivée réelle de la cellule
 - X est la date prévue
 - Initialement $X = t_0$ (date d'arrivée de la première cellule)



Notes :

Bursts instantanés ?

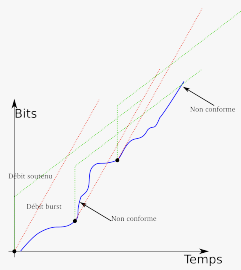
2 Un exemple : la modélisation du débit

- Le problème
- Un exemple de mise en œuvre : le token bucket
- Un autre exemple de mise en œuvre : le GCRA
- Bursts instantanés ?

Notes :

Bursts instantanés ?

- Problème avec les outils précédents
 - Le débit des bursts n'est pas limité
 - Il faut donc les limiter en amplitude
- Autre possibilité : utiliser deux jeux de paramètres
 - Un pour le débit moyen (*sustainable rate*)
 - Un pour le débit crête (*peak rate*)



Notes :

La qualité de service dans les réseaux circuit

8 La qualité de service dans les réseaux circuit

Notes :

La qualité de service dans les réseaux circuit

- Typiquement, réseau téléphonique
- Service fourni
 - Transfert de la voix
- Paramètres de qualité de service importants
 - Disponibilité du système (probabilité de refus faible en conditions normales)
 - Qualité : 8 bits 8000 fois par seconde, full duplex
 - Temps d'aller retour constant, très faible
- Techniques
 - Circuits commutés
 - Demande d'appel accepté ou refusé
 - Ressources réservées en cas d'acceptation
 - Réseaux dimensionnés en fonction de matrice de trafic
 - Qualité des transmissions surveillée
 - Multiplexage temporel synchrone

Notes :

La qualité de service dans les réseaux paquet

4 La qualité de service dans les réseaux paquet

Notes :

La qualité de service dans les réseaux paquet

- Par exemple Internet, mais pas uniquement
- Service fourni
 - Acheminement fiable de données
- Paramètres de qualité de service importants
 - Fiabilité
 - Aussi vite que possible, mais sans contrainte (historiquement)
 - Débit très variable, et non prédictible
- Techniques
 - Multiplexage temporel asynchrone
 - Reprise sur perte/erreur
 - Mise en place de circuits virtuels
 - Acceptation ou refus en fonction de l'état des ressources
 - Permet une garantie statistique
 - Coûteux sur des communications dynamiques
 - Impossible avec IP

Notes :

La qualité de service sur Internet

5 La qualité de service sur Internet

Notes :

Le problème

Réseaux à commutation de paquets : multiplexage temporel asynchrone

- Comportement best-effort
- Routage de proche en proche
- Pas de circuit virtuel

Quelles garanties de service de bout en bout ?

Notes :

Les enjeux

Convergence des technologies

- Réseaux aux caractéristiques très hétérogènes
- Interconnexion toujours plus forte

Convergence des services

- L'utilisateur ne veut qu'une interface ("Triple play")
- Les fournisseurs aussi

Un vieux serpent de mer !

- ISDN, ATM, ...

Notes :

Au delà des rêves de l'utilisateur

En fait, pour les opérateurs

- Utiliser des technologies plus économiques
- Maîtriser la congestion
- Plus généralement ingénierie de trafic
- Fournir un service apparenté à du circuit
 - Pour pouvoir le garantir ?
 - Pour pouvoir le facturer ?

Chassez l'ITU-T par le paquet il revient par la QoS ...

Notes :

Qualité de Service et IP ! ?

Mariage apparemment impossible

- IP sans connexion, sans état
 - IP fournit un service "best effort"
- QoS naturellement orientée connexion
 - Gestion de contrats de trafic (SLA)
 - Contrôle d'admission
 - Réservation de ressources ?
 - Différenciation des flots

Notes :

Qualité de Service et IP

- Quelle est la cause majeure du problème ?
 - Réseau IP : réseau à files d'attente
 - Les files d'attente des routeurs sont les principales responsables des pertes/délais/jigues
- Un objectif fondamental
 - Maintenir les files d'attente stables et peu remplies
 - Utile (voire nécessaire) mais non suffisant
- C'est l'objectif du contrôle de congestion
 - Depuis la deuxième moitié des années 80 dans IP
 - Essentiellement par des mécanismes réactifs
 - Jusque là sans vision globale
- C'est également l'un des buts de l'ingénierie de trafic
 - Optimiser l'utilisation des ressources à plus ou moins long terme

Notes :

Les différents axes de la QoS

Des architectures

- Quelle structure ?
- Comment s'organise-t-on ?

Des protocoles

- Quelle communication entre les entités ?
- Comment dialogue-t-on ?

Des mécanismes

- Quelle mise en œuvre ?
- Comment fait-on ?

Notes :

Des architectures

Répartition des responsabilités

- Qui décide ?
- Qui implante ?

Interfaces entre les entités

- Ajout de primitives de service supplémentaires
- Nécessité de décrire les paramètres

Répartition des entités sur le réseau

- Où sont les entités de décision, de mise en œuvre ?

Notes :

Des protocoles

Plan de contrôle

- Routage
- Description de ressources
- Description de politiques (d'admission, ...)
- Contrôle de congestion

Plan de données

- Transport
- Fiabilisation
- Estampillage
- Marquage
- Contrôle de congestion

Notes :

Des mécanismes

- Réserve de ressources
 - Afin de pouvoir garantir un service
- Classification
 - Des paquets en fonction de spécifications de trafic
- Mesure
 - Afin de vérifier la conformité au contrat
- Conditionnement
 - Afin d'assurer la conformité au contrat
- Contrôle de congestion
 - Afin de gérer le comportement du réseau

Notes :

[1] Andreas Vogel, Brigitte Kerhervé, Gregor v. Bochmann, and Jan Gecsei.
Distributed multimedia applications and quality of service : a survey.
In *CASCON '94 : Proceedings of the 1994 conference of the Centre for Advanced Studies on Collaborative research*, page 71. IBM Press, 1994.

Notes :

Notes :

Notes :
