



HAL
open science

État des lieux sur la qualité de service dans les réseaux ad hoc

Claude Chaudet, Isabelle Guérin-Lassous

► **To cite this version:**

Claude Chaudet, Isabelle Guérin-Lassous. État des lieux sur la qualité de service dans les réseaux ad hoc. Colloque Francophone sur l'Ingénierie des Protocoles - CFIP 2006, Eric Fleury and Farouk Kamoun, Oct 2006, Tozeur/Tunisia, 24 p. inria-00111731

HAL Id: inria-00111731

<https://inria.hal.science/inria-00111731>

Submitted on 20 Nov 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

État des lieux sur la qualité de service dans les réseaux *ad hoc*

Claude Chaudet*, Isabelle Guérin Lassous**

* GET/ENST ; LTCI-UMR CNRS 5141 – 37-39, rue Darreau – 75014 Paris – France
Claude.Chaudet@enst.fr

** Inria Ares / Citi, Insa de Lyon – 69621 Villeurbanne Cedex – France
Isabelle.Guerin-Lassous@inrialpes.fr

RÉSUMÉ. Dans cet article, nous proposons un état des lieux sur la qualité de service dans les réseaux ad hoc telle que nous la percevons actuellement. Après une discussion sur les performances des réseaux ad hoc en termes de bande passante, délai et taux de pertes, nous apportons un point de vue critique sur certaines solutions de qualité de service proposées dans la littérature. Enfin, nous terminons cet article par une discussion sur les points durs et ouverts de ce domaine.

ABSTRACT. In this article, we give our point of view on quality of service for ad hoc networks. After a discussion on performances of ad hoc networks in terms of bandwidth, delay and loss rate, we describe and discuss some solutions proposed in the literature. Finally, we end this article with a discussion on the key points that are still open in this research area.

MOTS-CLÉS : réseaux ad hoc, qualité de service, discussion

KEYWORDS: ad hoc networks, quality of service, discussion

1. Introduction

Il n'est nul besoin de présenter à nouveau les réseaux *ad hoc*. Il s'agit de réseaux sans fil, sans infrastructure fixe, *a priori* mobiles dans lesquels la totalité de la gestion du réseau est distribuée. De nombreux problèmes surviennent dans ces réseaux du fait de ces caractéristiques. Certaines questions récurrentes sont du type : comment organiser l'espace d'adressage afin de réduire la complexité du routage ? Comment acheminer les trames ? Comment offrir des garanties aux applications ?

Le routage a fait, et fait toujours, l'objet de nombreuses publications. De nombreuses approches ont été examinées, de nombreux protocoles ont été proposés. L'IETF, par le biais du groupe de recherche MANET¹, a élu en 2003 quatre protocoles pour une normalisation éventuelle. Depuis, d'autres problématiques dans les réseaux *ad hoc* sont les sujets d'étude de certains organismes de normalisation. La qualité de service fait partie de ces problématiques. Par exemple, elle a été définie comme un des axes du groupe de recherche ANS-Research² de l'IRTF. Au niveau de l'IEEE, la publication récente de l'amendement IEEE 802.11e définit les mécanismes de niveau accès au médium permettant de définir plusieurs niveaux de priorités entre trames ainsi qu'un mode centralisé permettant de mettre en œuvre des garanties strictes. Ce standard n'est pas destiné *a priori* à une utilisation directe dans un réseau *ad hoc* multisaut, mais peut constituer une base pour la conception de mécanismes de qualité de service aux niveaux des couches supérieures.

Tout d'abord, on peut se demander pourquoi offrir des garanties aux applications. En effet, le succès du logiciel de téléphonie pair à pair Skype³ semble plaider en faveur d'un sur-dimensionnement du réseau et d'une optimisation des algorithmes de compression des flux multimédias. Toutefois, cette approche est antagoniste à une bonne utilisation des ressources du réseau. En effet, si l'on veut, sans mécanisme dédié, que les paramètres de performance, tels que la bande passante, le délai ou la gigue soient conformes aux pré-requis des applications, il est nécessaire, dans cette approche, que le réseau soit sous-utilisé. Par ailleurs, dans un environnement sans fil, la technique de sur-dimensionnement ne s'applique pas. En effet, la bande de fréquences utilisables pour une technologie est limitée. Le développement de nouveaux mécanismes au niveau physique ou accès au médium permet d'optimiser l'utilisation de cette bande de fréquence, mais la théorie de l'information prouve qu'il existe une limite. Par ailleurs, il est difficile d'effectuer plusieurs transmissions en parallèle dans la même bande de fréquences, en conséquence la bande passante est souvent partagée au sein d'une même zone géographique. Enfin, les performances des réseaux sans fil, même si elles sont en constante amélioration, restent en deçà des performances de leurs homologues filaires du fait des contraintes mentionnées ci-dessus. Or, les applications sont le plus souvent définies pour une utilisation dans un réseau filaire commuté. C'est pourquoi,

1. <http://www.ietf.org/html.charters/manet-charter.html>
2. <http://www.flarion.com/ans-research/>
3. <http://www.skype.com>

dans une optique de déploiement de réseaux sans fil à grande échelle tels que les réseaux *ad hoc*, la gestion d'une certaine qualité de service semble nécessaire.

Les applications présentant une contrainte liée aux performances du réseau sont en plein essor. Selon l'observatoire des marchés de l'ARCEP⁴, les abonnements de téléphonie en voix sur IP (VoIP) représentaient 9 % du volume total d'abonnements au 4e trimestre 2005 pour un volume de communications atteignant les 11 %⁵. En dehors de la téléphonie, de nombreux services de divertissement tels que les jeux massivement multijoueurs ou la diffusion de contenus multimédias (vidéo à la demande, radios en ligne, etc.) sont en plein essor. Le jeu en ligne le plus populaire aujourd'hui rassemblait plus de 5 millions d'abonnés à travers le monde à la fin de l'année 2005. Il est donc naturel de s'intéresser à la possibilité de mettre en œuvre ces applications dans un contexte mobile et *a fortiori* dans un contexte de réseau *ad hoc*. Par exemple, lors de l'organisation d'un événement, on pourrait vouloir profiter d'un service de téléphonie autonome, ne nécessitant ni abonnement ni câblage. Pour les applications vocales de type téléphonie sur IP, les algorithmes de compression usuels génèrent en général de petites trames (de 80 bits à 960 bits pour des débits de 8 kbit/s à 64 kbit/s). La qualité de la transmission est perçue comme très bonne pour un délai de bouche à oreille ne dépassant pas les 200 ms et un taux de pertes des trames inférieur à 15 %. Les applications de type jeux en ligne requièrent en général un débit assez faible, bien que plus élevées que celles de la transmission vocale. Un délai d'aller-retour en dessous de 200 ms ainsi qu'un taux de pertes inférieur à 15 % sont considérés comme très bon. Les contraintes en termes de débit des applications de vidéo temps réel sont intimement liées à la taille et à la qualité de la vidéo transmise, et donc du type de terminal utilisé pour la visualisation. Les débits de données peuvent aller de 64 kb/s à 10 Mb/s selon la qualité requise. Nous voyons bien ici que les qualités à offrir en termes de bande passante, délai ou taux de pertes sont différentes d'une application à une autre.

Dans cet article, nous nous intéressons essentiellement aux réseaux *ad hoc* construits au moyen de la norme IEEE 802.11 [IEE 97] et ses dérivées [IEE 99a, IEE 03, IEE 99b]. En effet, même si cette norme n'est pas très adaptée à un contexte multisaut, elle est toutefois représentative des approches aléatoires. Le grand nombre de matériels disponibles rend par ailleurs indispensable son évaluation. Les approches nécessitant une allocation de codes, d'intervalles de temps ou de fréquences semblent moins adaptées à un contexte *ad hoc* du fait du problème d'allocation des canaux peu aisé à réaliser dans un réseau distribué multisaut. Toutefois, certaines de ces approches (Bluetooth, WiMax) intègrent une gestion de la qualité de service qu'il faudra évaluer dans un tel contexte.

Cet article ne fait pas la promotion d'une solution. Il s'agit plutôt d'une étude sur la pertinence des solutions actuellement proposées. Nous commençons par examiner, avec la section 2, les performances que le standard IEEE 802.11 peut atteindre dans des réseaux à un saut et multisauts. Dans les sections 3, 4 et 5, nous examinons et commentons certaines solutions publiées récemment. Enfin dans la section 6, nous

4. <http://www.art-telecom.fr/observatoire/stat/index-statrim4-2005.htm>

5. Ces chiffres ne prennent pas en compte la téléphonie pair à pair (Skype).

discutons des challenges actuels qui restent encore des problèmes ouverts pour le domaine de la qualité de service dans les réseaux *ad hoc*.

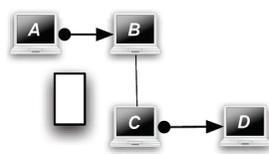
2. Performance attendues des réseaux *ad hoc*

La norme IEEE 802.11 et ses dérivées (IEEE 802.11b, IEEE 802.11a et IEEE 802.11g notamment) ont été très largement étudiées ces dernières années. Pour une description précise du protocole d'accès au médium, le lecteur est invité à se référer aux nombreux ouvrages existants à ce jour. De manière synthétique, lorsqu'un nœud souhaite accéder au médium partagé en utilisant la fonction de coordination distribuée (DCF), il commence par scruter le médium. Si celui-ci est libre, la trame est émise après un temps constant *DIFS* pendant lequel le médium doit rester libre. Dans le cas contraire (*i.e.* il reçoit un signal dont la puissance est supérieure à un seuil, appelé seuil de détection de porteuse), il attend la libération du médium, patiente un temps *DIFS* plus un temps aléatoire (*backoff*), multiple d'un temps unitaire défini par la norme, durant lesquels le médium doit rester libre. En cas de prise du médium par un autre émetteur, le processus est suspendu et reprendra à la valeur actuelle du compteur. Lorsque le compteur atteint une valeur nulle, la trame est émise. À la fin de la transmission, le récepteur patiente un temps *SIFS* constant avant de transmettre un acquittement à l'émetteur. En cas d'échec de la transmission, par exemple en cas de collision, le protocole ré-émet la trame en choisissant le *backoff* dans un intervalle (appelé fenêtre de contention) de taille double. Ce processus est répété jusqu'à la transmission avec succès de la trame ou jusqu'à un nombre maximal de retransmissions au-delà duquel la trame est détruite.

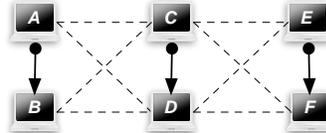
Dans un réseau radio multisaut basé sur la technologie sans fil 802.11 (mais aussi sur toute technologie sans fil utilisant un protocole d'accès de type CSMA/CA), le partage du médium n'est pas simple. En effet, pour qu'une station considère le médium comme libre, il faut qu'il n'y ait aucune transmission de station se trouvant dans sa zone de détection de porteuse. Cela implique que deux stations émettrices dans une même zone de détection de porteuse se partagent le médium radio (et cela implique aussi que deux stations à portée de communications se partagent le médium radio). Mais le partage du médium ne se fait pas seulement à l'émission. Il peut se faire aussi à la réception : deux flux dont les émetteurs sont indépendants peuvent entrer en collisions et dans ce cas, ils partagent aussi le médium radio puisqu'ils ne peuvent pas être émis en même temps. La nature multisaut des réseaux *ad hoc* complique ce partage et, selon les configurations, les flux obtiendront une proportion du médium différente les uns des autres. Les figures 1(a) et 1(b) donnent deux exemples de configurations multisauts (ces configurations sont expliquées en détail dans [CHA 05]). Dans la figure 1(b), la paire centrale est à portée de communication des deux paires extérieures alors que les deux paires extérieures sont indépendantes. Dans ce scénario, la paire centrale n'accède presque jamais au médium radio car les paires extérieures émettent de manière désynchronisée et occupent presque tout le temps le médium pour la paire centrale qui voit très peu le médium libre. Dans la figure 1(a) (présen-

tée dans [BHA 94]), les deux émetteurs ont en permanence un paquet à émettre, mais n'ont pas conscience de leur existence réciproque. *A* et *C* émettent des trames en permanence et alternent phases d'attente aléatoire et phases d'émission. Seul *B* peut souffrir de collisions et dans ce cas, seul *A* augmente la taille de sa fenêtre de contention. En conséquence, pour qu'une trame soit transmise avec succès de *A* à *B*, il est nécessaire, si le mécanisme RTS-CTS n'est pas activé, que la durée de sa transmission ne dépasse pas le temps de silence de *C*. Par conséquent, ces deux flux partagent le médium sans le savoir et le débit de *C* va être beaucoup plus élevé que celui de *A*. L'utilisation des RTS-CTS va augmenter légèrement le débit de *A*, mais n'aura néanmoins pas l'efficacité escomptée car l'asymétrie du scénario fait que *A* n'a pas conscience des paquets émis par *C* et que par conséquent les paquets subiront toujours beaucoup de collisions.

Nous allons maintenant regarder de plus près les paramètres débit, délai et taux de pertes.



(a) Scénario issu de [BHA 94] souffrant de collisions multiples au niveau de *B*.



(b) Scénario présentant une contention asymétrique entre émetteurs.

Figure 1. Scénarios multisauts asymétriques

2.1. Débit

Afin d'offrir des garanties sur le débit aux applications, il est nécessaire de comprendre et d'évaluer le mécanisme de partage du médium de la norme IEEE 802.11, afin de déterminer, ne serait-ce qu'au sein d'un même voisinage acheminant un certain trafic, si l'introduction d'un nouveau flux est possible ou non.

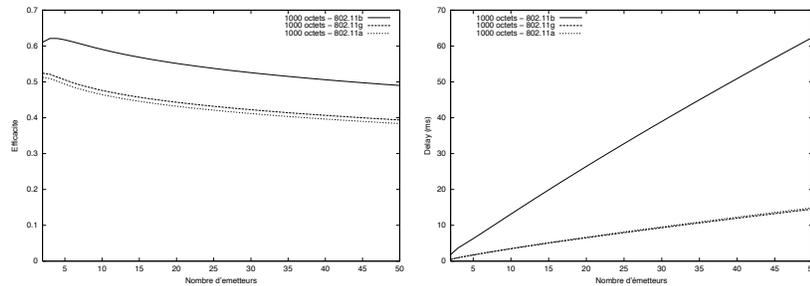
Entre un émetteur et un récepteur, à portée et sans trafic perturbateur, la transmission d'une trame de longueur P au débit physique D nécessite un temps $DIFS + \overline{backoff} + Th_{PHY} + (H + P)/D + SIFS + ACK$, où Th_{PHY} représente le temps nécessaire à la transmission des en-têtes physiques, $\overline{backoff}$ le backoff moyen, H la taille des en-têtes ajoutés par les différentes couches protocolaires et ACK étant le temps nécessaire à la transmission d'un acquittement. En considérant des trames de 2312 octets, taille maximale autorisée par le standard, et un débit physique de 11 Mb/s, l'efficacité du protocole ne dépasse pas 71 %, soit 7,88 Mb/s pour IEEE 802.11b en considérant le cas le plus optimiste (en-tête PLCP court, acquittement

transmis à 2 Mb/s). Pour IEEE 802.11g et IEEE 802.11a, cette valeur descend à 68 % (37 Mb/s) dans le cas de IEEE 802.11g (54 Mb/s, ERP-OFDM) et à 67 % (36 Mb/s) dans le cas d'IEEE 802.11a. Pour des trames de longueur plus typique d'une application VoIP (20 octets par exemple), cette efficacité chute à environ 2 % dans les 3 cas, ce qui milite en faveur de l'agrégation des trames.

L'un des premiers travaux visant à déterminer les performances de ce protocole a été réalisé par Bianchi dans [BIA 00]. Dans cet article, l'auteur évalue le débit que peut supporter un réseau mono-saut de n stations cherchant à émettre en permanence. L'auteur exprime la valeur du débit total d'un tel réseau, S en fonction de la probabilité P_{tr} qu'une station au moins transmette durant un intervalle de temps unitaire σ , de la probabilité de succès P_s d'une telle transmission, de la taille moyenne de la charge utile des trames $E[P]$, et des temps durant lesquels le médium est occupé par une transmission réussie (T_s) et par une collision (T_c) :

$$S = \frac{P_s \cdot P_{tr} \cdot E[P]}{(1 - P_{tr}) \cdot \sigma + P_{tr} \cdot P_s \cdot T_s + P_{tr} \cdot (1 - P_s) \cdot T_c}$$

Au moyen d'une chaîne de Markov en temps discret, l'auteur exprime P_{tr} en fonction de P_s , ainsi que P_s en fonction de P_{tr} , ramenant ainsi l'expression du débit total d'une cellule à la résolution d'un système non linéaire de deux équations à deux inconnues. L'article indique ces valeurs pour la version initiale de la norme IEEE 802.11. La figure 2(a) représente l'adaptation de ce résultat aux révisions récentes du standard (l'efficacité représentant le débit utile agrégé divisé par le débit physique). On peut constater qu'après un faible accroissement de la capacité dû à la décrémentation parallèle des compteurs de *backoff*, les performances décroissent avec le nombre d'émetteurs.



(a) Efficacité en termes de débit des différentes versions du protocole (b) Délai d'accès au médium dans un réseau à un saut

Dans un contexte multisaut, peu de résultats existent du fait de la complexité de la modélisation de la propagation radio et de la complexité du partage multi-saut. Dans l'idéal, la probabilité de transmission correcte d'une trame est fonction de l'ensemble des trames transmises en parallèle dans le réseau, de la configuration des lieux, du bruit ambiant, etc., et les valeurs des paramètres utilisés dans [BIA 00] (P_s , P_{tr} , etc.) vont être très différents d'un nœud à un autre. Cependant des bornes

existent. Dans [GUP 00], Gupta et Kumar montrent que pour un réseau dont la capacité du canal est de W bit/s, comportant n nœuds répartis aléatoirement dans un cercle, les couples émetteur-récepteur étant déterminés aléatoirement, le débit par nœud, $\lambda(n)$, dans le cas d'un ordonnancement optimal des trames, peut être borné par $\lambda(n) = \Theta\left(W/\sqrt{n \cdot \log(n)}\right)$. Lorsque le positionnement des nœuds est déterminé de façon optimale, cette borne devient $\Theta(W/\sqrt{n})$. Si cette étude montre qu'il est possible d'accroître les performances d'un réseau grâce à une certaine diversité spatiale, les études réelles dans des réseaux peu étendus montrent cependant que dans des cas pratiques, il n'en est rien.

Les études précédentes se positionnent soit dans des environnements aléatoires, soit dans des environnements extrêmes (saturation du médium). Toutefois, ces situations ne sont pas générales. En effet, de nombreux problèmes d'équité peuvent survenir dans les réseaux *ad hoc* du fait de la perception asymétrique de l'occupation du médium, comme nous l'avons vu précédemment. Ces problèmes surviennent au niveau microscopique et ne remettent pas en cause les résultats obtenus sur de très grands réseaux. Toutefois, lorsqu'on souhaite offrir des garanties à certains flux, il est nécessaire de s'assurer localement que chaque saut sera à même de satisfaire la demande jusqu'à la fin du transfert, ce qui au vu des résultats présentés ici est loin d'être trivial. Par conséquent, ces comportements microscopiques dans le partage de la bande passante sont très importants à comprendre et à évaluer pour une mise en place efficace de la qualité de service dans les réseaux *ad hoc*.

2.2. Délai

Le délai de transmission d'une trame lors d'un saut dans le réseau peut être décomposé en délai dû à l'application (encodage audio ou vidéo par exemple), délai dans la file d'attente et délai d'accès au médium. Le délai d'encodage est fonction du codec utilisé. Le délai d'accès au médium est un paramètre très difficile à calculer dans un réseau *ad hoc*. Il est fortement dépendant de la topologie et des flux environnants. La borne inférieure de ce délai, définie comme le temps entre la fin de la réception de cette trame et l'arrivée de la trame en tête de file d'attente, peut être calculée aisément en considérant les caractéristiques du protocole MAC. En revanche, la borne supérieure de ce délai est potentiellement infinie. En effet, la probabilité que le paquet subisse des collisions à répétition et soit supprimé par la couche MAC est non nulle.

Une approximation du délai moyen d'accès au médium dans un réseau à un saut fonctionnant à saturation peut toutefois être calculée en utilisant les résultats de l'article de Bianchi [BIA 00] (cf. paragraphe précédent). Plaçons nous dans le cas d'un réseau composé de n stations et considérons une station particulière. Lorsqu'elle souhaite émettre une trame, elle tire un *backoff* dans sa fenêtre de contention courante. À chaque intervalle de temps, elle décrémente cette valeur de 1 avec une probabilité $(1 - P_{tr})$, P_{tr} étant calculé sur les $n - 1$ autres émetteurs potentiels. Avec une probabilité P_{tr} , elle suspend le processus de décrémentation pour une durée égale

au temps d'occupation du médium. Ce temps sera égal à T_s avec une probabilité P_s et égal à T_c avec une probabilité $(1 - P_s)$. En prenant en considération ce comportement, on peut exprimer le temps moyen avant lequel une station aura terminé la décrémentation d'un *backoff* de valeur b pour arriver à un *backoff* nul : $\delta(b) = b \cdot ((1 - P_{tr}) \cdot \sigma + P_{tr} \cdot ((1 - P_s) \cdot T_c + P_s \cdot T_s))$. Le tirage du *backoff* étant uniforme dans la fenêtre de contention, le temps moyen avant d'arriver à un *backoff* nul pour une fenêtre de contention de taille W peut s'exprimer par : $\beta(W) = \sum_{b=1}^{W-1} \frac{\delta(b)}{W} = \frac{W-1}{2} \cdot ((1 - P_{tr}) \cdot \sigma + P_{tr} \cdot ((1 - P_s) \cdot T_c + P_s \cdot T_s))$.

Lorsque le *backoff* atteint une valeur nulle, la station émet sa trame et souffre d'une collision avec une probabilité P_{tr} . Le cas échéant, elle recommence le processus avec une fenêtre de contention de taille double. La valeur moyenne du délai avant que la trame ne soit transmise ou supprimée à cause d'un trop grand nombre de retransmissions peut donc s'exprimer par (où CW_{min} est la fenêtre de contention initiale) :

$$E[\text{délai}] = \beta(CW_{min}) + (1 - P_{tr}) \cdot T_s + P_{tr} \cdot (T_c + \beta(2 \cdot CW_{min}) + (1 - P_{tr}) \cdot T_s + P_{tr} \cdot (\dots))$$

En notant m l'étage maximal de *backoff* ($CW_{max} = 2^m \cdot CW_{min}$), on peut exprimer ce délai par :

$$E[\text{délai}] = \beta(CW_{min}) + \sum_{i=1}^m P_{tr}^i \cdot \beta(2^i \cdot CW_{min}) + P_{tr}^{i-1} (1 - P_{tr}) \cdot T_s + P_{tr}^i \cdot T_c$$

La trame est alors perdue avec une probabilité P_{tr}^m . La figure 2(b) représente ce délai en fonction du nombre d'émetteurs pour différentes valeurs de tailles de paquets et dans le cas de différentes technologies. La distribution de ce délai peut, elle aussi, être déterminée au moyen des mêmes résultats.

Dès lors que le délai d'accès au médium est connu, l'expression de ce délai d'accès peut être utilisé dans une file d'attente de type M/G/1/K ou G/G/1/K, selon le profil de trafic, en adaptant le taux d'inter-arrivées de manière à ce que la file ne soit jamais vide afin de respecter les hypothèses de [BIA 00]. Le médium étant partagé il peut être considéré comme un serveur unique. Il est alors possible d'en déduire la longueur moyenne de la file d'attente et d'estimer le délai de transmission sur un lien.

Si cette analyse s'applique dans le cas d'un réseau à un saut, il n'en est toutefois pas de même dans le cas d'un réseau *ad hoc*, du fait des dissymétries pouvant exister dans l'accès au médium, ce qui complique la tâche de l'évaluation du délai dans un tel réseau. En effet, chaque station doit être en mesure d'évaluer le nombre de stations avec lesquelles elle partage le médium, or cette estimation est difficile à réaliser dès que le partage se fait dans la zone de détection de porteuse ou sur des configurations de stations cachées asymétriques comme celle de la figure 1(a). De plus, chaque station aura des paramètres, comme par exemple P_{tr} ou T_s , propres. Toutefois, cette analyse fournit un ordre de grandeur pour le délai d'accès au médium dans un réseau basé sur

la norme IEEE 802.11. Si avec cet ordre de grandeur, il est possible d'évaluer le délai de bout en bout subi par une trame, ce résultat peut être utilisé afin d'évaluer s'il est possible de lancer une application sensible au délai (VoIP, jeux) dans le réseau. Il faut néanmoins noter que le délai passé dans les files d'attente des nœuds n'est pas ici pris en compte. Or ce délai semble très difficile à modéliser dans un contexte multisaut, car il dépend d'une part des profils de trafic des applications au niveau des sources mais aussi de l'évolution des différents paquets dans tout le réseau. Par conséquent, la détermination des profils d'évolution des files d'attente sur les différents nœuds semble loin d'être triviale.

2.3. Taux de pertes

Deux facteurs principaux peuvent engendrer la perte d'une trame. Le premier concerne la présence de bruit ou d'interférences sur le canal et le deuxième concerne l'occurrence successive de collisions conduisant à un dépassement du nombre de transmissions maximal de la trame. Dans le cas d'une émission de paquets en mode diffusion, seul le premier facteur aura une influence sur le taux de pertes puisque ces paquets ne sont pas acquittés et donc jamais ré-émis. Dans le cas d'une émission de paquets en mode point-à-point, le premier impact a un impact sur la bonne réception ou non d'un paquet, mais c'est seulement la combinaison de ces deux facteurs qui indique si un paquet est rejeté ou non. Le premier facteur est essentiellement dépendant de la modulation utilisée et du niveau de bruit sur le canal, tandis que le deuxième facteur a, comme nous l'avons vu précédemment, une influence sur le délai et peut être évalué dans le cas d'un réseau à un saut, en utilisant les formules de Bianchi.

Plusieurs études évaluent la probabilité d'erreur sur un paquet pour différentes modulations et en considérant différents modèles de canal (cf. la thèse de Manshaei pour une étude détaillée sur le sujet [MAN 05]). Par exemple, pour les modulations OFDM 64-QAM (IEEE 802.11a et IEEE 802.11g en mode ERP-OFDM) et dans l'hypothèse d'un canal AWGN (bruit gaussien additif), le taux d'erreur bit en fonction du rapport E_b/N_0 est le suivant :

$$BER_{OFDM,64-QAM} = \frac{7}{12} \cdot Q\left(\sqrt{\frac{2 \cdot E_b}{7 \cdot N_0}}\right) + \frac{1}{2} \cdot Q\left(3 \cdot \sqrt{\frac{2 \cdot E_b}{7 \cdot N_0}}\right)$$

$$Q \text{ étant la fonction de Marcum : } Q(x) = \frac{1}{\pi} \int_0^{\pi/2} \exp\left(-\frac{x^2}{2 \cdot \sin^2 \theta}\right) d\theta.$$

À partir de l'expression du taux d'erreur-bit et en considérant que la norme IEEE 802.11 n'intègre aucun mécanisme de correction d'erreur, il est possible de déterminer la probabilité qu'un paquet contienne une erreur et d'en déduire la probabilité qu'un paquet soit perdu après le nombre de retransmissions prévu par le standard. Néanmoins, une telle évaluation repose sur le calcul du taux d'erreur bit qui nécessite de connaître *a priori* le rapport signal sur bruit, ce qui ne semble pas évident et sur le calcul de P_{tr}^m décrit précédemment qui n'est pas direct dans un contexte multisaut.

2.4. Résumé pour la qualité de service

Les caractéristiques propres aux réseaux *ad hoc* font qu'il est difficile de mettre en œuvre une certaine qualité de service pour ces réseaux. Premièrement, la nature spécifique du médium radio implique des taux de pertes importants ainsi qu'une variation importante de la capacité du canal radio. De telles propriétés entravent la fiabilité des protocoles qui reposent sur des échanges explicites de messages de contrôle. De plus, beaucoup de cartes sans fil actuelles adaptent leur débit physique d'émission en fonction de la capacité du canal rencontré. Par conséquent, cette capacité peut être très dynamique. Les solutions de qualité de service peuvent donc difficilement reposer sur l'hypothèse d'une capacité de canal radio fixe et connue.

Deuxièmement, la nature multisaut des réseaux *ad hoc* complique grandement le partage du médium radio. L'émission d'un flux peut donc avoir un impact sur tous les autres flux avec lesquels il partage le médium radio, et ce sans que l'un ou l'autre des flux en ait conscience. Or, lors de l'examen de l'introduction d'un nouveau flux, il est nécessaire de vérifier si le réseau dispose d'une capacité résiduelle suffisante pour acheminer le flux, et ce sans provoquer de violation de garanties précédemment accordées. Pour réaliser cette opération il serait en théorie nécessaire de connaître le niveau de bruit généré par chaque émetteur au niveau de chaque nœud, et de déterminer précisément quelles trames sont émises en parallèle. Une telle connaissance est bien évidemment impossible à obtenir, et comme nous venons de le voir, les formules (relativement) simples qu'on peut établir sur un réseau à un saut ne peuvent être utilisées directement. Les solutions proposées se doivent donc d'évaluer au mieux ces contraintes tout en requérant un minimum de communications. En effet, plus les nœuds s'échangeront des informations précises à des distances et des fréquences élevées, plus l'évaluation sera précise, mais plus le processus sera coûteux. Le bon compromis, s'il existe, reste encore à déterminer.

Troisièmement, la mobilité possible des réseaux *ad hoc* peut avoir un fort impact sur la qualité de service offerte. Une telle mobilité peut impliquer une violation des garanties de qualité de service qui étaient valables au départ. Comment réagir à une telle mobilité est toujours un challenge pour lequel peu de solutions apportent des réponses satisfaisantes.

3. Solutions actuelles

Il est possible de classifier les protocoles visant à offrir une certaine qualité de service selon plusieurs critères. La classification de Wu et Harms, souvent citée [WU 01], sépare les solutions de qualité de service en quatre grandes classes : les modèles de qualité de service qui définissent des architectures, les protocoles de signalisation qui définissent les messages nécessaires à la mise en place et au maintien de la qualité de service, les protocoles de routage avec qualité de service dont l'objectif est de déterminer des routes satisfaisant un ou plusieurs critères de QoS et les protocoles de QoS

au niveau MAC dont l'objectif est d'introduire des priorités au niveau de l'accès au médium.

Cette classification permet d'identifier les composants importants en termes de qualité de service. Néanmoins, chaque solution dans cette classification, considérée isolément, est souvent peu complète, et une solution efficace pour les réseaux *ad hoc* semble plutôt être une combinaison de certains de ces différents composants. Par exemple, pourquoi dissocier le routage avec QoS de la signalisation ? Très souvent, de tels protocoles de routage ont leur propre signalisation qui est très dépendante du routage proposé. Il semble aussi très probable de pouvoir offrir de meilleures garanties si une optimisation conjointe entre le routage et le protocole MAC est réalisée. De très nombreuses solutions de qualité de service pour les réseaux *ad hoc* sont d'ailleurs difficiles à ranger dans cette classification.

Nous avons choisi d'adopter, dans la suite de cet article, la classification usuelle séparant solutions offrant des garanties statistiques (type *DiffServ*) des approches cherchant à offrir des garanties quantitatives (type *IntServ*). Plus précisément, il serait possible de placer chacune des solutions proposées sur un axe de «qualité» des garanties allant de garanties totalement statistiques (appartenance des paquets à une classe de trafic, ce qui correspond plus ou moins à différents niveaux de *best effort*) à des garanties très strictes (aussitôt que la route ne peut plus assurer la QoS demandée, une nouvelle route est empruntée si elle existe). Entre ces deux extrêmes, tous les niveaux sont envisageables.

Dans toutes les solutions, on peut distinguer deux aspects : l'aspect architectural qui concerne essentiellement l'identification des acteurs et de leurs rôles (les nœuds impliqués, les couches OSI concernées) et l'aspect «réalisation» qui va permettre de mettre en œuvre la qualité de service. Pour schématiser, le premier aspect correspond à «qui fait quoi» alors que le deuxième correspond à «comment on fait le quoi».

Par la suite, nous allons décrire quelques-unes des solutions existantes. Nous ne faisons pas un survol de toutes ou d'une grande partie des propositions (et elles sont nombreuses), mais nous préférons nous concentrer sur quelques protocoles. Le choix des solutions est toujours un acte difficile. Nous avons privilégié des solutions de références qui sont citées dans presque tous les articles qui traitent de la qualité de service dans les réseaux *ad hoc*, des solutions qui présentent des avancées significatives dans le domaine et dont nous pensons qu'elles deviendront des références, des solutions isolées qui traitent d'un sujet souvent peu abordé ou encore des solutions que nous avons beaucoup étudiées dans nos laboratoires. Ce choix reste bien sûr subjectif, mais nous pensons avoir un échantillon assez représentatif⁶ pour que tout lecteur qui se soit déjà intéressé à ce sujet connaisse la plupart des solutions présentées ici.

6. La plupart des solutions ont d'ailleurs un indice de citation assez élevé sur Google Scholar. Même si cet outil n'est pas parfait, il donne une idée de l'activité autour d'un papier.

4. Qualité de service statistique

La qualité de service statistique pourrait être une approche privilégiée car elle est plus facile à mettre en œuvre et souffre moins des problèmes de passage à l'échelle comparé aux solutions qui cherchent à apporter des garanties. Ce n'est pourtant pas l'approche privilégiée dans le monde des réseaux *ad hoc*. Une telle approche est assez simple à mettre en place et ne requiert pas un volume de signalisation important. En revanche, le report des politiques de voisinage en voisinage est difficile, ce qui rend complexe la mise en place de politiques de priorités au niveau global.

4.1. IEEE 802.11e

La révision *e* du standard IEEE 802.11 regroupe en une norme un ensemble de mécanismes permettant d'affecter des priorités lors de l'accès au médium aux différents paquets émis dans le réseau. Dans ce dessein, il est possible de jouer sur un ensemble de paramètres tels que le temps d'attente fixe (*DIFS*, renommé en *AIFS*), les valeurs minimale et maximale de la fenêtre de contention ainsi que la durée allouée à un émetteur gagnant la contention. Différentes valeurs de ces paramètres permettent de faciliter l'accès au médium pour les classes de trafic prioritaires. Le standard définit 8 classes de trafic.

Cet amendement définit assez précisément les mécanismes de qualité de service pouvant être mis en œuvre dans un réseau administré par une station de base, mais l'utilisation dans un contexte *ad hoc* multisaut souffre des mêmes avantages et limitations que la DCF de la norme initiale. Selon la topologie du réseau, il est possible qu'un flux de priorité faible obtienne l'accès au médium plus souvent qu'un flux temps-réel, comme dans le scénario de la figure 1(a) par exemple. Si le flux de *C* vers *D* a une priorité faible, il est fort probable que les temps inter-trames soient malgré tout trop courts en moyenne pour permettre l'émission d'une trame de *A* à *B*. Dans le cas contraire, l'efficacité globale du protocole se trouverait réduite puisque le surcoût du protocole augmenterait. A contrario, il est possible de tirer parti de cette différenciation pour résoudre de tels problèmes liés à la topologie du réseau en adaptant dynamiquement les durées allouées à chaque émetteur et à chaque classe par exemple. Il est cependant nécessaire pour cela de déterminer quelles informations doivent être connues, comment et à quelle distance les transmettre.

4.2. Blackburst

Sobrinho et Krishnakumar proposent dans [SOB 96] une approche inverse à 802.11e, plus proche du fonctionnement d'Hiperlan 1. Les stations de priorité faible accèdent au canal en utilisant la norme IEEE 802.11. Une station souhaitant émettre un paquet appartenant à un flux prioritaire brouillera le canal dès que celui-ci devient libre, et ce pendant un temps proportionnel au temps depuis lequel elle attend. Ce brouillage bloque le mécanisme de détection de porteuse ainsi que le *backoff* des

autres stations. La station la plus prioritaire brouillera le canal le plus longtemps et pourra y accéder librement dès la fin de l'émission de bruit. En choisissant correctement les temporisateurs du protocole, les auteurs montrent qu'il est possible d'éviter les collisions. Cependant, ce mécanisme souffre aussi des mêmes limitations que les approches de type DCF dans un contexte multisaut. Lorsque la perception de l'état du médium est asymétrique, il est possible qu'un nœud de faible priorité monopolise le canal au détriment d'un nœud de forte priorité. En considérant une fois encore le scénario de la figure 1(a), les émetteurs étant ignorants de leur présence mutuelle, le signal de brouillage ne permettra pas simplement de rendre le flux de *A* à *B* prioritaire. Enfin, les auteurs de [LIN 03] montrent que le protocole peut engendrer un problème de famine

4.3. AEDCF

L'idée d'AEDCF ([ROM 03]) est de modifier l'algorithme BEB utilisé dans 802.11e afin de l'adapter à chacune des classes. Après une collision, la taille de la fenêtre de contention est augmentée en fonction d'un facteur qui dépend de la priorité des classes et qui assure que les flux qui sont entrés en collision garderont la même priorité. Après une transmission réussie, la taille de la fenêtre de contention est diminuée d'un facteur dépendant du taux moyen de collisions subies ainsi que de la priorité de la file afin de préserver l'ordre des priorités dans l'accès au médium tout en limitant les collisions consécutives.

Si une telle approche présente de bonnes performances sur des réseaux à un saut ou en présence de stations cachées, elle reste inefficace sur certains scénarios simples comme celui présenté sur la figure 1(a). Comme indiqué plus haut, l'adaptation du temps d'attente aléatoire, qu'elle soit faite en agissant sur la fenêtre de contention ou sur l'évolution du *backoff*, n'est pas suffisante pour résoudre ce problème tout en garantissant une bonne utilisation des ressources.

5. Garanties quantitatives

C'est l'approche que l'on retrouve le plus couramment dans le monde des réseaux *ad hoc*. Il est possible de diviser cette catégorie en deux sous approches. La première concerne les solutions qui cherchent à construire des routes qui respecteront les garanties. À cause de la mobilité potentielle, il est possible que certaines routes ne soient plus en mesure de fournir les garanties initiales. S'il y a dégradation, certaines solutions repartent de zéro, en cherchant de nouvelles routes par exemple, tandis que d'autres vont essayer de retrouver certaines garanties en pénalisant certains flux. La deuxième sous approche ne se base pas sur le routage. L'argument avancé par ces solutions est le suivant : puisque le réseau est mobile, il est inutile de tenter de construire une route offrant certaines garanties puisqu'il est très probable que celles-ci seront violées dans le temps ; il vaut donc mieux considérer les routes telles qu'elles sont et

modifier en continu l'environnement pour tenter d'offrir des garanties aux flux qui en ont besoin.

Les avantages des solutions de cette approche sont d'une part, une bonne connaissance des ressources allouées à chaque application, ce qui permet d'adapter le débit du flux de données (*layered video*, par ex.) sans retour d'information nécessaire, et d'autre part la possibilité d'appliquer des techniques d'optimisation au niveau du réseau, flux par flux. Les inconvénients de l'approche sont un passage à l'échelle difficile⁷, des problèmes d'évaluation précise des ressources disponibles et la difficulté de gérer la mobilité.

5.1. Approche a posteriori

Nous allons présenter deux protocoles qui ne se basent pas sur un routage avec qualité de service spécifique, mais qui régulent le débit de certains flux pour pouvoir offrir certaines garanties à d'autres.

SWAN fait partie de ces protocoles mythiques cités dans presque tous les articles traitant de la qualité de service dans les réseaux *ad hoc* [AHN 02]. La philosophie du protocole est de ne pas garder d'état dans les nœuds. SWAN assure une différenciation de service entre le trafic temps réel et le trafic *Best Effort*. Néanmoins un contrôle d'admission est réalisé pour le trafic temps réel afin d'assurer certaines garanties. Ce contrôle d'admission se base sur le calcul de la bande passante disponible sur le chemin qu'empruntera le flux (ce chemin étant déterminé par un protocole de routage indépendant de SWAN). Or ce calcul présente deux inconvénients : la bande passante disponible au niveau de chaque nœud est déduite du trafic temps réel perçu par chaque nœud et ne prend donc pas en compte la bande passante consommée au niveau de la zone de détection de porteuse. Par ailleurs, le calcul de la bande passante disponible sur le chemin ne prend pas en compte le fait que les paquets qui traversent un lien consomment aussi de la bande passante sur certains autres liens du chemin. Par conséquent, SWAN effectue une sur-estimation de la bande passante disponible.

La régulation du trafic *Best Effort* est effectuée grâce à un lissage de type AIMD : tant que le délai n'a pas atteint une valeur seuil, le débit est augmenté de manière additive ; lorsque la valeur seuil est atteinte ou dépassée, le débit est diminué de manière multiplicative. Cette valeur seuil est importante. Selon les configurations considérées et la géographie des flux *Best Effort*, ces derniers pourront avoir des impacts différents sur les débits des trafics temps réel. Par exemple, un scénario où un flux *Best Effort* est à portée de communication d'un flux temps réel n'aura pas le même impact qu'un scénario, tel que celui de la figure 1(a), où le flux best effort (de *C* vers *D*) va entrer en collision avec le flux temps réel (de *A* vers *B*). Par conséquent, une bonne utilisation des ressources implique des valeurs seuils différentes sur les nœuds selon les configurations. Or cet aspect n'est pas discuté dans cet article.

7. Toutefois, si cette critique est souvent formulée, il en est de même pour la majorité des solutions *ad hoc*, ne serait-ce que pour le routage...

En cas de détection d'une congestion, le nœud concerné marque les bits ECN dans l'en-tête de ses paquets de données temps réel. La destination qui reçoit de tels paquets notifie la source pour qu'elle ré-établit, éventuellement, la session temps réel concernée. Plusieurs solutions sont proposées pour le choix des flux à ré-établir. L'idée est d'introduire un mécanisme aléatoire dans le choix des flux à stopper afin de ne pas arrêter trop de flux en même temps dans les endroits congestionnés. Si cette approche est assez simple, le fait d'être indépendant d'un protocole de routage avec qualité de service peut conduire à un rejet de certains flux, puisque les routes reconstruites seront très probablement les mêmes qu'avant la congestion et que par conséquent il ne sera pas possible d'acheminer sur ces mêmes routes tous les flux qui ont provoqué précédemment une congestion.

QPART se situe dans le même état d'esprit que SWAN [YAN 04] : différenciation de services entre différents types de trafic et garanties pour certains trafics sans garder d'états dans les nœuds. QPART considère trois types de trafic : le trafic sensible au délai, le trafic sensible à la bande passante et le trafic *Best Effort*. Garantir les contraintes et assurer la différenciation sont réalisés par des modifications des tailles des fenêtres de contention pour les différents types de trafic. En ce qui concerne les trafics sensibles au délai, la taille de la fenêtre de contention sur chaque nœud dépend du délai que le paquet vient de connaître sur le lien précédent et de celui qu'il aurait dû avoir. L'inconvénient de la solution est que chaque lien est sensé avoir un délai similaire, ce qui n'est pas une hypothèse réaliste. De plus, le traitement sur le délai ne prend aucunement en compte l'impact qu'il peut avoir sur les délais des autres flux existants dans le réseau. Pour les trafics sensibles à la bande passante, la modification de la fenêtre de contention est fonction de la taille de la file d'attente dédiée à ce flux. Cette file doit rester de taille constante pour garantir le débit du flux. Cela implique donc que si un flux subit beaucoup de collisions, sa file d'attente va grossir et il va donc diminuer sa fenêtre de contention pour réduire la taille de sa file, ce qui peut générer encore plus de collisions. Une telle approche peut être consommatrice de ressources inutilement. Pour les flux *Best Effort*, la taille de la fenêtre de contention est régulée en fonction d'un seuil de congestion. Ce seuil correspond à un temps libre moyen entre deux périodes d'occupation du médium. Si le temps libre moyen détecté est inférieur à ce seuil, alors la station considère que le médium est congestionné et va donc augmenter sa fenêtre de contention ; si en revanche le temps libre moyen détecté est supérieur à ce seuil, alors la station considère qu'il reste de la bande passante de libre et qu'elle peut donc émettre plus de trafic *Best Effort* en diminuant sa fenêtre de contention. Avec ce type d'approche, la même remarque que pour SWAN peut être faite : ce seuil est très dépendant de la configuration et ne peut être choisi de manière fixe comme c'est le cas dans QPART.

QPART n'effectue ni routage avec contraintes ni contrôle d'admission. QPART n'agit donc que sur la résolution de conflits, lorsqu'une congestion apparaît, en arrêtant certains flux. Le choix de ces flux se fait en fonction de leur priorité qui correspond à l'âge d'un flux : plus le flux est vieux, plus il est prioritaire (la mise à jour est réalisée périodiquement), jusqu'à atteindre une priorité maximale. Chaque priorité correspond à un seuil d'admission qui lui-même correspond à un temps libre moyen entre deux

périodes d'occupation du médium. Lorsqu'une station détecte un temps libre moyen inférieur à ce seuil, alors le flux est arrêté pendant un certain temps. Si au bout de ce temps, le temps libre moyen perçu pas la station est toujours inférieur à ce seuil alors le flux est définitivement rejeté, sinon il est redémarré. Les choix de ces paramètres et de ces seuils sont bien sûr les points clés de ce type d'approche. Or ils sont souvent peu discutés et quand les auteurs de QPART soulignent que l'ordre des priorités est préservé lors de l'arrêt des flux, c'est sans compter sur nos scénarios multisaits préférés. Reprenons l'exemple de la figure 1(a) et supposons que le flux de *A* soit le plus âgé et donc plus prioritaire que le flux de *C*. Pourtant, avec un protocole comme QPART, le flux de *A*, qui va subir beaucoup de collisions, va diminuer sa fenêtre de contention pour tenter de maintenir sa file d'attente constante et va donc atteindre très vite un petit temps libre moyen entre deux périodes d'occupation du médium. Pendant ce temps, tout se passe bien pour le flux de *C* qui ne détecte aucune congestion. Par conséquent c'est le flux de *A* qui sera arrêté avant le flux de *C* alors que ce premier était prioritaire.

5.2. Routage avec qualité de service

De très nombreux protocoles rentrent dans cette section. Nous n'allons pas tous les lister car ils sont trop nombreux. De plus, beaucoup de ces protocoles se concentrent surtout sur la partie routage, ce qui n'est pas très intéressant *in fine*, puisqu'on retrouve les mêmes problématiques que pour le routage classique et que ces aspects commencent à être bien connus. De notre point de vue, une partie vraiment difficile dans le routage avec qualité de service est l'évaluation des ressources. Or trop d'articles omettent cette partie en supposant connues les ressources disponibles. Nous avons donc choisi ici de décrire en détail une solution qui explique avec soin les mécanismes d'évaluation de ressources.

CACP est un protocole de routage couplé à un mécanisme de contrôle d'admission cherchant à garantir la bande passante des flux [YAN 05]. Le routage utilisé est un routage à la source réactif dont la phase de construction de route cherche à déterminer une route possédant suffisamment de bande passante. L'approche réactive est relativement naturelle dès qu'on cherche à offrir des garanties, car la phase d'inondation pour la construction de route permet d'explorer le réseau afin d'établir des routes candidates. Nous n'allons pas nous appesantir sur ce routage qui est naturel.

La partie contrôle d'admission est beaucoup plus intéressante. Les auteurs mettent en avant le fait que le partage du médium ne se fait pas seulement dans la zone de communication mais aussi dans la zone de détection de porteuse, et qu'il est donc fondamental de prendre en compte la bande passante consommée (ou bande passante libre) dans la zone de détection de porteuse contrairement à ce que font la plupart des

protocoles proposés sur le sujet⁸. Pour estimer sa bande passante de libre, un nœud écoute le médium pendant un intervalle de temps et considère qu'il est occupé dès qu'il reçoit un signal de puissance supérieure à son seuil de détection de porteuse. Ainsi il peut en déduire la proportion de temps libre pendant cet intervalle de temps, puis le temps libre moyen grâce à l'utilisation d'une fenêtre glissante. Le temps libre calculé tient bien compte de l'utilisation du médium dans la zone de détection de porteuse. Ensuite, pour savoir si une station est en mesure de transmettre des paquets à un certain débit, il lui faut aussi connaître la bande passante libre de chacun de ses voisins se trouvant dans sa zone de détection de porteuse. En effet, même si elle a suffisamment de bande passante, son émission peut provoquer des dépassements de capacité sur certains autres nœuds de sa zone de détection de porteuse et ainsi dégrader des flux existants et acceptés préalablement. Pour savoir si une station peut émettre à un certain débit, elle doit donc vérifier qu'elle est en mesure d'émettre à ce débit (i.e. elle a suffisamment de bande passante libre) et qu'elle ne va pas saturer les nœuds avec lesquels elle partage le médium radio, (i.e. les nœuds se trouvant dans sa zone de détection de porteuse). Elle doit donc acquérir toutes ces informations sur la bande passante libre. Le problème, pour cette acquisition, est que les stations en zone de détection de porteuse ne sont pas nécessairement à portée de communication et dans ce cas ne sont pas en mesure d'échanger ces informations. CACP propose trois méthodes pour essayer de réaliser cette acquisition. La première approche repose sur le fait (validé d'ailleurs expérimentalement dans d'autres articles) que les nœuds à deux sauts d'une station se trouvent dans la zone de détection de porteuse de cette station⁹ pour un débit physique à 2Mbps. Chaque nœud va donc envoyer à deux sauts de lui sa quantité de bande passante libre. L'inconvénient d'une telle approche est que la zone de communication à deux sauts n'est qu'une approximation de la zone de détection de porteuse et qu'il ne sera pas possible de récupérer l'information de certains nœuds de la zone de détection de porteuse qui n'ont pas de voisin à un saut pour relayer l'information. Certaines informations peuvent donc manquer. La deuxième approche propose d'augmenter la puissance lors de l'envoi de l'information sur la bande passante afin de toucher tous les nœuds de sa zone de détection de porteuse. L'inconvénient de cette approche est que les envois de ces messages à puissance plus élevée vont générer plus d'interférences sur le réseau. La troisième méthode consiste à diminuer le seuil de détection de porteuse afin que chaque station puisse écouter toutes les transmissions qui vont pénaliser les nœuds de sa zone de détection de porteuse et ainsi déterminer la bande passante disponible sur toute la zone de détection de porteuse. Les auteurs affirment qu'une telle approche fournit une sous-estimation de la bande passante disponible, or ce n'est pas complètement vrai. Considérons l'exemple représenté en figure 2 donné dans l'article présentant CACP. Supposons, comme dans l'article, que la capacité du médium est 2 Mb/s, que F envoie 1 Mb/s de données à E et que C envoie 1 Mb/s de données à B . Comme le disent les auteurs, avec cette troisième méthode, la station

8. Les auteurs oublient néanmoins de citer certains papiers comme [CHA 02] qui mettait déjà en lumière ce phénomène. Mais une polémique sur cette paternité est hors de propos dans cet article...

9. Cette approche était déjà décrite dans [CHA 02].

centrale A peut penser que la bande passante disponible est inférieure à 1 Mb/s si ces deux communications ne se font pas toujours en même temps, alors qu'en fait cette bande passante disponible est bien de 1 Mb/s. Or contrairement à ce qu'indiquent les auteurs le recouvrement ou non de ces communications a une importance. Imaginons qu'elles soient parfaitement alternées et que A cherche à envoyer des données à un récepteur R se trouvant dans sa zone de communication. D'après les auteurs, A pourrait envoyer ses données à 1 Mb/s. Or dès que A va émettre, ses données vont soit entrer en collision en E ou en B puisqu'un des deux est forcément en train de recevoir dans notre hypothèse. Par conséquent l'émission de A va forcément dégrader les débits de flux déjà acceptés.

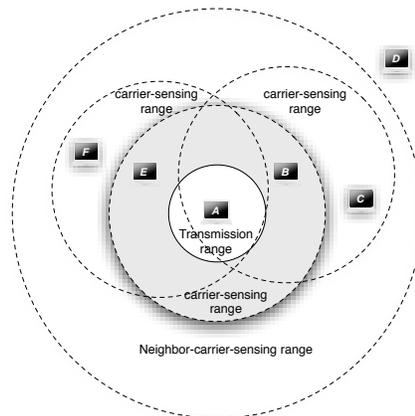


Figure 2. Scénario issu de [YAN 05].

6. Discussion

Une discussion sur la qualité de service dans les réseaux *ad hoc* a déjà été initiée dans la section précédente lors des descriptions de certains protocoles. Nous allons ici discuter un peu plus longuement sur les points durs qui nous semblent importants et qui restent, de notre point de vue, encore à travailler. Nous pensons qu'il reste encore des progrès à réaliser au niveau de l'évaluation des ressources consommées et disponibles, de la gestion de la mobilité et d'une gestion efficace des ressources.

6.1. Évaluation des ressources

Contrairement à ce que beaucoup de papiers semblent indiquer, le point dur n'est pas de savoir quels messages échanger. Quand il y a une information à transmettre (demande de réservation, cassure de route, etc.), on est bien obligé de communiquer.

On peut épargner légèrement les ressources en faisant de la signalisation en bande, mais le point difficile reste, de notre point de vue, comment évaluer les ressources et l'impact de l'introduction d'un nouveau flux sur les ressources consommées et libres. En effet, il ne suffit pas de savoir quelles sont les ressources libres pour le flux qu'on veut émettre, mais il faut aussi être en mesure d'évaluer l'impact qu'aura l'introduction de ce nouveau flux sur les flux déjà existants dans le réseau. Or comme nous l'avons vu en filigrane tout au long de cet article cet impact peut être très important dans les réseaux à médium partagé comme les réseaux *ad hoc*.

Des avancées ont été faites sur le sujet, comme notamment la prise en compte du partage du médium au niveau de la zone de détection de porteuse. Néanmoins, il ne faut pas perdre de vue que le partage du médium radio ne se fait pas qu'à l'émission mais aussi à la réception. Le scénario de la figure 1(a) est un exemple typique d'un tel partage à la réception. Or ce type de partage vient par exemple fausser tous les protocoles basés sur l'inéquation «somme des débits des flux émis inférieure à la capacité du médium» dans chaque zone de contention. Une telle équation est nécessaire mais non suffisante car c'est sans compter sur les collisions éventuelles qui peuvent apparaître au niveau des récepteurs. Par conséquent, pour évaluer les ressources sur un lien, il est nécessaire de connaître le partage à l'émission mais aussi le partage à la réception. Des recherches sont donc encore à mener sur par exemple l'estimation du recouvrement des périodes de silence de l'émetteur et du récepteur pour permettre une communication¹⁰ ou encore l'estimation des collisions au niveau des récepteurs. Pour ce dernier problème, la réalisation de l'estimation va être différente selon si on estime les collisions qu'un nouveau flux à émettre va subir ou si on estime les collisions qu'un nouveau flux à émettre va faire subir aux flux déjà existants. Le problème clé est bien sûr de comment réaliser ces estimations avec le moins d'informations locales possibles.

La plupart des protocoles qui se sont penchés sur les problèmes d'évaluation de ressources, ce sont surtout concentrés sur le paramètre bande passante libre. Les paramètres délai, taux de pertes et gigue sont souvent moins étudiés. Or comme nous l'avons indiqué en introduction, ces paramètres ont une importance conséquente en termes de qualité de service. Les problématiques resteront identiques à celles identifiées pour la bande passante : quel délai, taux de pertes ou gigue le réseau peut offrir, est-il possible de trouver une route qui aura tel délai, taux de pertes ou gigue ou encore quel est l'impact de l'introduction d'un nouveau flux en termes de délai, taux de pertes ou gigue sur les flux existants ? En revanche, les techniques pour faire ces évaluations seront peut-être différentes.

Enfin, une fois que des mécanismes précis d'évaluation de ressources seront au point, il sera très probablement possible de dimensionner correctement et dynamiquement les protocoles, notamment par exemple sur les valeurs seuils qui permettent de déterminer l'augmentation ou la diminution des débits des flux ou encore le rejet des flux.

10. Ce problème a fait l'objet d'une étude dans [SAR 05].

6.2. Gestion de la mobilité

La gestion de la mobilité doit définir comment on réagit aux changements dans la disponibilité des ressources dans le réseau dus à la mobilité. Le problème actuel n'est pas dans l'absence de solutions pour gérer la mobilité mais plutôt dans l'absence d'évaluations approfondies et de comparaisons entre les différentes méthodes. Par exemple, nous ne connaissons pas de travaux qui dégagent les différentes performances des gestions proposées selon les configurations et les modèles de mobilité comme on peut le trouver dans le domaine du routage. Par conséquent, il est difficile de se faire une idée sur ce qui est performant ou non.

Parmi les méthodes de gestion de la mobilité, on retrouve essentiellement trois approches. La première, qui est souvent donnée dans les protocoles basés sur un routage avec qualité de service, concerne le re-routage. Si une route n'est pas en mesure d'assurer les garanties qu'elle pouvait initialement offrir, alors on cherche une autre route qui respectera les contraintes. Très souvent, la reconstruction se fait à la source ou en local, mais le choix d'une telle reconstruction n'est jamais justifié. Il pourrait être intéressant d'avoir une étude qui évalue l'impact des deux méthodes (à la source ou local) sur la qualité de service. Dans le cas de flux sensibles au délai et/ou à la gigue, en particulier, il est important d'évaluer le temps de reconstruction d'une route afin de déterminer s'il est plus intéressant de maintenir des routes de secours, d'effectuer une reconstruction locale ou une reconstruction à la source.

La deuxième méthode se retrouve plutôt dans les protocoles qui ne se basent pas sur un routage avec qualité de service. Comme l'idée de ces protocoles est d'être le plus simple et le plus léger possible, l'apparition de congestion va se traiter par l'arrêt (puis éventuellement par la restauration) de certains flux. Le choix des flux à arrêter est une question clé. Une approche simple et équitable est d'arrêter tous les flux en même temps dans une zone de congestion. Néanmoins cette approche peut mener à une sous-utilisation des ressources pendant un certain temps et générer des oscillations. Le choix de certains flux peut se faire aléatoirement ou selon certains critères comme les priorités par exemple. Le problème, là encore, est d'obtenir des décisions pertinentes avec seulement une connaissance locale. Par exemple, nous avons vu précédemment que conserver un bon ordre global sur les priorités n'était pas si simple. La troisième consiste à relâcher les contraintes sur les garanties et à dégrader certains flux (puis éventuellement à les restaurer si cela est possible). On retrouve ici les mêmes problématiques que pour l'approche précédente : quelle politique appliquer sur les différents flux sachant qu'on n'a qu'une connaissance locale et que la dégradation peut avoir un impact sur tout le réseau ? Une telle question revient d'ailleurs à une problématique de compromis équité (dans le sens où l'on ne pénalise pas seulement un seul flux) - efficacité à obtenir avec seulement des infos locales. Cette dernière est un sujet d'étude très en vogue dans les réseaux *ad hoc* et est encore un problème d'actualité. Les solutions qui se basent sur cette approche oublient souvent la partie restauration des flux or cette partie est très importante car elle empêche une sous-utilisation des ressources et permet d'honorer au mieux les garanties initiales.

6.3. Gestion efficace des ressources du réseau

L'idéal, en termes d'utilisation du réseau, est de pouvoir offrir des garanties à un maximum de flux tout en ayant une bonne utilisation des ressources libres pour les flux type *Best Effort*. Cette notion d'efficacité est très difficile à atteindre dans un réseau *ad hoc* et là encore peu de solutions se concentrent sur cette problématique.

Une approche possible est de déterminer au préalable les ressources qui pourront être utilisées par chaque type de trafic. Si cette approche a l'avantage de ne pas gérer la dynamique dans l'allocation des ressources, elle peut s'avérer être relativement inefficace si certains types de trafic sont inexistant dans le réseau, puisqu'une partie des ressources ne seront pas utilisées. De plus, l'allocation statique n'est pas si simple car elle peut être très différente d'une configuration à une autre.

Certaines solutions régulent le trafic des flux *Best Effort* afin d'utiliser les ressources restantes. Pour cela, le débit des flux *Best Effort* est augmenté régulièrement jusqu'à éventuellement atteindre un seuil de saturation ; à partir de là le débit des flux *Best Effort* est diminué jusqu'à revenir à une solution sans saturation. Il existe plusieurs inconvénients à cette approche. Premièrement, comme nous l'avons vu précédemment, il n'est pas simple de déterminer le seuil de congestion qui est, de plus, dépendant des scénarios. Deuxièmement, cette approche peut aussi conduire à limiter le nombre de flux privilégiés acceptés dans le réseau. En effet, imaginons un scénario où il y a un seul flux privilégié qui consomme peu de ressources. Cette approche va augmenter les ressources allouées aux flux *Best Effort* jusqu'à utiliser le maximum possible de bande passante (sans dégrader le flux privilégié existant si c'est un bon protocole de qualité de service). Imaginons qu'un nouveau flux ayant une contrainte de bande passante cherche à être émis. Comme la plupart des protocoles d'évaluation de ressources ne font pas de différenciation des flux, l'émetteur du nouveau flux va considérer que toute la bande passante est occupée et qu'il ne peut pas trouver de route respectant les contraintes de bande passante. Or il suffirait peut-être de diminuer le débit de certains flux *Best Effort* pour permettre à ce nouveau flux d'être émis. Le problème ici est d'arriver à estimer qu'il y a suffisamment de ressources pour un flux privilégié même si les ressources sont occupées et d'arriver à diminuer en conséquence les débits des flux *Best Effort*. Une solution possible est d'introduire une différenciation des flux dans l'évaluation des ressources disponibles. Mais une telle approche sera toujours mise en défaut dès que le partage se fera entre des flux qui n'ont pas moyen de communiquer. Un tel problème est un réel challenge auquel peu de papiers se sont intéressés.

Enfin, à notre connaissance, il n'existe pas de travaux qui mènent une comparaison approfondie entre les approches basées sur le routage et les approches sans routage. Or les questions du type «est-ce que le routage avec qualité de service permet une bonne utilisation des ressources ?» ou encore «quel est le type de mobilité qui justifie de ne pas utiliser un routage avec qualité de service ?» nous semblent être des questions fondamentales dont les réponses devraient permettre de poser des jalons importants dans le domaine.

7. Conclusion

Dans cet article, nous donnons, au moyen d'études sur des réseaux sans fil à un saut, quelques indications sur les performances escomptées des réseaux *ad hoc* basés sur une approche de type IEEE 802.11. Beaucoup reste à faire dans ce domaine, les évaluations présentées n'étant pas applicables directement dans un contexte multisaut. Comprendre les interactions entre les différents flux transportés par un réseau n'est pas simple, traduire cette compréhension en politiques locales de gestion des ressources l'est encore moins.

Par la suite, nous décrivons quelques solutions représentatives de la recherche dans le domaine de la qualité de service dans ces réseaux et tâchons d'en extraire les thématiques de recherche encore trop peu explorées. La plupart des solutions peuvent être classées selon le niveau de garanties offert aux applications, des garanties totalement statistiques aux garanties strictes et quantitatives, mais aussi selon le niveau de trafic de contrôle qu'elles génèrent. Il existe, comme toujours, un compromis évident entre volume d'informations échangées (interférence entre flux, volumes et profils de trafics au niveau de chaque nœud, mobilité des nœuds, etc.) et précision des garanties. Plus les nœuds seront au fait de l'état du réseau, plus ils seront capables d'évaluer la capacité résiduelle de ce dernier, la mobilité des terminaux, etc.

Si la communauté s'est beaucoup intéressée aux aspects de performance des solutions, telles que la capacité de différenciation ou la stabilité des routes choisies par un protocole de routage QoS, certains aspects, tels que la *TCP friendliness*, restent à ce jour totalement ignorés et devront être évalués. La question de la qualité de service dans les réseaux *ad hoc* est donc loin d'être close et, en dehors des choix philosophiques (qui préférera une approche statistique du fait du passage à l'échelle), un travail important serait de fournir une base d'évaluation standard des différentes solutions définissant scénarios, outils et paramètres de simulations, plates-formes d'expérimentations, etc.

Enfin, tout au long de cet article, nous avons délibérément adopté un point de vue critique incisif. Nous tenons à souligner ici que des avancées significatives ont été réalisées dans le domaine de la qualité de service pour les réseaux *ad hoc* et que certaines des solutions proposées ne sont pas des traductions directes de solutions filaires, mais pensées pour les caractéristiques des réseaux *ad hoc*. Néanmoins, il est intéressant de remarquer qu'un scénario très simple comme celui de la figure 1(a) met la majorité des protocoles en défaut. Nous sommes sensibles au fait qu'un tel scénario est peut-être un cas très particulier, cependant, comme nous l'avons dit précédemment, les problèmes microscopiques ne peuvent pas être omis quand la qualité de service est en jeu. D'autre part, nous ne savons actuellement rien sur l'apparition d'un tel scénario dans des configurations réelles de réseaux *ad hoc*. Regarder le nombre d'occurrences de cette sous-configuration dans un réseau de type Mesh ou un réseau déployé aléatoirement est par exemple un problème intéressant. Enfin, un tel scénario, problématique pour la qualité de service, n'est peut-être que la partie visible de l'iceberg, et que la combinaison de configurations clés rendra le problème beaucoup plus ardu. Ceci

montre qu'une compréhension fine des propriétés des réseaux *ad hoc*, ainsi que l'intégration de cette compréhension dans les solutions de qualité de service restent encore à réaliser. C'est finalement une bonne nouvelle pour les chercheurs, car il reste encore beaucoup de problèmes ouverts dans le domaine de la qualité de service pour les réseaux *ad hoc*.

8. Bibliographie

- [AHN 02] AHN G.-S., CAMPBELL A. T., VERES A., SUN L.-H., « Supporting Service Differentiation for Real-Time and Best-Effort Traffic in Stateless Wireless Ad Hoc Networks (SWAN) », *IEEE Transactions on Mobile Computing*, vol. 1, n° 3, 2002, p. 192-207, IEEE press.
- [BHA 94] BHARGHAVAN V., DEMERS A., SHENKER S., ZHANG L., « MACAW : A media access protocol for wireless LANs », *Proceedings of the ACM SIGCOMM '94 Conference on Communications Architectures, Protocols and Applications*, London, UK, août 1994, p. 212-225.
- [BIA 00] BIANCHI G., « Performance Analysis of the IEEE 802.11 Distributed Coordination Function », *IEEE Journal on Selected Areas in Communications*, vol. 18, n° 3, 2000, p. 535-547, IEEE press.
- [CHA 02] CHAUDET C., GUÉRIN LASSOUS I., « BRuIT : Bandwidth Reservation under Interferences influence », *European Wireless*, Florence, Italie, février 2002, p. 466-472.
- [CHA 05] CHAUDET C., DHOUTAUT D., GUÉRIN LASSOUS I., « Performance Issues with IEEE 802.11 in Ad Hoc Networking », *IEEE Communication Magazine*, vol. 43, n° 7, 2005.
- [GUP 00] GUPTA P., KUMAR P., « The capacity of wireless networks », *IEEE Transactions on Information Theory*, vol. IT-46, n° 2, 2000, p. 388-404, IEEE press.
- [IEE 97] IEEE STANDARD FOR INFORMATION TECHNOLOGY – TELECOMMUNICATIONS AND INFORMATION EXCHANGE BETWEEN SYSTEMS, « Local and Metropolitan Area Network – Specific Requirements – Part 11 : Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications », 1997.
- [IEE 99a] IEEE STANDARD FOR INFORMATION TECHNOLOGY – TELECOMMUNICATIONS AND INFORMATION EXCHANGE BETWEEN SYSTEMS, « Local and Metropolitan Area Network – Specific Requirements – Part 11 : Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications — Higher-speed physical layer extension in the 2.4 GHz band », 1999.
- [IEE 99b] IEEE STANDARD FOR INFORMATION TECHNOLOGY – TELECOMMUNICATIONS AND INFORMATION EXCHANGE BETWEEN SYSTEMS, « Local and Metropolitan Area Network – Specific Requirements – Part 11 : Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications — High-speed physical layer in the 5 GHz band », 1999.
- [IEE 03] IEEE STANDARD FOR INFORMATION TECHNOLOGY – TELECOMMUNICATIONS AND INFORMATION EXCHANGE BETWEEN SYSTEMS, « Local and Metropolitan Area Network – Specific Requirements – Part 11 : Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications — Amendment 4 : Further Higher Data Rate Extension in the 2.4 GHz Band », 2003.

- [LIN 03] LINDGREN A., ALMQUIST A., SCHELÉN O., « Quality of Service Schemes for IEEE 802.11 Wireless LANs - An Evaluation », *Mobile Networks and Applications*, vol. 8, n° 3, 2003, p. 223-235, Kluwer Academic Press.
- [MAN 05] MANSHAEI M. H., « Cross Layer Interactions for Adaptive Communications in IEEE 802.11 Wireless LANs », PhD thesis, Université de Nice - Sophia Antipolis, décembre 2005.
- [ROM 03] ROMDHANI R., NI Q., TURLETTI T., « AEDCF : Enhanced Service Differentiation for IEEE 802.11 Wireless Ad-Hoc Networks », *IEEE WCNC*, New Orleans, Louisiana, USA, mars 2003.
- [SAR 05] SARR C., CHAUDET C., CHELIUS G., GUÉRIN LASSOUS I., « A node-based available bandwidth evaluation in IEEE 802.11 ad hoc networks », *International Journal of Parallel, Emergent and Distributed Systems*, , 2005, To Appear.
- [SOB 96] SOBRINHO J. L., KRISHNAKUMAR A. S., « Real-Time Traffic over the IEEE 802.11 Medium Access Control Layer », *Bell Labs Technical Journal*, vol. 1, n° 2, 1996, p. 172-187.
- [Uni 03] UNION INTERNATIONALE DES TÉLÉCOMMUNICATIONS, « Connexions et circuits téléphoniques internationaux – Généralités sur la qualité de transmission d'une connexion téléphonique internationale complète – Temps de transmission dans un sens », rapport n° G.114, 2003, UIT-T.
- [Uni 05] UNION INTERNATIONALE DES TÉLÉCOMMUNICATIONS, « Le modèle E : modèle de calcul utilisé pour la planification de la transmission », rapport n° G.107, 2005, UIT-T.
- [WU 01] WU K., HARMS J., « QoS Support in Mobile Ad Hoc Networks », *Crossing Boundaries- the GSA Journal of University of Alberta*, novembre 2001.
- [YAN 04] YANG Y., KRAVETS R., « Distributed QoS Guarantees for Realtime Traffic in Ad Hoc Networks », *First IEEE International Conference on Sensor and Ad hoc Communications and Networks (SECON)*, juin 2004.
- [YAN 05] YANG Y., KRAVETS R., « Contention-Aware Admission Control for Ad Hoc Networks », *IEEE Transactions on Mobile Computing*, vol. 4/4, 2005, p. 363–377.