

Modélisation du trafic multimedia dans les réseaux de télécommunication

Samir Tohmé
Laboratoire PRISM

Objectifs

- Valider les spécifications fonctionnelles des protocoles
- Modéliser de la demande (caractériser le trafic) en vue de
 - Dimensionner le réseau
 - Concevoir un contrôle d'admission
 - Vérifier la consistance et instancier des normes
 - Faire évoluer le réseau en fonction de la demande

Modélisation du réseau

- Comme la demande est aléatoire et les ressources sont dynamiquement partagés, un grand réseau de télécommunication peut être modélisé par un réseau de files d'attente de grande dimension (ouvert, fermé)
- Les utilisateurs du réseau peuvent être fixes ou mobiles
- Importance de modéliser en relation avec l'échelle de temps appropriée

Echelles de temps

- L'heure de pointe dans le mois, dans la semaine, sur 24 heures
- L'appel en mode circuit : durée 2 à 3 minutes avec la téléphonie
- L'activité :
 - 0.1 Erlang pour la téléphonie
 - temps de réponse : quelques secondes dans les systèmes transactionnels
- La rafale (flots TCP : longs, courts quelques dizaines de millisecondes dans les réseaux à haut débit)
- Le paquet (microseconde dans les réseaux à haut débit)

Modèles de files d'attente de grandes dimensions

- En général, les réseaux de files d'attente qui apparaissent dans la modélisation ne sont pas tractables analytiquement
- Sous certaines hypothèses (restrictives), les réseaux sont à forme produit ce qui permet une résolution analytique
- Les réseaux à forme produit ouverts requièrent la résolution de systèmes linéaires de très grande dimension
- Les réseaux à forme produit fermés requièrent le calcul de la constante de normalisation

Quelques outils de simulation spécialisés dans l'analyse des réseaux

- Block Oriented Network Simulator BONeS
Designer
- Optimized Network Engineering Tools
OPNET
- Queuing Network Analysis Package
QNAP
- Network Simulator NS
- NIST (développé par National Institute of
Standards and Technology)

Modèles de sources

- Source voix
- Source WWW
- Source Email
- Source FTP
- Source SMS
- Source vidéo

Source voix

- Codage Adaptative Multi-Rate AMR
- Le codeur AMR change de débit en temps réel en fonction de la qualité du canal
- La longueur d'une trame de parole nécessaire au codage est de 20 ms
- Les débits standardisés du codeur AMR ainsi que les tailles des trames sont :

Débit (kbit/s)	12,20	10,20	7,95	7,40	6,70	5,90	5,15	4,75
Taille de trame (bit)	244	204	159	148	134	118	103	95

Les modes du codeur AMR

Trame de durée 20 ms

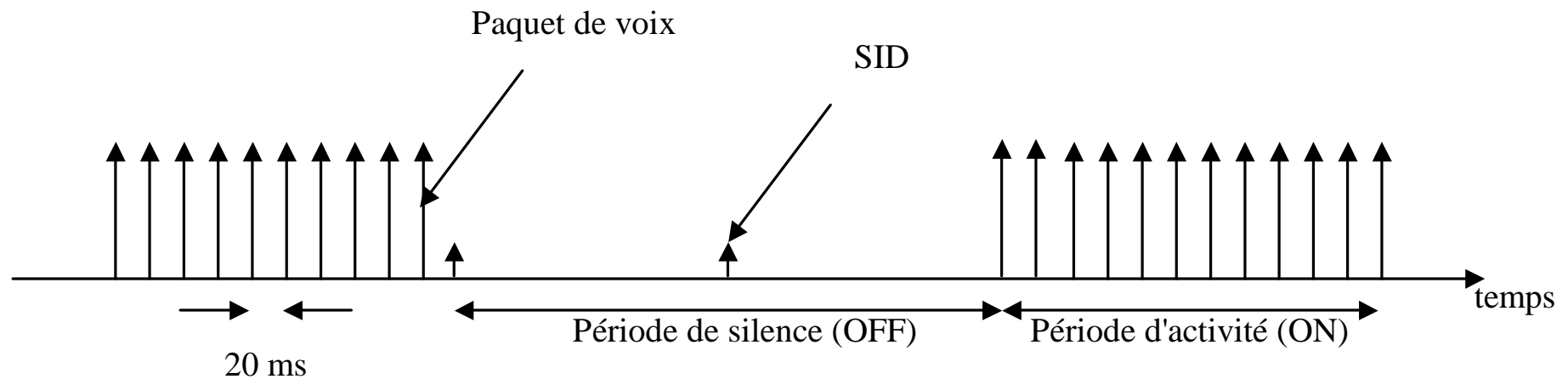
Modèle de source de voix (suite)

- Une source de voix AMR peut être représentée par un modèle ON/OFF
- Les périodes ON et OFF correspondent respectivement aux périodes d'activité et de silence de la source
- La durée de la période ON (resp OFF) est exponentiellement distribuée de moyenne 0.4 secondes (resp 0.6 secondes)
- Pendant la période ON, des paquets de taille constante sont envoyés toutes les 20 ms

Modèle de source de voix

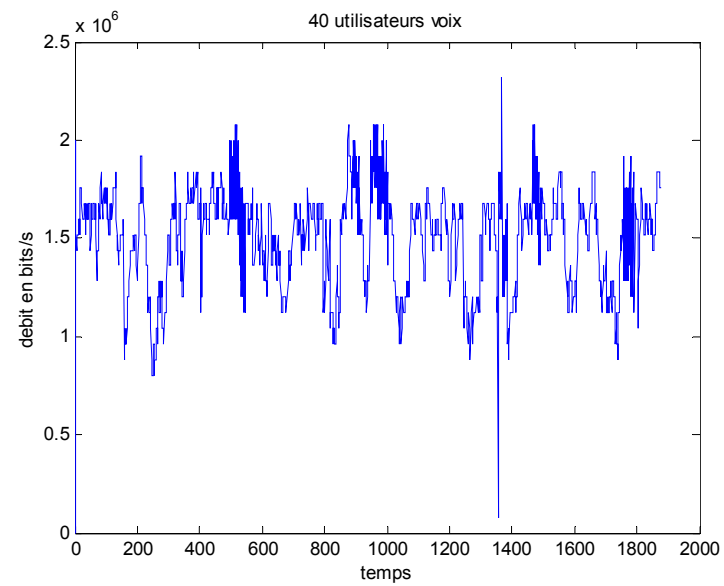
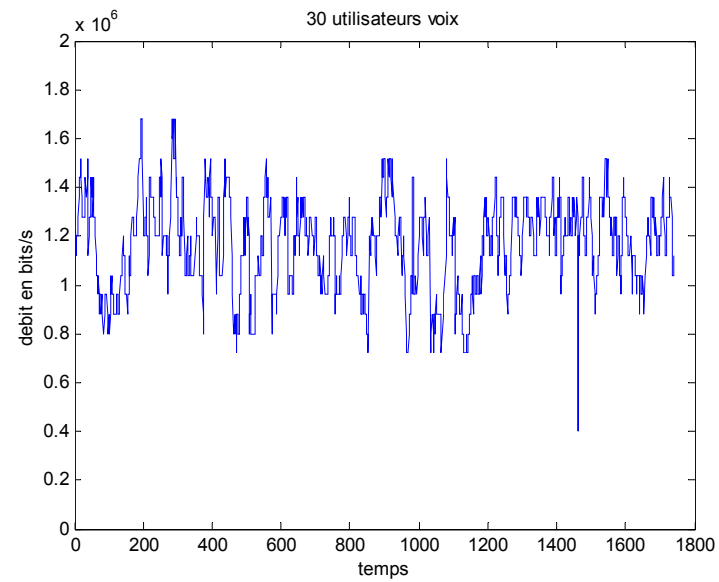
- Après la fin de la période d'activité de la source de voix, le codeur AMR envoie un paquet appelé SID (*Silence Descriptor*) pour indiquer le début d'une période de silence
- Ce paquet est envoyé après 20 ms de l'envoi du dernier paquet de voix
- Pendant la période de silence, des SID sont envoyés pour indiquer les changements du niveau du bruit de fond
- Si ce niveau ne change pas pendant cette période, alors aucun SID n'est envoyé
- La taille d'un SID est de 39 bits auxquels 1 bit de bourrage est ajouté

Modèle d'une source de voix AMR



Modèle d'agrégation de sources de voix

- La superposition de N processus ON/OFF indépendants homogènes est un processus de naissance et de mort à $N+1$ états
- Il est possible de calculer les taux de transitions infinitésimaux entre deux états consécutifs
- Cela permet d'éviter la simulation lourde de N sources en activité



Flux agrégé à la sortie de la couche MAC

Rappel sur la loi de Pareto

- A l'origine il s'agit des distributions de probabilité gamma
- Cette loi permet de modéliser la taille d'un segment TCP (rafale) dans sa version tronquée
- Elle permet également de modéliser la longueur d'un paquet dans sa version tronquée également

Distribution de Pareto standard

$$\left\{ \begin{array}{l} f_X(x) = \frac{\alpha \cdot k^\alpha}{x^{\alpha+1}} \quad ; \quad \forall x \geq k \\ F_X(x) = 1 - \left(\frac{k}{x}\right)^\alpha \quad ; \quad \forall x \geq k \\ \mu = \frac{k \cdot \alpha}{\alpha - 1} \quad ; \quad \alpha > 1 \\ \sigma^2 = \frac{k^2 \cdot \alpha}{(\alpha - 2)(\alpha - 1)^2} \quad ; \quad \alpha > 2 \end{array} \right.$$

Distribution de Pareto tronquée

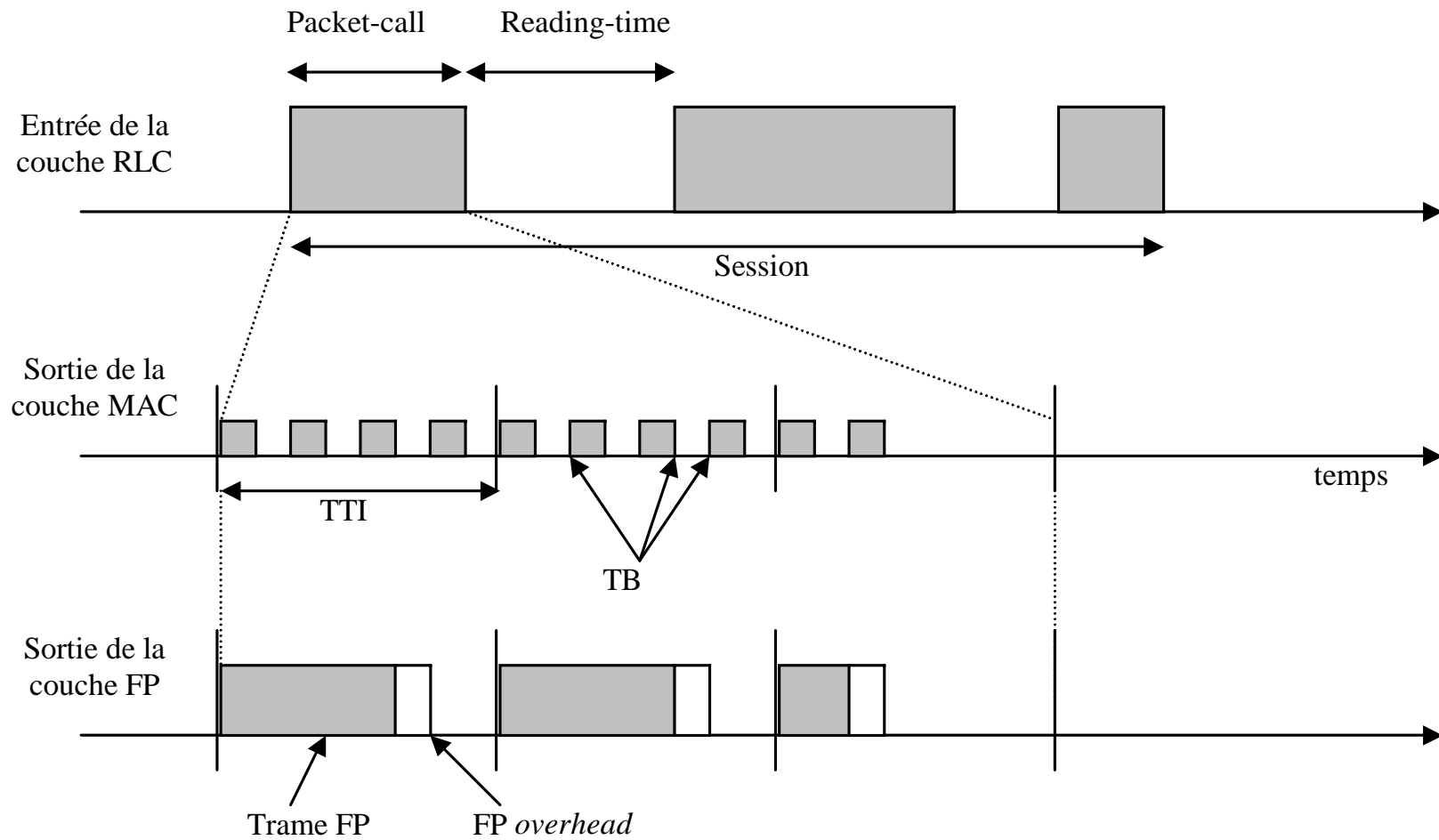
$$f_S(x) = \begin{cases} \frac{\alpha \cdot k^\alpha}{x^{\alpha+1}} & ; \quad k \leq x < m \\ \frac{k^\alpha}{m^\alpha} & ; \quad x \geq m \end{cases}$$

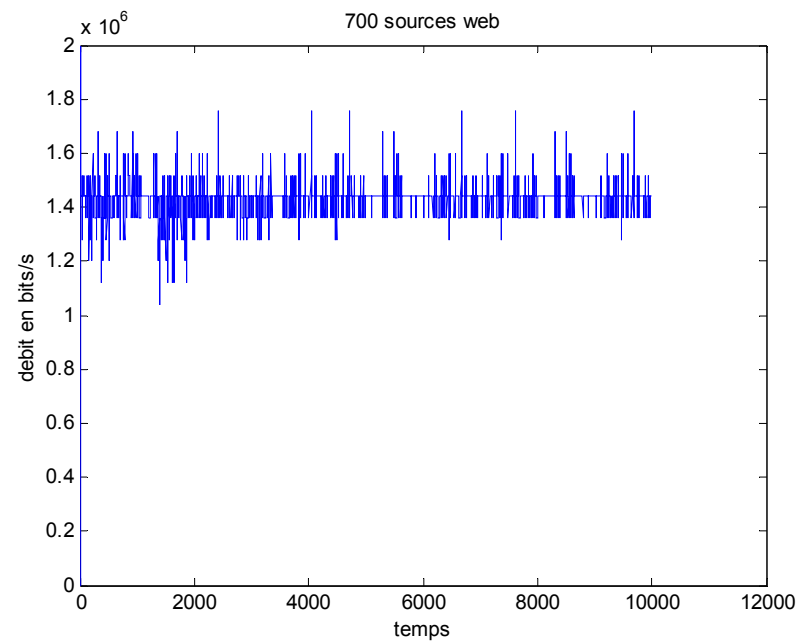
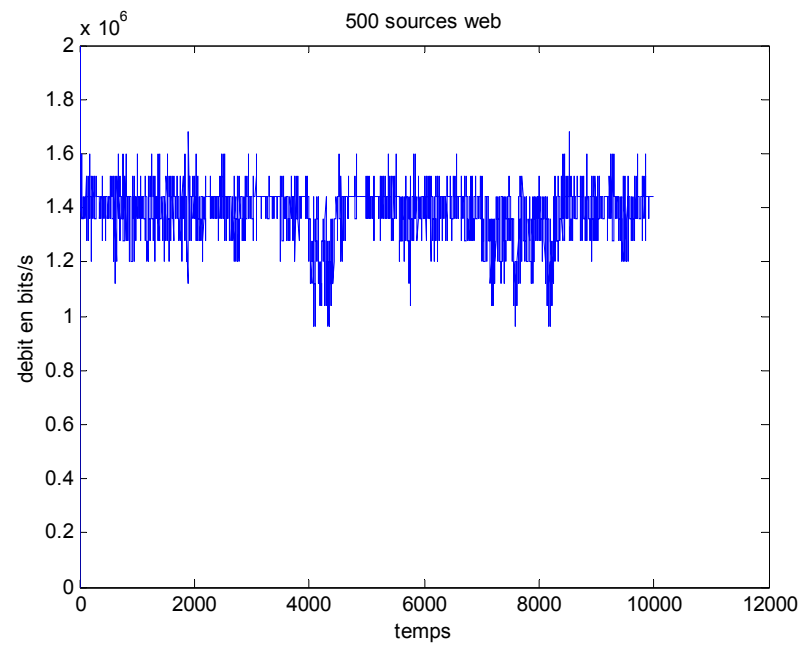
$$\mu = \frac{\alpha \cdot k - m \left(\frac{k}{m}\right)^\alpha}{\alpha - 1} \quad ; \quad \alpha > 1$$

Paramètres pour une source WEB

- μ est la taille moyenne d'un *packet-call*
- k est la taille minimale d'un *packet-call*
- Dans le rapport technique du 3GPP TR 25.933 [92] : $\alpha = 1,1$; $k = 1858$ octets et $m = 5000\ 000$ octets
- On trouve ainsi que la taille moyenne d'un *packet-call* est de 12 000 octets
- Deux *packet-calls* successifs sont séparés par un intervalle de temps appelé "*reading-time*" ou "*thinking-time*" exponentiellement distribué de moyenne 12 secondes
- Le nombre de *packet-calls* dans une session WWW est géométriquement distribué avec une moyenne de 5

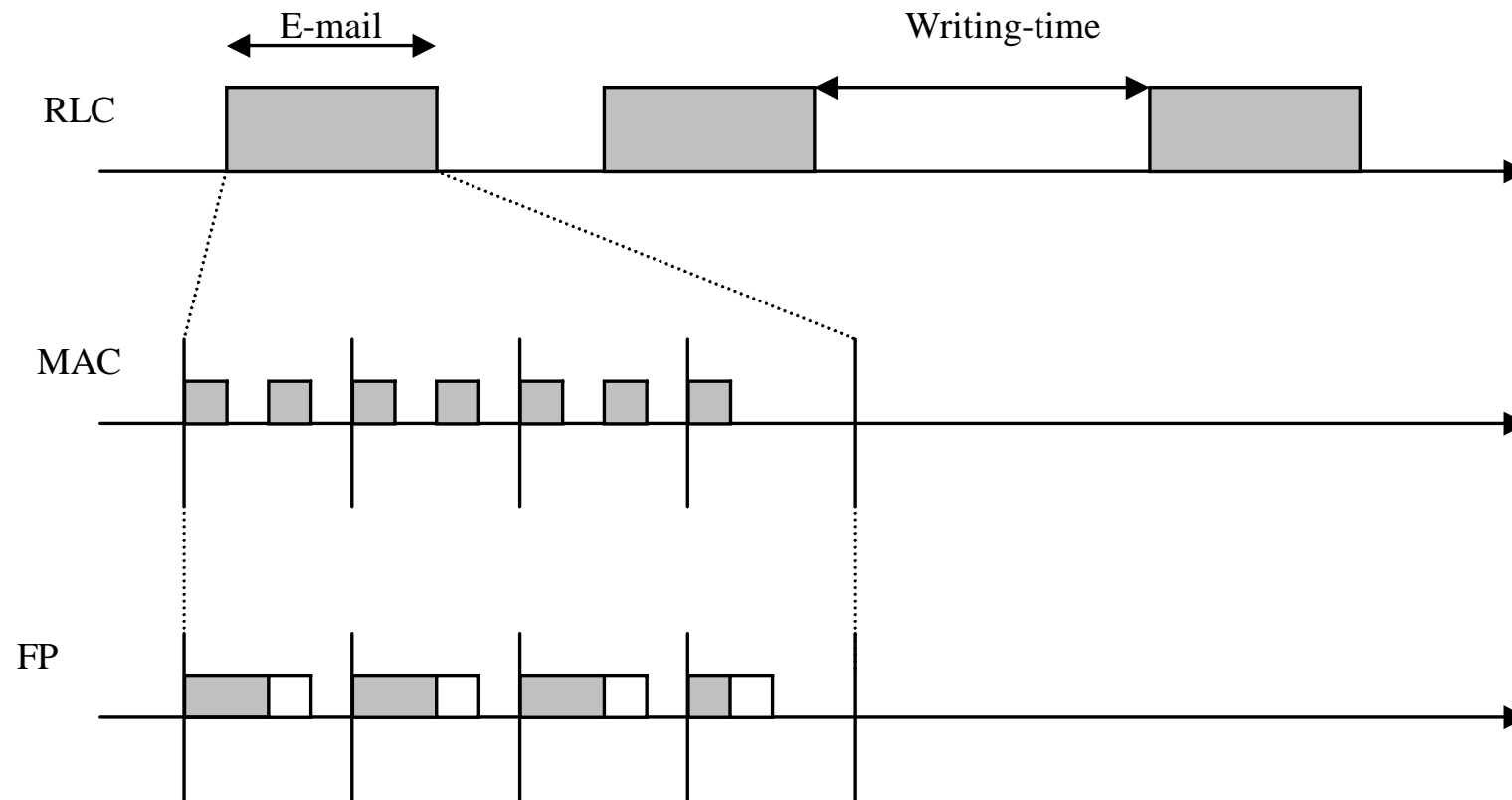
Exemple de l'UMTS

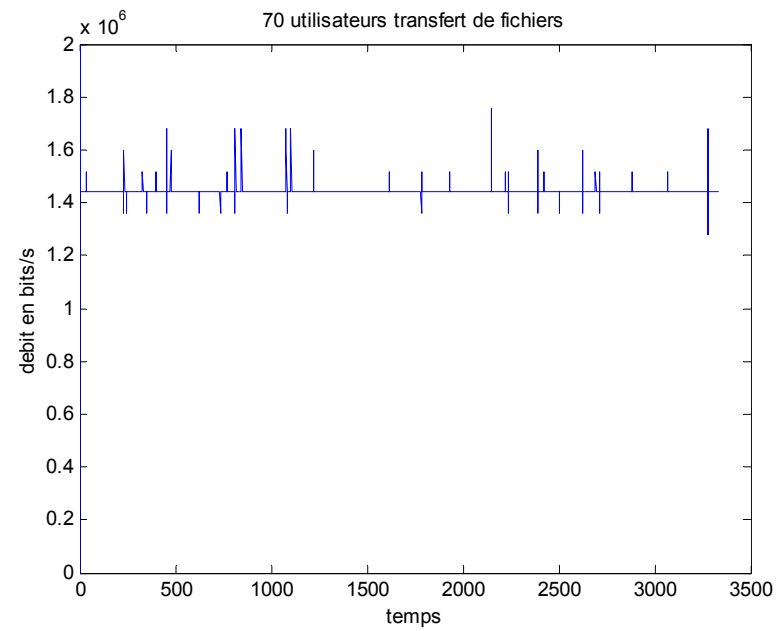
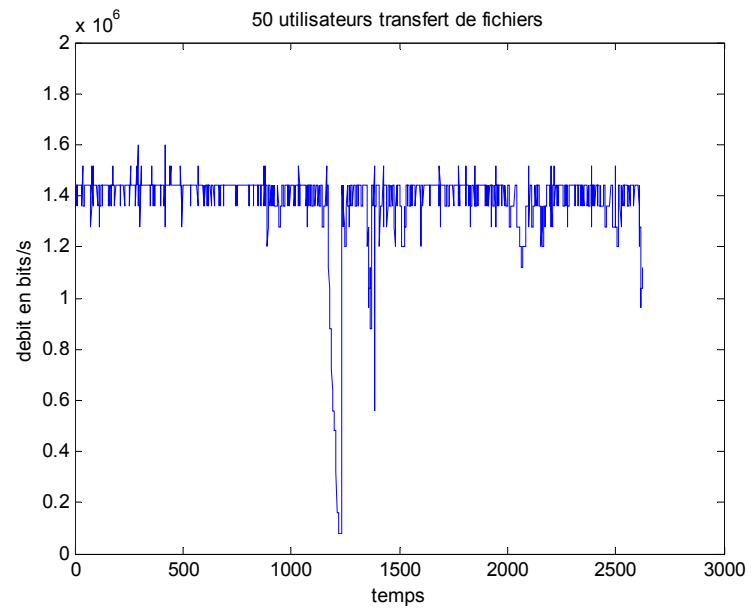




Superpositions de sources WEB (sortie couche MAC)

Profil du trafic E-mail / FTP





Superposition de sources FTP

Autres modèles de sources de données (suite)

- ON/OFF (IPP) où le processus d'arrivée de paquets durant la période ON est Poisson et pas d'arrivées dans la phase OFF
- La distribution de la longueur de paquet suit une loi de Pareto tronquée (ex : 1500 octets, 64 ko)
- La superposition de n sources de ce type est un MMPP dont le processus de commande est de type naissance et de mort à $n+1$ états

Modèles d'agrégation de sources de données

- MMPP général :
 - Processus de modulation : Markov à temps continu et à états finis (phases)
 - Quand le processus de modulation est en phase i , le MMPP est un Poisson d'intensité qui dépend de la phase i
 - Le temps de séjours du MMPP dans la phase i est exponentiellement distribué de paramètre qui dépend de la phase
 - La probabilité de saut entre les phases (i,j) est donnée par la matrice de la chaîne incluse du processus de modulation

Modèles d'agrégation de sources de données (suite)

- Par rapport au modèle de superposition de IPP le processus de modulation est général d'où une portée plus générale du modèle
- Les paramètres à déterminer : le générateur infinitésimal du processus de modulation et les intensités du Poisson dans chaque phase
- La superposition de MMPP est MMPP mais avec un espace de phase multiplicatif... Le modèle MMPP peut difficilement servir pour modéliser une source individuelle

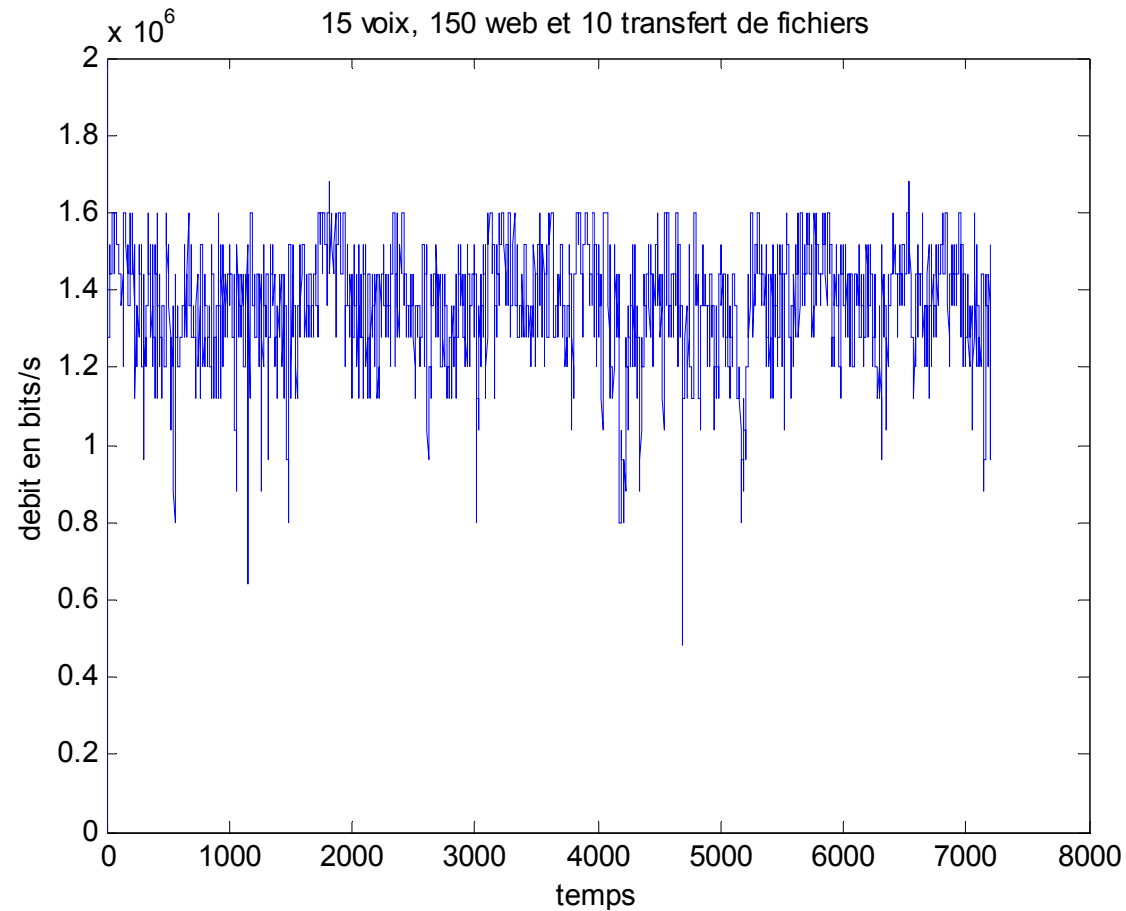
Modèle de source vidéo

- Le trafic est périodique avec une période de 40 ms donc pas de période OFF
- Trames I (Intra), P (predicted), B (backward) avec :
 - Taille de trame I # 3 fois la taille de la trame P
 - Taille de trame P # 2 fois la taille de la trame B
- Les trames sont groupées en petits ensembles GoP, récurrence du GoP IBBPBBPBBPBB (pic de corrélation)
- Paramètres : L distance entre deux trames I consécutives, Q distance entre la trame I et la première trame P du GoP

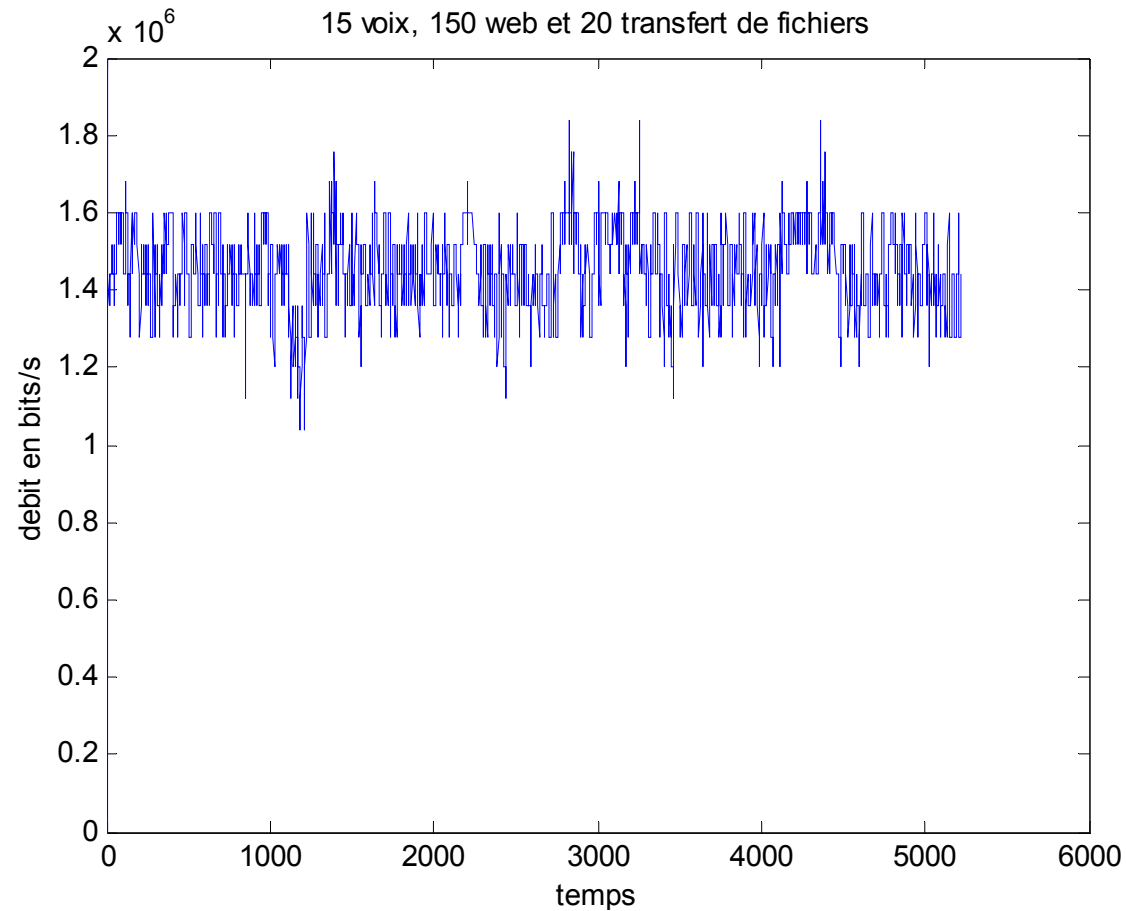
Source vidéo MPEG 4

- Existence de paliers où la taille des trames varie peu
- Notion de scène
- Notion de régime d'activité : 2 identifiés (haut, normal)
- Modélisation de la taille de trames (B, P) par des va de distribution log-normale de paramètres dépendant du régime d'activité
- Modèle : DBMAP, le temps est discret (base de temps de 40 ms), l'espace de phase dépend du régime d'activité et du type de trame
- La superposition de DBMAP est DBMAP mais l'espace de phase est multiplicatif
- Nouveau standard MPEG4 : sujet ouvert

Multiplexage des services hétérogènes



Multiplexage des services hétérogènes



Estimation des paramètres

- Processus MMPP : le générateur infinitésimal du processus de modulation et les intensités du Poisson dans chaque phase (chaîne de Markov cachée HMM)
- Choix du nombre de phases en tenant compte de la particularité de la couche MAC dans l'UMTS qui « lisse » les flux
- Méthode du maximum de vraisemblance (MLE pour HMM) : utiliser l'algorithme EM (maximiser l'espérance) itératif → nécessité d'un choix approprié de valeurs initiales pour accélérer la convergence

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

$$Q = \begin{pmatrix} -\sigma_1 & \sigma_1 \\ \sigma_2 & -\sigma_2 \end{pmatrix}$$

	λ_1		λ_2		σ_1		σ_2	
	Réelle	Estimé	Réelle	Estimé	Réelle	Estimé	Réelle	Estimé
Cas 1	100	101,48	10	9,48	10	10,09	1	1,24
Cas 2	10	10,17	1	0,94	0,5	0,42	0,2	0,22

Tableau 1 : Estimation des paramètres

Processus MMPP2

	λ_1	λ_2	σ_1	σ_2	Constant e
30 Voix	17129.09	44578.39	55.38	173.03	9.52e-5
40 Voix	22765.67	42850.33	58.97	166.59	6.89e-5
Multiplexage 15 voix 150 Web 20 ftp	37728.97	64594.60	37.314	135.58	8.69e-5

Tableau 2 : Valeurs des paramètres pour les différents scénarios.
Trafic à la sortie de la couche MAC de l'UMTS

Transport des flux dans le réseau :

Réseaux de files d'attente

- Réseaux ouverts : spécifiés par
 - Processus d'arrivées exogènes
 - Processus de service requis
 - Politique de service, taille des mémoires, nombre de serveurs
 - Règles de routage
- Réseaux fermés : spécifiés par
 - Processus de service requis
 - Politique de service, nombre fini de clients, nombre de serveurs
 - Règles de routage

Approximation Markovienne

- Dans une file GI/GI/1 le temps d'attente (donc de transit) moyen dépend de la variance et de la durée moyenne de service des processus d'inter-arrivée et du service requis (donc du coefficient de variation des lois associées)
- La loi exponentielle est intermédiaire entre celle d'Erlang (somme d'exponentielles) et celle de l'hyper-exponentielle (choix aléatoire d'exponentielle)
- Dans un certain nombre de situations l'approximation exponentielle donne des résultats pessimistes
- Il n'y a pas de théorie générale de réseaux de files d'attente
- Les équations de trafic (conservation des flux) permettent de calculer l'intensité du flux entrant dans chaque file d'attente

Réseaux Markoviens ouverts

- Le processus d'arrivées du monde extérieur : Poisson
- Le processus du service requis est exponentiel de paramètre qui dépend du nœud
- Le routage est indépendant de l'état du réseau
- Nombre de serveur quelconque
- Politique de service FIFO

Réseaux Markoviens fermés

- Pas d'arrivées extérieures
- Nombre de clients finis
- Le reste suit les mêmes spécification que dans le cas des réseaux ouverts

Réseaux à forme produit

- Les réseaux Markoviens ouverts / fermés sont à forme produit sous la condition de stabilité
- Problèmes numériques liés à la dimension de l'espace d'état
 - Algorithme de convolution pour déterminer la constante de normalisation dans le cas d'un réseau fermé
 - Décomposition hiérarchique