

Union internationale des télécommunications

UIT-T

SECTEUR DE LA NORMALISATION
DES TÉLÉCOMMUNICATIONS
DE L'UIT

G.1030

(11/2005)

SÉRIE G: SYSTÈMES ET SUPPORTS DE
TRANSMISSION, SYSTÈMES ET RÉSEAUX
NUMÉRIQUES

Qualité de service et de transmission – Aspects
génériques et aspects liés à l'utilisateur

**Evaluation de la qualité de fonctionnement
de bout en bout dans les réseaux IP pour les
applications de transmission de données**

Recommandation UIT-T G.1030



RECOMMANDATIONS UIT-T DE LA SÉRIE G
SYSTÈMES ET SUPPORTS DE TRANSMISSION, SYSTÈMES ET RÉSEAUX NUMÉRIQUES

CONNEXIONS ET CIRCUITS TÉLÉPHONIQUES INTERNATIONAUX	G.100–G.199
CARACTÉRISTIQUES GÉNÉRALES COMMUNES À TOUS LES SYSTÈMES ANALOGIQUES À COURANTS PORTEURS	G.200–G.299
CARACTÉRISTIQUES INDIVIDUELLES DES SYSTÈMES TÉLÉPHONIQUES INTERNATIONAUX À COURANTS PORTEURS SUR LIGNES MÉTALLIQUES	G.300–G.399
CARACTÉRISTIQUES GÉNÉRALES DES SYSTÈMES TÉLÉPHONIQUES INTERNATIONAUX HERTZIENS OU À SATELLITES ET INTERCONNEXION AVEC LES SYSTÈMES SUR LIGNES MÉTALLIQUES	G.400–G.449
COORDINATION DE LA RADIODÉLÉPHONIE ET DE LA TÉLÉPHONIE SUR LIGNES	G.450–G.499
CARACTÉRISTIQUES DES SUPPORTS DE TRANSMISSION	G.600–G.699
EQUIPEMENTS TERMINAUX NUMÉRIQUES	G.700–G.799
RÉSEAUX NUMÉRIQUES	G.800–G.899
SECTIONS NUMÉRIQUES ET SYSTÈMES DE LIGNES NUMÉRIQUES	G.900–G.999
QUALITÉ DE SERVICE ET DE TRANSMISSION – ASPECTS GÉNÉRIQUES ET ASPECTS LIÉS À L'UTILISATEUR	G.1000–G.1999
CARACTÉRISTIQUES DES SUPPORTS DE TRANSMISSION	G.6000–G.6999
DONNÉES SUR COUCHE TRANSPORT – ASPECTS GÉNÉRIQUES	G.7000–G.7999
ASPECTS RELATIFS AU PROTOCOLE ETHERNET SUR COUCHE TRANSPORT	G.8000–G.8999
RÉSEAUX D'ACCÈS	G.9000–G.9999

Pour plus de détails, voir la Liste des Recommandations de l'UIT-T.

Recommandation UIT-T G.1030

Evaluation de la qualité de fonctionnement de bout en bout dans les réseaux IP pour les applications de transmission de données

Résumé

La présente Recommandation définit un ensemble d'outils permettant de déterminer la qualité de fonctionnement des réseaux IP, d'évaluer la qualité de fonctionnement des applications d'utilisateur et d'appliquer des modèles de perception pour mesurer le degré de satisfaction de l'utilisation à l'égard de la qualité de fonctionnement de bout en bout.

La qualité de fonctionnement des applications de transmission de données sur les réseaux par paquets, telle qu'elle est perçue par l'utilisateur, dépend de nombreux facteurs, parmi lesquels la qualité de fonctionnement de bout en bout du réseau par paquets, la dépendance de l'application à l'égard du réseau de communication, la qualité de fonctionnement des terminaux et des autres dispositifs ne relevant pas de l'opérateur (ou des opérateurs) de réseau, ainsi que la tâche que doit accomplir l'utilisateur et l'importance de l'interaction de celui-ci avec l'application. Les concepteurs de réseau tiennent compte de ces facteurs pour satisfaire l'utilisateur. Une fois que la qualité de fonctionnement a été évaluée, des modèles de perception peuvent être appliqués pour interpréter le niveau de qualité de fonctionnement atteint.

La présente Recommandation pose en principe que le lecteur sera à même d'établir à tout le moins un certain nombre d'indications détaillées sur chacun des principaux facteurs susmentionnés, puis d'utiliser l'ensemble d'outils défini ici pour évaluer la qualité de fonctionnement de bout en bout.

Source

La Recommandation UIT-T G.1030 a été approuvée le 29 novembre 2005 par la Commission d'études 12 (2005-2008) de l'UIT-T selon la procédure définie dans la Recommandation UIT-T A.8.

AVANT-PROPOS

L'UIT (Union internationale des télécommunications) est une institution spécialisée des Nations Unies dans le domaine des télécommunications. L'UIT-T (Secteur de la normalisation des télécommunications) est un organe permanent de l'UIT. Il est chargé de l'étude des questions techniques, d'exploitation et de tarification, et émet à ce sujet des Recommandations en vue de la normalisation des télécommunications à l'échelle mondiale.

L'Assemblée mondiale de normalisation des télécommunications (AMNT), qui se réunit tous les quatre ans, détermine les thèmes d'étude à traiter par les Commissions d'études de l'UIT-T, lesquelles élaborent en retour des Recommandations sur ces thèmes.

L'approbation des Recommandations par les Membres de l'UIT-T s'effectue selon la procédure définie dans la Résolution 1 de l'AMNT.

Dans certains secteurs des technologies de l'information qui correspondent à la sphère de compétence de l'UIT-T, les normes nécessaires se préparent en collaboration avec l'ISO et la CEI.

NOTE

Dans la présente Recommandation, l'expression "Administration" est utilisée pour désigner de façon abrégée aussi bien une administration de télécommunications qu'une exploitation reconnue.

Le respect de cette Recommandation se fait à titre volontaire. Cependant, il se peut que la Recommandation contienne certaines dispositions obligatoires (pour assurer, par exemple, l'interopérabilité et l'applicabilité) et considère que la Recommandation est respectée lorsque toutes ces dispositions sont observées. Le futur d'obligation et les autres moyens d'expression de l'obligation comme le verbe "devoir" ainsi que leurs formes négatives servent à énoncer des prescriptions. L'utilisation de ces formes ne signifie pas qu'il est obligatoire de respecter la Recommandation.

DROITS DE PROPRIÉTÉ INTELLECTUELLE

L'UIT attire l'attention sur la possibilité que l'application ou la mise en œuvre de la présente Recommandation puisse donner lieu à l'utilisation d'un droit de propriété intellectuelle. L'UIT ne prend pas position en ce qui concerne l'existence, la validité ou l'applicabilité des droits de propriété intellectuelle, qu'ils soient revendiqués par un Membre de l'UIT ou par une tierce partie étrangère à la procédure d'élaboration des Recommandations.

A la date d'approbation de la présente Recommandation, l'UIT n'avait pas été avisée de l'existence d'une propriété intellectuelle protégée par des brevets à acquérir pour mettre en œuvre la présente Recommandation. Toutefois, comme il ne s'agit peut-être pas de renseignements les plus récents, il est vivement recommandé aux responsables de la mise en œuvre de consulter la base de données des brevets du TSB.

© UIT 2006

Tous droits réservés. Aucune partie de cette publication ne peut être reproduite, par quelque procédé que ce soit, sans l'accord écrit préalable de l'UIT.

TABLE DES MATIÈRES

		Page
1	Domaine d'application	1
1.1	Evaluation de la qualité de fonctionnement du réseau	1
1.2	Evaluation de la qualité de fonctionnement de l'application	2
1.3	Modèles de perception.....	2
1.4	Ensemble de modèles du processus d'évaluation de la qualité de fonctionnement de bout en bout	2
2	Références normatives.....	3
3	Abréviations.....	4
4	Evaluation de la qualité de service pour l'utilisateur final	4
	Annexe A – Modèle d'opinion pour les applications de navigation sur le Web	4
A.1	Domaine d'application	4
A.2	Introduction	4
A.3	Essais subjectifs d'évaluation de la qualité de navigation sur le Web et résultats	6
A.4	Résultats de modélisation	9
A.5	Qualité perçue de sessions de navigation sur le web d'une page et d'événements de temporisation isolés (non normatif)	13
A.6	Référence à l'Annexe A	15
	Appendice I – Evaluations de la qualité de fonctionnement du réseau au moyen d'informations restreintes.....	15
I.1	Introduction	15
I.2	Connexion de référence	15
I.3	Concaténation des valeurs de qualité de fonctionnement pour le transfert de paquets	16
I.4	Bande passante engorgée.....	17
I.5	Durée de la prise de contact.....	20
I.6	Exemple d'application: transaction HTTP	22
I.7	Résumé	24
I.8	Références à l'Appendice I	24

Introduction

La qualité de fonctionnement des applications de transmission de données sur les réseaux par paquets, telle qu'elle est perçue par l'utilisateur, dépend de nombreux facteurs, parmi lesquels il y a lieu de citer, en particulier:

- 1) La qualité de fonctionnement de bout en bout du réseau par paquets (en termes de connectivité, de perte de paquets, de délai et de capacité de transfert de paquets, par exemple), compte tenu de la conception du réseau et de la charge du trafic utilisateur. La qualité de fonctionnement du réseau constitue communément l'élément premier de la qualité de transmission de bout en bout.
- 2) La dépendance de l'application à l'égard du réseau de communication (par exemple, le nombre d'échanges de paquets nécessaires pour mener à bien une transaction, le contrôle de flux pour le transfert de données) et des autres entités de réseau assurant des services à l'application (serveurs de noms de domaine, par exemple).
- 3) La qualité de fonctionnement des dispositifs connexes ne relevant pas de l'opérateur (ou des opérateurs) de réseau (équipement d'utilisateur, serveurs, par exemple).
- 4) La tâche que doit accomplir l'utilisateur et l'importance de l'interaction de celui-ci avec l'application.

Les concepteurs de réseau tiennent compte de ces facteurs pour satisfaire l'utilisateur. Un modèle de qualité de fonctionnement pour les applications de transmission de données doit inclure le plus grand nombre possible de ces facteurs. Une fois que la qualité de fonctionnement a été évaluée, des modèles de perception peuvent être appliqués pour interpréter le niveau de qualité de fonctionnement atteint.

La présente Recommandation pose en principe que le lecteur sera à même d'établir à tout le moins un certain nombre d'indications détaillées sur chacun des principaux facteurs susmentionnés, puis d'utiliser un ensemble d'outils (dont beaucoup sont définis ici) pour évaluer la qualité de fonctionnement de bout en bout.

Recommandation UIT-T G.1030

Evaluation de la qualité de fonctionnement de bout en bout dans les réseaux IP pour les applications de transmission de données

1 Domaine d'application

La présente Recommandation traite du processus d'évaluation de la qualité de fonctionnement de bout en bout des applications fonctionnant sur réseaux IP, qui utilise:

- la qualité de fonctionnement du réseau IP considéré, déterminée selon des mesures appropriées ou d'après les résultats de modélisation de réseau obtenus;
- les spécifications de l'application considérée, pour ce qui est du ou des protocoles qui la régissent avec les options définies, ou un modèle de l'application utilisant comme données d'entrée la qualité de fonctionnement du réseau et la qualité de fonctionnement de l'équipement d'abonné et permettant d'obtenir comme résultat une valeur essentielle de qualité de fonctionnement pour cette application (le temps de téléchargement de fichier, par exemple);
- un modèle de perception conçu pour les applications considérées, permettant d'interpréter la qualité de fonctionnement de l'application comme constituant une estimation de la qualité perçue par une population type d'utilisateurs.

La Figure 1 illustre le processus général d'évaluation de la qualité de fonctionnement de bout en bout.

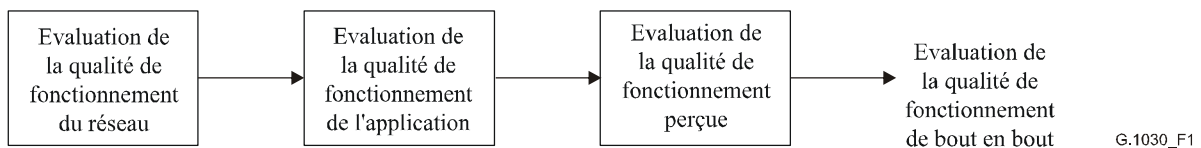


Figure 1/G.1030 – Processus d'évaluation de la qualité de fonctionnement de bout en bout

A noter que l'étape d'évaluation de la qualité de fonctionnement du réseau et l'étape d'évaluation de la qualité de fonctionnement de l'application peuvent parfois être associées, dans le cas par exemple où une simulation permet de mesurer la qualité de fonctionnement d'une session donnée, ou d'un ensemble de sessions.

1.1 Evaluation de la qualité de fonctionnement du réseau

La qualité de fonctionnement du réseau peut être évaluée en fonction des paramètres de performance pour le transfert de paquets définis dans la Rec. UIT-T Y.1540 et dans d'autres normes applicables (documents RFC mis au point par le Groupe de travail des mesures de performance IP du Groupe de travail d'ingénierie Internet (IEFT, *Internet engineering task force*)). Outre les mesures usuelles de performance pour le transfert dans un seul sens, on pourra parfois recourir à la méthode la plus simple pour l'évaluation de la qualité de fonctionnement de l'application, qui fait appel à des mesures dans les deux sens (aller-retour).

Il existe deux sources principales d'informations, de données de mesure et de modélisation sur la qualité de fonctionnement du réseau.

Les mesures sur le réseau permettent à l'évaluateur de traiter le réseau comme une boîte noire et de produire des informations qui pourront être utiles dans les étapes restantes de la chaîne de

modélisation. Toutefois, plusieurs éléments importants doivent être pris en considération pour la conception des mesures, à savoir:

- 1) la discipline imposée en matière d'émission doit correspondre à l'application considérée, dans certains cas. Par exemple, le contrôle de flux exercé par le protocole de commande de transmission (TCP, *transmission control protocol*) est fonction des conditions du réseau et a généralement pour effet d'allonger la file d'attente lorsque celui-ci est engorgé, d'où un délai de transfert supérieur à celui qui serait mesuré en temps normal;
- 2) la charge de trafic en l'absence de mesures doit être analogue, en volume et en nature, aux conditions dans lesquelles les évaluations de la qualité de fonctionnement vont s'appliquer. Des mesures sur un réseau exempt de toute charge de trafic ne présentent pas d'intérêt particulier.

La modélisation du réseau peut aider à définir la qualité de fonctionnement nécessaire lorsque la construction du réseau n'est pas encore achevée (par exemple, les noeuds et les liaisons sont en place, mais le protocole indispensable à leur fonctionnement n'est pas en service), ou lorsque les principaux éléments à prendre en considération pour les mesures ne peuvent être réunis. Il existe un grand choix d'outils de modélisation, parmi lesquels des produits commerciaux et des outils de recherche du domaine public. Les outils de modélisation exigent un haut niveau de connaissance technique du réseau considéré ainsi qu'une quantité non négligeable d'informations sur celui-ci pour être utilisés de manière efficace. En lieu et place de cette méthode de modélisation relativement précise, on peut aussi utiliser les méthodes simplifiées définies dans l'Appendice I pour évaluer la qualité de fonctionnement de bout en bout d'un réseau IP. Cependant, la précision de cette méthode dépendra grandement de la précision des informations fournies.

1.2 Evaluation de la qualité de fonctionnement de l'application

Les modèles d'application intègrent comme données d'entrée les évaluations de la qualité de fonctionnement du réseau ainsi que les informations relatives à la qualité de fonctionnement du dispositif d'application, et produisent comme données de sortie une ou plusieurs des principales mesures de la qualité de fonctionnement de l'application.

Un modèle d'application de ce type, pour les échanges de paquets en mode conversation caractéristiques des transactions par carte de crédit et des autres transactions au point de vente, est défini dans la Rec. UIT-T G.1040.

Pour le transfert de fichiers à longue durée de vie par le service de transfert d'octets fiable du protocole TCP, les modèles présentés dans l'Appendice I donnent des résultats précis, pour autant que les données d'entrée soient exactes.

1.3 Modèles de perception

La présente Recommandation expose des modèles de perception permettant d'obtenir la qualité de fonctionnement de bout en bout souhaitée.

L'Annexe A définit un modèle de navigation sur le web.

Dès qu'il sera accessible, tout nouveau modèle de navigation fera l'objet de nouvelles annexes de la présente Recommandation.

1.4 Ensemble de modèles du processus d'évaluation de la qualité de fonctionnement de bout en bout

La Figure 2 donne un aperçu des options permettant d'évaluer la qualité de fonctionnement de bout en bout des applications sur les réseaux IP. Il ressort de cette figure qu'il existe maintes options pour mener à bien cette évaluation, bien que dans la pratique l'évaluateur soit tenu d'associer d'entre elles les options qui sont compatibles avec le but d'une évaluation de bout en bout (et compatibles entre elles).

Données d'entrée (outre l'application considérée)	Evaluation de la qualité de fonctionnement du réseau	Evaluation de la qualité de fonctionnement de l'application	Evaluation de la qualité de fonctionnement perçue	
Accès au réseau pour essai en trafic réel	Mesures <ul style="list-style-type: none"> De diagnostic de la connectivité réseau (Ping, <i>packet Internet-network grouper</i>) D'échantillon non polarisé Du flux d'application De l'outil de transfert en masse 	Modèle applicable aux flux TCP à longue durée de vie (Appendice I) Evaluation directe au moyen d'un outil d'application	Modèle de navigation sur le Web (Annexe A)	Evaluation de la qualité de fonctionnement de bout en bout
Conception du réseau (indications détaillées)	Outils de modélisation <ul style="list-style-type: none"> Produits commerciaux Outils de recherche 	Evaluation intégrée dans l'outil de modélisation du réseau	Futur modèle de téléconférence vidéo	
Conception du réseau (indications restreintes)	Evaluation simplifiée (Appendice I)	Rec. UIT-T G.1040 (Note)	Autres modèles	

NOTE – La Rec. UIT-T G.1040 indique la contribution du réseau à la qualité de fonctionnement des applications de transaction. Sans procéder à une évaluation complète de la qualité de fonctionnement des applications, elle donne un aperçu de la qualité de fonctionnement du réseau axé sur les applications.

Figure 2/G.1030 – Ensemble de méthodes d'évaluation de la qualité de fonctionnement de bout en bout dans les réseaux IP

2 Références normatives

La présente Recommandation se réfère à certaines dispositions des Recommandations UIT-T et textes suivants qui, de ce fait, en sont partie intégrante. Les versions indiquées étaient en vigueur au moment de la publication de la présente Recommandation. Toute Recommandation ou tout texte étant sujet à révision, les utilisateurs de la présente Recommandation sont invités à se reporter, si possible, aux versions les plus récentes des références normatives suivantes. La liste des Recommandations de l'UIT-T en vigueur est régulièrement publiée. La référence à un document figurant dans la présente Recommandation ne donne pas à ce document, en tant que tel, le statut d'une Recommandation.

- Recommandation UIT-T G.1010 (2001), *Catégories de qualité de service multimédia pour l'utilisateur final*.
- Recommandation UIT-T G.1040 (2006), *Contribution du réseau à la durée des transactions*.
- Recommandation UIT-T Y.1540 (2002), *Service de communication de données par protocole Internet – Paramètres de performance pour le transfert de paquets IP et la disponibilité de ce service*.

3 Abréviations

IP protocole Internet (*Internet protocol*)

QS qualité de service

TCP protocole de commande de transmission (*transmission control protocol*)

4 Evaluation de la qualité de service pour l'utilisateur final

Après avoir évalué la qualité de fonctionnement physique de bout en bout, les planificateurs de réseau devront impérativement évaluer la qualité de service pour l'utilisateur final en prenant en compte les aspects relatifs à la perception.

La Rec. UIT-T G.1010 définit les catégories de qualité de service pour les applications de données ainsi que pour les applications audio et vidéo. Par ailleurs, l'Annexe A définit un modèle d'opinion permettant d'évaluer la qualité des applications de navigation sur le Web telle qu'elle est perçue par les utilisateurs. Ce modèle permet de procéder à une évaluation diagnostique d'après les valeurs mesurées ou prévues du délai de transmission de bout en bout en fonction de l'opinion des utilisateurs.

Annexe A

Modèle d'opinion pour les applications de navigation sur le Web

A.1 Domaine d'application

La présente annexe définit un modèle permettant d'établir un mappage entre les temps de réponse et de téléchargement mesurés dans le réseau ou calculés d'après la durée des transactions HTTP et la qualité perçue d'une session de navigation sur le Web. Le modèle est axé sur des essais pour lesquels les temps de réponse et de téléchargement au cours d'une session de navigation sur le Web ont été manipulés [1], [2]. Son domaine d'application se limite pour l'instant à des sessions de navigation sur le Web en deux étapes: présentation d'une demande de recherche pour la première étape et affichage d'une page de résultats pour la seconde. Une extension simple à des événements de temporisation isolés, avec modélisation de l'incidence du délai d'attente par page, est également présentée.

A.2 Introduction

En ce qui concerne la modélisation de la qualité perçue d'une session de navigation sur le Web, il est à noter que la qualité perçue est principalement fonction de la durée maximale *prévue* de cette session. Si l'on s'attend à une durée de session de 100 secondes, une session de 10 secondes sera perçue comme étant d'une qualité bien supérieure que si l'on s'attend à une durée de session de 1 seconde. En conséquence, le modèle intègre une approche modulable en fonction du contexte en utilisant trois échelles de temps différentes, de 6, 15 et 60 secondes, correspondant respectivement à des contextes de réseau rapide, modéré et lent.

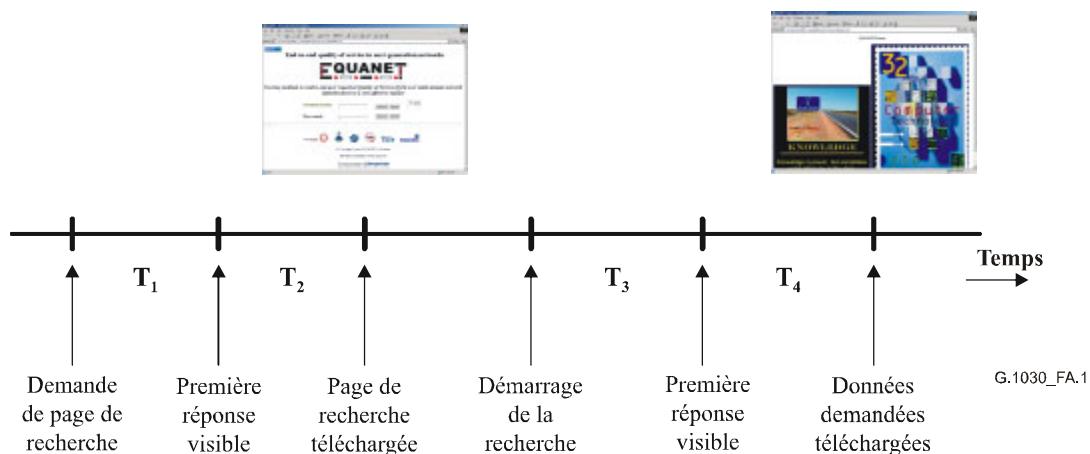
En règle générale, la perception de la qualité liée au temps de réponse peut être classée conformément aux trois régions de perception suivantes [3]:

- 1) **expérience instantanée**: un temps de réponse de 0,1 seconde constitue approximativement la limite à ne pas dépasser pour avoir l'impression que le système réagit instantanément, limite qu'il importe de respecter dans le cas de services de conversation (discussion en ligne, par exemple).

- 2) **expérience ininterrompue:** un temps de réponse de 1,0 seconde constitue approximativement la limite à ne pas dépasser pour que l'utilisateur ne perde pas le fil de ses pensées, même s'il perd la sensation que le service se déroule en direct, limite qu'il importe de respecter dans le cas de services interactifs (jeux, par exemple).
- 3) **perte d'attention:** un temps de réponse de 10 secondes constitue approximativement la limite à ne pas dépasser pour que l'utilisateur reste attentif au dialogue. Dans le cas de temps de réponse plus longs, les utilisateurs veulent effectuer d'autres tâches en attendant que l'ordinateur termine l'opération en cours et celui-ci devrait donc leur indiquer quand cette opération devrait prendre fin. Ce retour de l'information pendant le délai d'attente est particulièrement important dans les cas où le temps de réponse peut être extrêmement variable, du fait que les utilisateurs ne sauront pas alors à quoi s'attendre.

En ce qui concerne les temps de téléchargement, les sujets sont enclins à adapter le jugement qu'ils portent sur la qualité en fonction du temps de téléchargement prévu [4]. Dès qu'ils sont informés de ce dernier, ils sont prêts à accepter des temps de téléchargement longs.

Le modèle défini dans la présente Recommandation indique la corrélation entre les différents temps de réponse et de téléchargement dans le cadre de sessions de navigation sur le web et la qualité perçue correspondante pour une durée de session *maximale* donnée, dans telle ou telle configuration de réseau ou de système. Ce modèle est applicable à un grand nombre de configurations de réseaux et de systèmes, ainsi qu'aux services de navigation sur le Web offerts à de multiples utilisateurs. Les essais subjectifs, sur la base desquels le modèle a été établi, ont reproduit le plus fidèlement possible une session de navigation sur le Web telle qu'elle se déroule dans la pratique. Trois essais subjectifs de navigation sur le Web, d'une durée de l'ordre de 6, 15 et 60 secondes, correspondant respectivement aux contextes de réseau rapide, modéré et lent, ont été utilisés pour la mise au point du modèle. Au cours de chaque session, un sujet commence par extraire une page de recherche puis une page affichant les résultats de la recherche. La Figure A.1 indique le déroulement temporel d'une telle session. Les deux premiers intervalles de temps, T_1 et T_2 , représentent les temps de réponse et de téléchargement non interactifs de la page de recherche. Les deux intervalles de temps suivants, T_3 et T_4 , représentent les temps de réponse et de téléchargement interactifs de la page de résultats.



T_1 est le temps de réponse non interactif qui a été manipulé en utilisant Java scripting. T_2 est le temps de téléchargement non interactif qui a été manipulé en utilisant le manipulateur de réseau. T_3 et T_4 sont les équivalents de la partie interactive. La somme de $T_1 + T_2 + T_3 + T_4$ représente la durée de la session.

Figure A.1/G.1030 – Temporisateurs d'essai

En raison de la différence de comportement connue entre utilisateurs qualifiés et utilisateurs non qualifiés, dits utilisateurs inexpérimentés, une distinction est établie entre ces deux groupes d'utilisateurs aux fins de la mise au point d'un modèle de prévision de la qualité de navigation sur le Web qui soit adapté à une population nombreuse d'utilisateurs.

A.3 Essais subjectifs d'évaluation de la qualité de navigation sur le Web et résultats

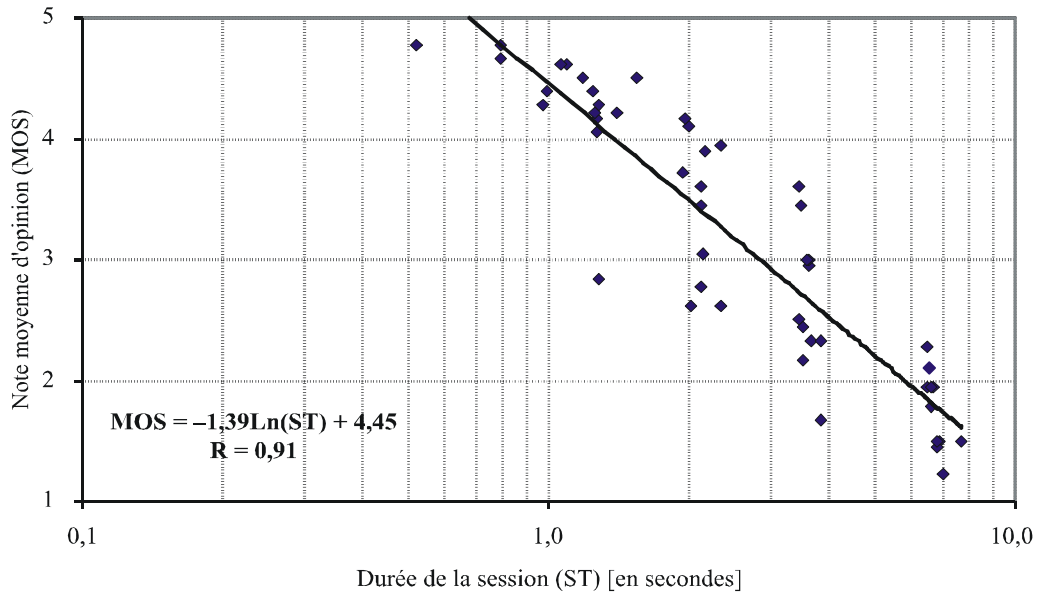
Pour les essais, et par conséquent dans le modèle, on a retenu l'échelle d'évaluation par catégories absolues (ACR, *absolute category rating*) de l'UIT-T [5] (échelle à cinq notes de qualité: 5 = excellente; 4 = bonne; 3 = passable; 2 = médiocre; 1 = mauvaise). Pour chaque essai, 49 sessions ont été présentées aux sujets, qui devaient pour chacune d'entre elles:

- demander, extraire et afficher une page de recherche;
- taper et soumettre un terme à rechercher sur cette page;
- extraire et afficher la page de résultats.

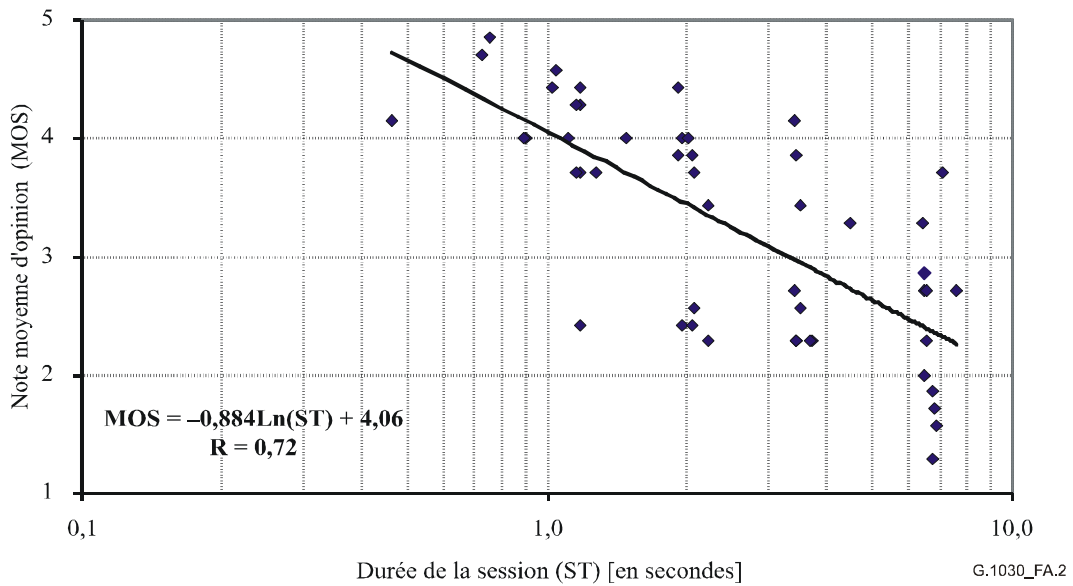
Pour obtenir des données cohérentes, on a utilisé exactement les mêmes pages de résultats pour chaque session en demandant aux sujets de taper la même requête de recherche. Du point de vue des sujets, le moteur de recherche doit d'abord trouver la page de résultats, qui doit ensuite être téléchargée. Pour chacune des 49 sessions, différentes combinaisons de T_1 à T_4 sont configurées, la somme $T_1 + T_2 + T_3 + T_4$, c'est-à-dire la durée de la session, variant de 0 à l'échelle des temps pour cette série d'essais.

Les résultats globaux sont indiqués dans les Figures A.2 à A.4 qui représentent la courbe de la durée de la session (c'est-à-dire $T_1 + T_2 + T_3 + T_4$) (portée en abscisse) par rapport à la valeur de note moyenne d'opinion (MOS) (portée en ordonnée) pour les trois essais. Pour le contexte de longue durée, voir la Figure A.4, les résultats obtenus par les sujets inexpérimentés et par les sujets expérimentés ont été pratiquement les mêmes et les valeurs MOS ont été calculées pour l'ensemble de la population. Pour les deux contextes de plus courte durée, voir les Figures A.2 et A.3, les sujets inexpérimentés et les sujets expérimentés se sont comportés de manière différente, et les corrélations entre les durées de session et la qualité perçue sont sensiblement différentes et sont donc indiquées séparément. Tous les résultats mettent en évidence le même phénomène: une diminution linéaire de la qualité perçue avec l'algorithme de la durée de la session. Dans le contexte expérimental de longue durée, la corrélation est suffisamment élevée ($> 0,9$) pour permettre l'obtention de prévisions de qualité fiables pour les sujets inexpérimentés et expérimentés. En règle générale, des corrélations supérieures à 0,9 sont recherchées aux fins de l'établissement de modèles psychophysiques de perception de la qualité [6]. Il ressort également des résultats que pour l'essai de 6 secondes avec des sujets inexpérimentés, la corrélation entre la durée de session et la qualité perçue est beaucoup trop faible (0,72) pour que l'on puisse se contenter d'utiliser un modèle axé uniquement sur la durée de la session.

Résultats pour les sujets expérimentés

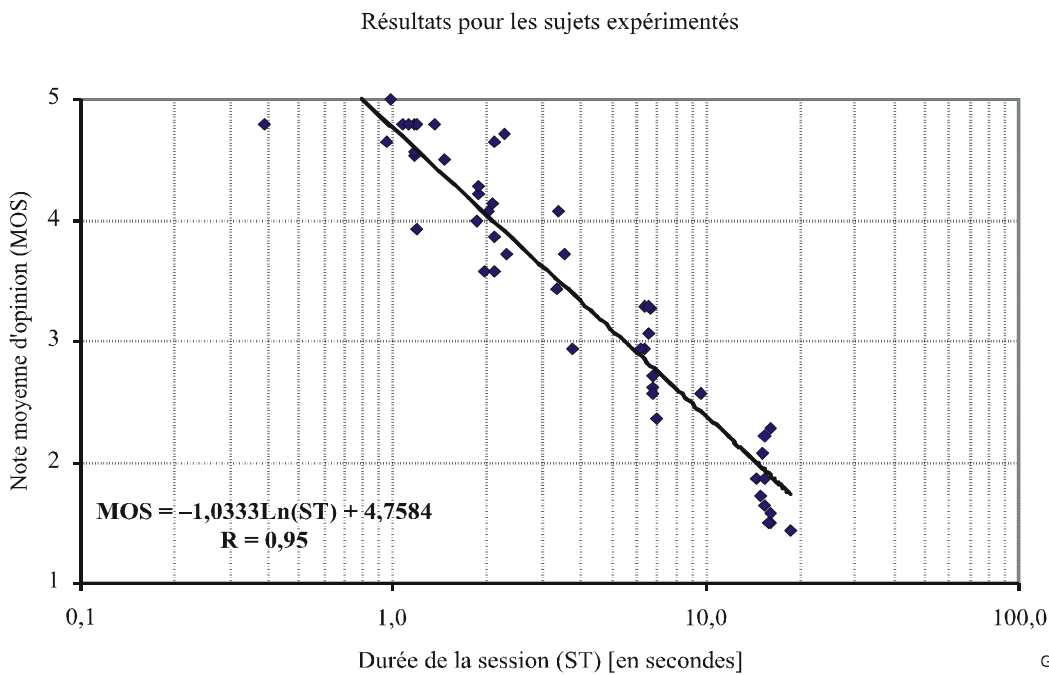
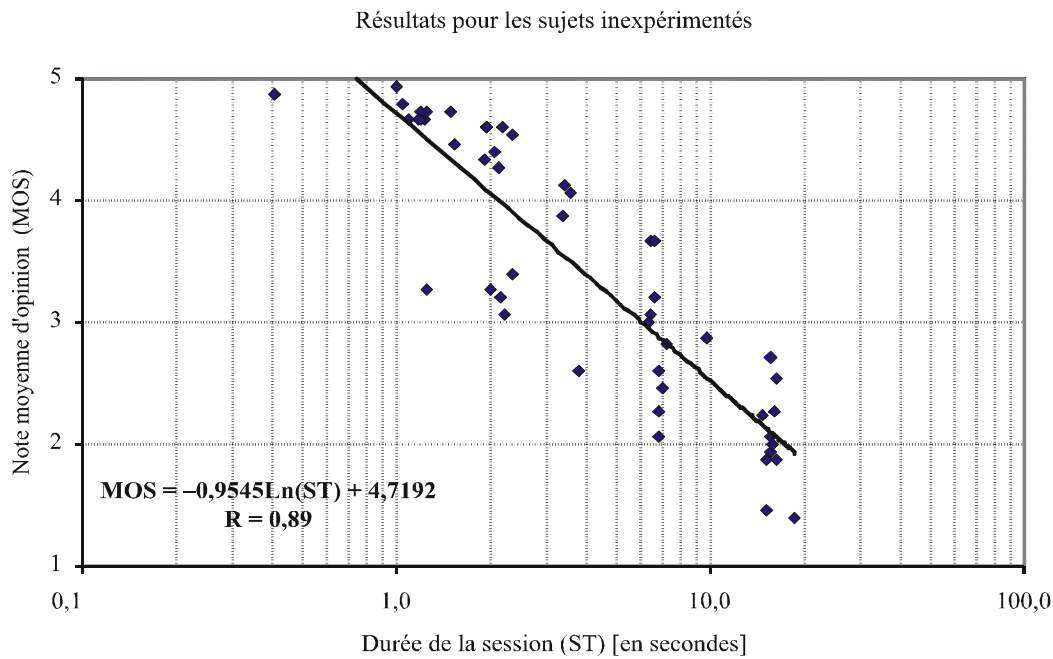


Résultats pour les sujets inexpérimentés



Résultats pour (7) sujets inexpérimentés et (18) sujets expérimentés avec l'échelle de temps de 6 secondes. Pour les sujets inexpérimentés, la corrélation est trop faible pour permettre des prévisions précises de la MOS. Pour les sujets expérimentés, la MOS peut être prévue à partir d'une interpolation logarithmique de la durée de la session entre 0,67 et 12 secondes.

Figure A.2/G.1030 – Résultats pour sujets avec l'échelle de temps de 6 secondes

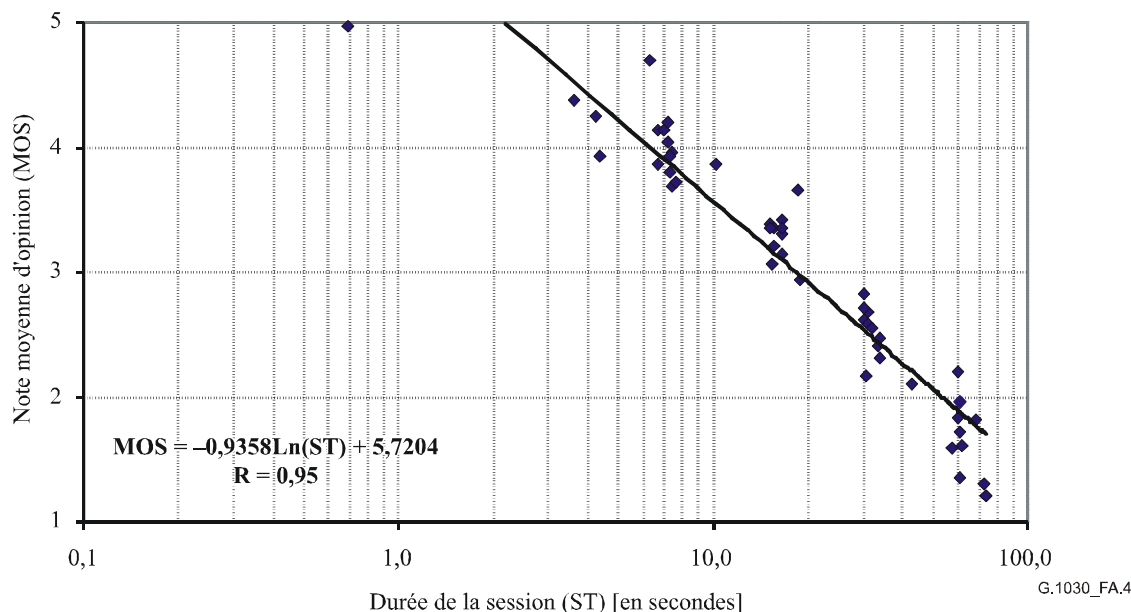


G.1030_FA.3

Résultats pour (15) sujets inexpérimentés et (14) sujets expérimentés avec l'échelle de temps de 15 secondes. Ces résultats font apparaître le même comportement. Pour les sujets inexpérimentés, la corrélation est légèrement trop faible pour permettre des prévisions précises de la MOS. Pour les sujets expérimentés, la MOS peut être prévue à partir d'une interpolation logarithmique de la durée de la session entre environ 0,79 et 38 secondes.

Figure A.3/G.1030 – Résultats pour sujets avec l'échelle de temps de 15 secondes

Résultats pour tous les sujets



Résultats pour tous les sujets (12 + 17 = 29) avec l'échelle de temps de 60 secondes. La MOS peut être prévue à partir de l'interpolation logarithmique de la durée de la session entre 2,16 et 155 secondes.

Figure A.4/G.1030 – Résultats pour sujets avec l'échelle de temps de 60 secondes

A.4 Résultats de modélisation

A.4.1 Contexte de 60 secondes

Pour le contexte de 60 secondes, les corrélations entre la durée de la session et la qualité subjective sont très bonnes (0,95) tant pour les sujets inexpérimentés que pour les sujets expérimentés et le mappage entre la durée de la session et la qualité subjective peut être directement utilisée dans le modèle objectif. La régression indique que les durées de session inférieures à environ 2 secondes donnent une valeur MOS maximale de 5 (voir la Figure A.4), cependant que la valeur MOS minimale de 1 est obtenue pour des durées de session supérieures à la durée de session la plus longue retenue pour l'essai dans la pratique. Un mappage général entre la durée de session et la qualité de navigation sur le Web pour le contexte de longue durée est établi en définissant une durée de session minimale (*Min*) et une durée de session maximale (*Max*) et en utilisant une interpolation logarithmique entre ces valeurs extrêmes de durées de session (*SessionTime*). Si nous posons: $MOS = a - b \cdot \ln(SessionTime)$, en mettant $MOS = 5$ pour $SessionTime = Min$ et $MOS = 1$ pour $SessionTime = Max$, on obtient pour les durées de session entre *Min* et *Max*:

$$MOS = \frac{4}{\ln(Min / Max)} \cdot (\ln(SessionTime) - \ln(Min)) + 5 \quad (1)$$

Pour l'essai de longue durée (voir la Figure A.4) la régression est:

$$MOS = 5,72 - 0,936 \cdot \ln(SessionTime) \quad (2)$$

les valeurs de MOS inférieures à 1 étant ramenées à 1 et les valeurs supérieures à 5 étant ramenées à 5.

A.4.2 Contextes de 6 secondes et de 15 secondes

Pour les contextes expérimentaux de 6 et 15 secondes, les valeurs de corrélation entre la durée de session et la qualité subjective sont nettement plus faibles que pour le contexte de 60 secondes, et un modèle plus élaboré pour la prévision de la qualité subjective pour les utilisateurs inexpérimentés et expérimentés est mis au point selon le principe que pour les durées de session plus brèves, le dernier temps de téléchargement (T_4 dans notre essai) a une incidence plus forte sur la qualité finale de navigation sur le web perçue que les autres temps de réponse et de téléchargement (T_1 , T_2 et T_3 dans notre essai). Le Tableau A.1 indique les facteurs de pondération qui ont été appliqués aux temps T_1 à T_4 pour obtenir une grandeur qui présente la corrélation la plus forte avec les valeurs MOS déterminées subjectivement. Cette grandeur ou durée de session pondérée (*weightedST*),

$$WeightedST = WT1 \cdot T1 + WT2 \cdot T2 + WT3 \cdot T3 + WT4 \cdot T4$$

peut être mappée à la valeur MOS en utilisant l'interpolation logarithmique entre les durées de session minimale et maximale qui a été utilisée en (1):

$$MOS = \frac{4}{\ln(Min / Max)} \cdot (\ln(WeightedST) - \ln(Min)) + 5 \quad (3)$$

Il ressort du Tableau A.1 que, pour le contexte à durée de session la plus courte, l'incidence du dernier temps de téléchargement est plus de deux fois plus importante que l'incidence des autres temps de téléchargement et de réponse. On observe également une différence sensible de comportement entre les sujets inexpérimentés et les sujets expérimentés, les coefficients de pondération optimaux pour les sujets inexpérimentés révélant une incidence plus forte du dernier temps de téléchargement que les coefficients de pondération optimaux pour les sujets expérimentés. Pour les sujets inexpérimentés, l'incidence du temps de téléchargement le plus long est plus de quatre fois plus importante que l'incidence des autres temps de téléchargement et de réponse. Pour les sujets expérimentés, cette incidence est environ deux fois plus importante, alors que pour les coefficients de pondération optimaux pour tous les sujets confondus elle est environ trois fois plus importante.

Tableau A.1/G.1030 – Pondération optimale du modèle pour T₁, T₂, T₃ et T₄ et valeurs de corrélation associées entre les résultats de temporisation objective et de MOS subjective

	WT1	WT2	WT3	WT4	<i>Min</i>	<i>Max</i>	Valeur de corrélation
Session de 6 s, sujets expérimentés	0,56	0,84	0,80	1,80			0,97
Session de 6 s, sujets inexpérimentés	0,37	0,40	0,60	2,63			0,93
Session de 6 s, tous sujets confondus	0,47	0,60	0,71	2,22	0,62	13,5	0,95
Session de 15 s, sujets expérimentés	0,63	0,77	1,11	1,49			0,98
Session de 15 s, sujets inexpérimentés	0,48	0,70	0,88	1,95			0,96
Session de 15 s, tous sujets confondus	0,54	0,72	0,98	1,76	0,81	39	0,97
Session de 60 s, sujets expérimentés	0,84	0,77	1,22	1,18			0,99
Session de 60 s, sujets inexpérimentés	0,64	1,01	1,12	1,24			0,98
Session de 60 s, tous sujets confondus	0,73	0,90	1,16	1,22	2,22	151	0,98
<i>Session de 60 s, tous sujets confondus, pas de pondération, voir Figure A.4</i>	<i>1,00</i>	<i>1,00</i>	<i>1,00</i>	<i>1,00</i>	<i>2,16</i>	<i>155</i>	<i>0,95</i>
NOTE – La somme des coefficients de pondération est normalisée à 4 pour permettre de comparer les durées de session normales (T ₁ + T ₂ + T ₃ + T ₄) avec les durées de session pondérées. La pondération utilisée dans le modèle normalisé ainsi que les durées <i>Min</i> et <i>Max</i> utilisées dans l'équation (3) sont indiquées en caractères gras. Pour les contextes à durée de session longue (>60 secondes), le modèle simple sans pondération peut être utilisé (caractères gras italiques).							

Il ressort également du Tableau A.1 que pour le contexte à durée de session intermédiaire, les facteurs de pondération pour les sujets expérimentés, les sujets inexpérimentés et pour tous les sujets confondus sont compris entre les facteurs de pondération pour les essais dans le contexte à durée de session courte et dans le contexte à durée de session longue. Cela démontre la validité de la méthode de pondération, permettant ainsi une interpolation entre les différentes *durées de session des contextes d'essai* (c'est-à-dire les échelles de temps de 6, 15 et 60 secondes) pour obtenir les coefficients de pondération applicables aux autres durées de session des contextes d'essai.

A.4.3 Résumé

La présente Recommandation définit un modèle simple normalisé pour le contexte à durée de session de 60 secondes, qui permet d'évaluer les sessions de navigation sur le Web d'une durée maximale d'environ 155 secondes au moyen de l'équation (2). La présente Recommandation définit en outre trois modèles évolués normalisés utilisant les facteurs de pondération optimaux pour tous les sujets confondus, indiqués dans le Tableau A.1, associés aux valeurs de mappage suivantes entre la durée de session pondérée (*WeightedSessionTime*) et la qualité de navigation perçue en fonction des notes moyennes d'opinion:

$$MOS = 4,38 - 1,30 \cdot \ln(WeightedSessionTime) \quad (4)$$

les valeurs de MOS inférieures à 1 étant ramenées à 1 et les valeurs supérieures à 5 étant ramenées à 5, pour les sessions de courte durée

$$MOS = 4,79 - 1,03 \cdot \ln(WeightedSessionTime) \quad (5)$$

les valeurs de MOS inférieures à 1 étant ramenées à 1 et les valeurs supérieures à 5 étant ramenées à 5, pour les sessions de durée intermédiaire

$$MOS = 5,76 - 0,948 \cdot \ln(WeightedSessionTime) \quad (6)$$

les valeurs de MOS inférieures à 1 étant ramenées à 1 et les valeurs supérieures à 5 étant ramenées à 5, pour les sessions de longue durée

Les résultats des ajustements de régression des équations (4), (5) et (6) sont indiqués dans les Figures A.5 à A.7.

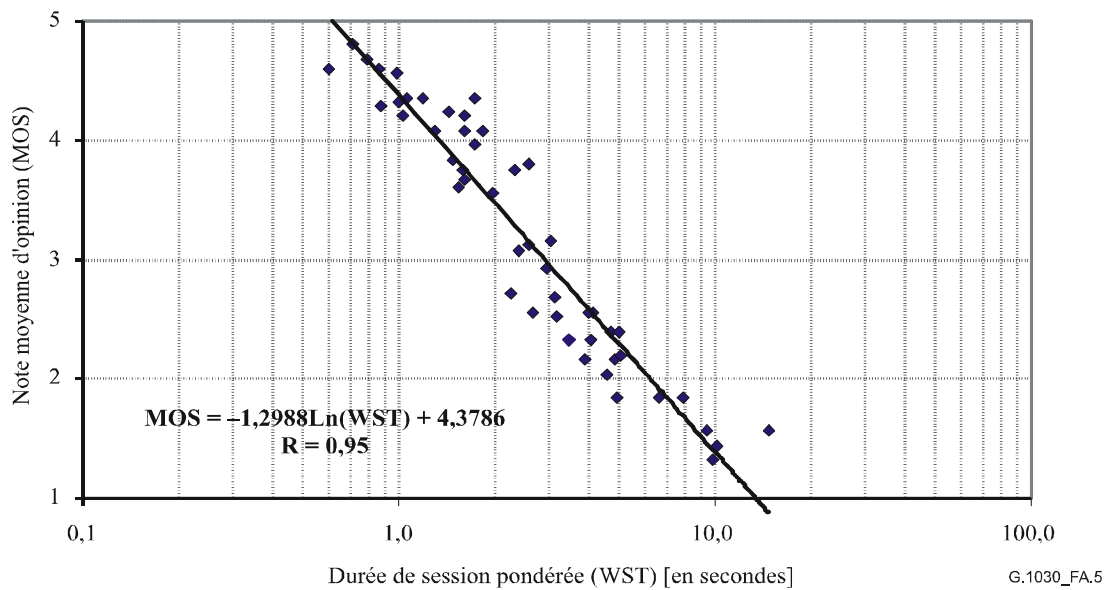


Figure A.5/G.1030 – Comparaison entre le modèle et les données pour tous les sujets avec l'échelle de temps de 6 secondes

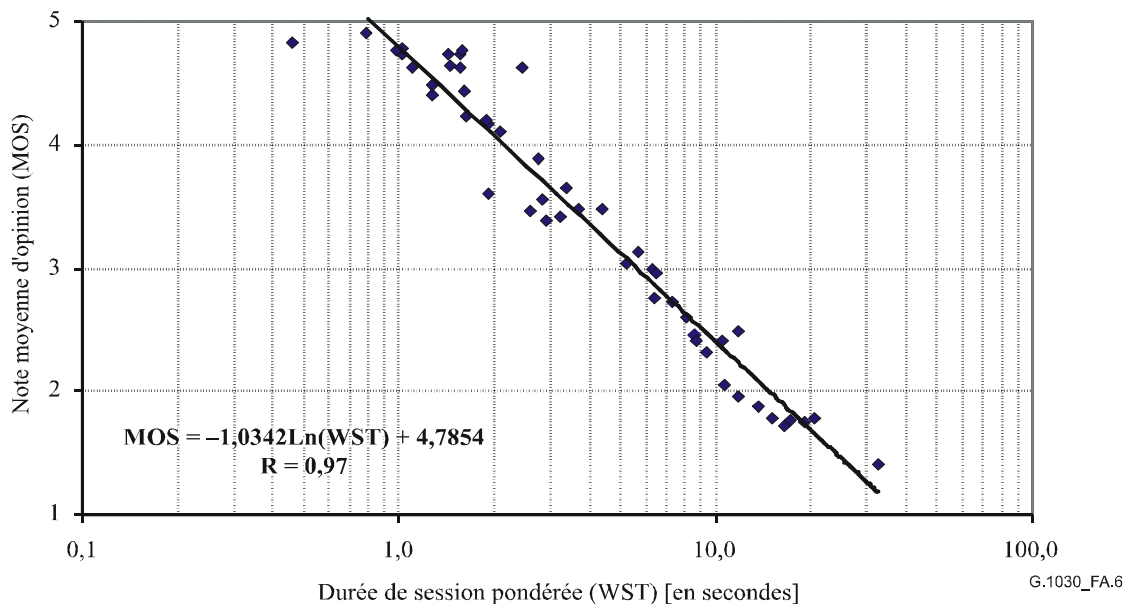


Figure A.6/G.1030 – Comparaison entre le modèle et les données pour tous les sujets avec l'échelle de temps de 15 secondes

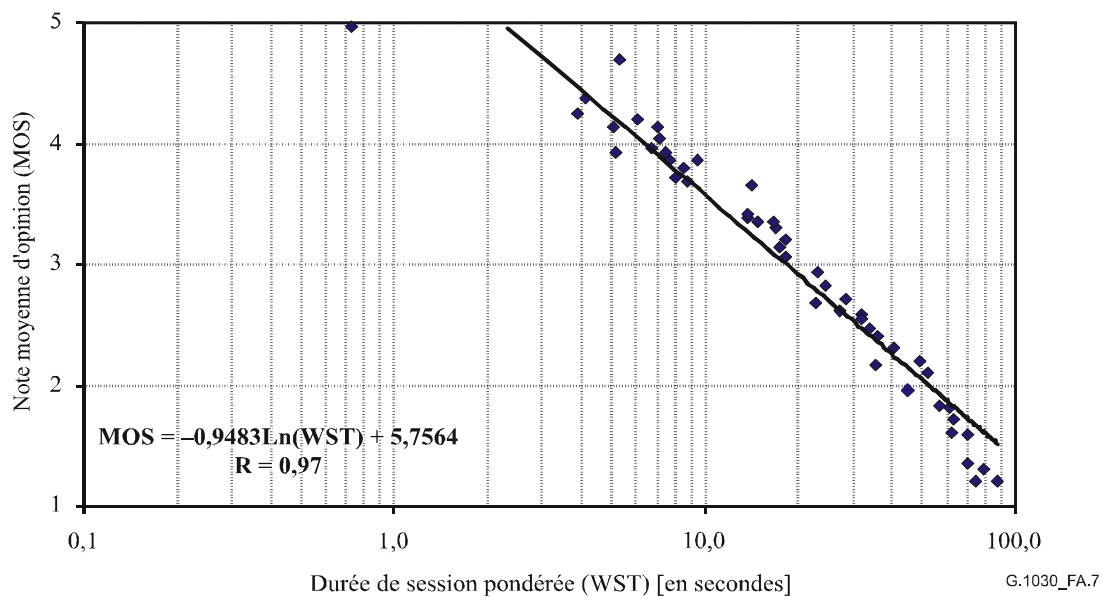


Figure A.7/G.1030 – Comparaison entre le modèle et les données pour tous les sujets avec l'échelle de temps de 60 secondes

A.5 Qualité perçue de sessions de navigation sur le web d'une page et d'événements de temporisation isolés (non normatif)

Les données d'essai sur la base desquelles ont été établies les équations (1) et (3) nous permettent également de déterminer la corrélation entre la durée de sessions de navigation sur le Web d'une seule page et la qualité perçue. D'autre part, nous examinerons également ici les sessions d'une seule page dont le temps de téléchargement est toujours égal à zéro, ce qui est le cas lorsque les informations extraites s'affichent instantanément sur l'écran de l'utilisateur après un certain temps d'attente. Cette dernière relation établit une correspondance entre un événement de temporisation isolé et la qualité perçue subjectivement en fonction d'une note moyenne d'opinion (MOS).

Le calcul des mappages établis pour des sessions d'une page commence par la constatation que la qualité diminue de manière linéaire avec l'algorithme des durées de session comprises entre une valeur minimale et une valeur maximale pour lesquelles les notes MOS sont respectivement de 5 et de 1 (voir les Figures A.2 à A.7). Il ressort des données de navigation de deux pages présentées dans les Figures A.2 à A.4 que la durée de session minimale dans les trois contextes d'essais subjectifs qui ont été retenus varie entre environ 0,7 et 2,2 et augmente à mesure que la durée maximale de l'essai augmente (voir l'état récapitulatif des résultats dans le Tableau A.2).

Tableau A.2/G.1030 – Durées de session minimale (*Min*) et maximale (*Max*) dans les modèles sans pondération pour l'essai de navigation de deux pages

	<i>Min</i> (s)	<i>Max</i> (s)
Session de 6 s, sujets expérimentés	0,67	12
Session de 15 s, sujets expérimentés	0,79	38
Session de 60 s, tous sujets confondus	2,16	155

Ces résultats nous permettent de définir une durée de session minimale (*Min*), pour laquelle une note MOS de 5 est obtenue, telle que $Min = 0,011 Max + 0,47$ où *Max* représente la durée de session maximale prévue. On peut ainsi établir un mappage général entre la durée de session et la note MOS pour l'essai de deux pages pour toute durée maximale (*Max*) prévue de la session de navigation sur le Web de deux pages:

$$MOS_{2-page} = \frac{4}{\ln((0,011Max + 0,47) / Max)} \cdot (\ln(SessionTime) - \ln(0,011Max + 0,47)) + 5 \quad (7)$$

On peut déduire des données d'essai présentées que cette relation vaudra probablement pour toutes les sessions de navigation sur le web de deux pages d'une durée comprise entre 10 et 200 secondes.

Pour les sessions d'une page arbitraires qui n'ont qu'un seul temps de réponse T_1 et un seul temps de téléchargement T_2 (voir la Figure A.1), la qualité la plus élevée est perçue pour une valeur minimale égale à environ la moitié de la valeur obtenue dans les sessions de deux pages. Cela nous amène à définir la durée de session minimale (*Min*) comme suit: $Min = 0,005Max + 0,24$, le mappage avec les valeurs MOS étant donné par l'équation (8):

$$MOS_{1-page} = \frac{4}{\ln((0,005Max + 0,24) / Max)} \cdot (\ln(SessionTime) - \ln(0,005Max + 0,24)) + 5 \quad (8)$$

On peut déduire des données d'essai présentées que cette relation vaudra probablement pour toutes les sessions de navigation sur le web d'une page d'une durée comprise entre 5 et 100 secondes.

De même, pour les sessions d'une page arbitraires dont le temps de téléchargement est toujours égal à zéro, ou pour tout événement de temporisation isolé, on obtient:

$$MOS_{single\ timing\ event} = \frac{4}{\ln((0,003Max + 0,12) / Max)} \cdot (\ln(SessionTime) - \ln(0,003Max + 0,12)) + 5 \quad (9)$$

On peut déduire des données d'essai présentées que cette relation vaudra probablement pour tous les événements de temporisation isolés d'une durée comprise entre 3 et 50 secondes. Notons que pour une interaction isolée de cette nature, la durée minimale est égale à 0,12 seconde, ce qui correspond au seuil de perception instantanée [3].

Notons qu'il convient de valider par des essais pratiques les modèles présentés ci-dessus pour l'évaluation de la qualité perçue de sessions de navigation sur le web d'une page et d'événements de temporisation isolés.

A.6 Référence à l'Annexe A

- [1] BEERENDS (J.G.), VAN DER GAAST (S.), AHMED (O.K.), Web browse quality modelling, *White contribution COM 12-C 3 to ITU-T Study Group 12*, novembre 2004.
- [2] VAN DER GAAST (S.), BEERENDS (J.G.), AHMED (O.K.), and MEEUWISSEN (H.B.), Quantification and prediction of end-user perceived web-browsing quality, submitted on 24 mars 2005.
- [3] NIELSEN (J.), Response Times: The Three Important Limits (1994). Available: <http://www.useit.com/papers/responsetime.html>
- [4] DELLAERT (G.C.), KAHN (B.E.), How Tolerable is Delay? Consumers' Evaluations of Internet Websites after Waiting (1998). Available: <http://greywww.kub.nl:2080/greyfiles/center/1998/64.html>
- [5] Recommandation UIT-T P.800 (1996), *Méthodes d'évaluation subjective de la qualité de transmission*.
- [6] Recommandation UIT-T P.862 (2001), *Evaluation de la qualité vocale perçue: méthode objective d'évaluation de la qualité vocale de bout en bout des codecs vocaux et des réseaux téléphoniques à bande étroite*.

Appendice I

Evaluations de la qualité de fonctionnement du réseau au moyen d'informations restreintes

I.1 Introduction

Le présent appendice fait le point des méthodes simplifiées que l'on peut utiliser pour évaluer la qualité de fonctionnement de bout en bout dans un réseau IP lorsqu'on dispose d'informations restreintes. La précision des évaluations effectuées à l'aide de ces méthodes dépend beaucoup de la qualité des informations à traiter. Le contenu du présent appendice est susceptible d'être modifié après complément d'étude et analyse.

I.2 Connexion de référence

Il est nécessaire de tenir compte des pare-feu, des proxys (destinés au Web et/ou à améliorer la qualité de fonctionnement), des traducteurs d'adresses de réseau (NAT, *network address translator*) ainsi que des commutateurs d'équilibrage de charge dans la connexion de référence, du fait que ces dispositifs sont présents aujourd'hui dans beaucoup de foyers, d'entreprises et de réseaux IP gérés. Ces dispositifs ou "boîtes intermédiaires" (*middleboxes*) assurent diverses fonctions, tout en éliminant la transparence de bout en bout qui constituait une caractéristique souhaitable de l'architecture de réseau IP. La Figure I.1 représente une connexion de référence appropriée.

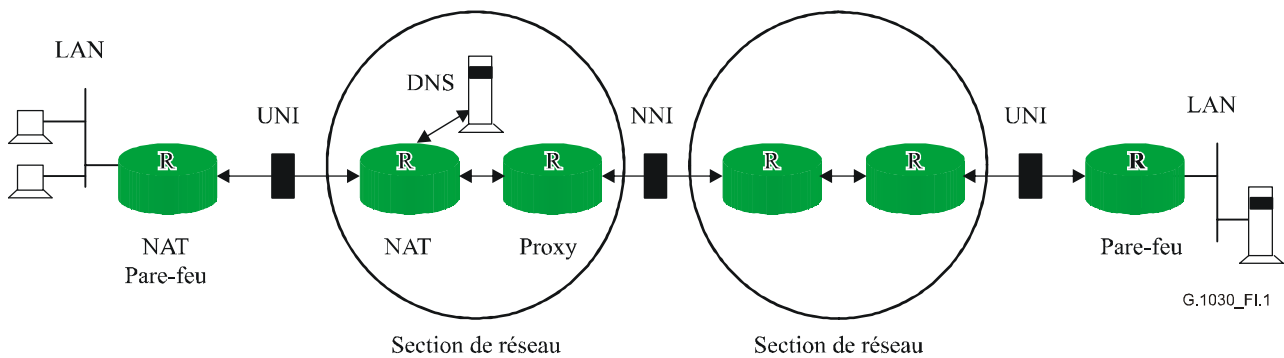


Figure I.1/G.1030 – Connexion de référence avec "boîtes intermédiaires" types

Cette extension permet d'évaluer la qualité de fonctionnement de protocoles de signalisation tels que le protocole d'ouverture de session (SIP) qui utilise la même configuration générale que les autres applications IP.

I.3 Concaténation des valeurs de qualité de fonctionnement pour le transfert de paquets

L'insertion de "boîtes intermédiaires" signifie que le trajet de bout en bout comportera un plus grand nombre de sections individuelles, d'où la nécessité de formules d'accumulation précises.

I.3.1 Délai de transfert

Les délais de transfert moyens pour les différentes sections s'ajoutent. Toutefois, il est à noter que le délai de transfert moyen représente le centre de gravité de la distribution de la variation du délai de transfert souvent associée au transport de paquets. Lorsque le délai de transfert de paquets à travers une section de réseau (ou le temps de traitement dans un serveur) est représenté par le délai de transfert moyen, un seul échantillon de la distribution du délai de transfert est alors remplacé par la valeur prévue pour celui-ci. Les valeurs moyennes du délai de transfert sont appropriées dans le cas présent, car chaque transaction échantillonnera la distribution du délai de transfert sous-jacent à de nombreux instants différents pendant l'intervalle de temps de transaction.

I.3.2 Perte

La qualité de fonctionnement de bout en bout exprimée par le rapport de pertes du protocole Internet (IPLR, *IP loss ratio*) est essentiellement la qualité de fonctionnement d'interface UNI à interface UNI. On peut considérer que les probabilités de perte des terminaux d'extrémité et des serveurs sont négligeables.

La méthode proposée pour la concaténation IPLR consiste à inverser la probabilité d'un transfert de paquets réussi à travers n sections de réseau, comme suit:

$$IPLR_{UNI-UNI} = 1 - \{(1 - IPLR_{NS1}) \times (1 - IPLR_{NS2}) \times (1 - IPLR_{NS3}) \times \dots \times (1 - IPLR_{NSn})\}$$

Cette équation, qui repose sur la théorie des probabilités conditionnelles, pose en principe que les probabilités de perte dans chaque section de réseau sont indépendantes. Pour un trajet d'interface UNI à interface UNI entre deux réseaux, A et B, où p_A et p_B sont les probabilités de perte:

Probabilité {de réussite du transfert de paquets sur les deux réseaux} = probabilité {de réussite du transfert sur le réseau B | de réussite du transfert sur le réseau A} \times probabilité {de réussite du transfert sur le réseau A}

$$= (1 - p_B) \times (1 - p_A)$$

Les probabilités de perte (p_A et p_B) calculées de manière empirique pour chacun des deux réseaux équivalent aux probabilités conditionnelles de réussite du transfert de paquets sur l'un ou l'autre des réseaux précédents – qui nécessite un événement d'entrée de paquets pour lancer l'opération de mesure.

1.3.2.1 Additivité des pertes de paquets en rafales

Ce point nécessite un complément d'étude.

1.3.3 Variation du délai de transfert

Les applications IP traditionnelles et maintes applications de transmission de données sont moins sensibles à la variation du délai de transfert des paquets que les applications nécessitant des vitesses de diffusion en continu (applications isochrones ou en temps réel, de type VoIP, par exemple). C'est la raison pour laquelle les applications IP traditionnelles sont parfois classées en catégories en fonction de l'élasticité de leurs flux de paquets, qui peuvent être compressés ou allongés durant leur transfert sans incidence sensible. En cas de modification de l'espacement des paquets VoIP en cours de transport, l'application doit renforcer la mise en mémoire tampon pour rétablir la capacité de diffusion en continu, laquelle mémoire contribue à allonger le délai de transfert des flux de paquets "non élastiques".

Il est difficile de concaténer de manière précise la variation du délai de transfert des différents processeurs et sections de réseau, étant donné que l'on connaît rarement la distribution complète du délai de transfert. Toutefois, une méthode d'approximation acceptable est indiquée au § 8/Y.1541.

1.4 Bande passante engorgée

L'utilisation de la bande passante d'accès comme facteur restrictif lors du calcul du temps de transfert des données de bout en bout peut conduire à surévaluer le débit.

1.4.1 Eléments essentiels à prendre en considération

Une approximation plus précise du temps de transfert effectif des données d'un fichier ou message de longueur fixe peut être obtenue en prenant en considération dans le modèle les éléments suivants:

- 1) le sens de transfert doit être indiqué. Les débits d'accès utilisés pour de nombreuses technologies nouvelles – telles que les lignes d'abonnés numériques (DSL) et les câblomodems – étant asymétriques, le résultat obtenu peut varier de 1 à 10, voire plus;
- 2) il convient d'utiliser la bande passante engorgée effective. Le débit d'accès ne correspond pas nécessairement au débit mesuré au point d'engorgement. Un des serveurs peut ralentir le débit de transmission (en ne transmettant pas suffisamment de données aux fins du processus TCP, ou en limitant la taille des fenêtres), les paramètres TCP du serveur peuvent limiter le débit et des "boîtes intermédiaires", de type proxys visant à améliorer la qualité de fonctionnement peuvent modifier les paramètres TCP pour accroître le débit, etc;
- 3) la transaction peut être entravée par des conflits d'accès à des ressources communes. Il convient d'utiliser la bande passante engorgée effective correspondant aux objectifs nominaux pour prendre en charge simultanément un certain nombre d'utilisateurs sur une liaison ou autre ressource commune;
- 4) l'engorgement de la bande passante peut varier dans le temps. Les limites de la bande passante détermineront une échelle de temps pour le transfert des données;
- 5) la redondance doit être calculée au point d'engorgement. Chaque couche inférieure à la couche IP ajoute une redondance d'en-tête. Il convient de retenir une taille de charge utile et de calculer le pourcentage de redondance, ce qui aura pour effet d'allonger encore le temps de transfert des données;

- 6) le protocole TCP ne permet pas d'obtenir une capacité de transfert égale à la bande passante engorgée. L'équation de Mathis appliquée au protocole TCP donne une bonne approximation de la capacité de transfert en régime permanent. Toutefois, les facteurs susmentionnés peuvent avoir une importance prédominante.

La plupart de ces éléments exigent un surcroît de préparation de la part de l'utilisateur du modèle ou du planificateur du réseau. Des précisions sur l'approximation de la capacité TCP sont données au § I.4.2.

I.4.2 Capacité TCP

Dans leur immense majorité, les applications IP traditionnelles utilisent les services de transfert de flux d'octet fiables du protocole de commande de transmission (TCP, *transmission control protocol*). Un modèle de comportement du contrôle de flux TCP [TCPCon] est décrit comme suit:

$$BW < \frac{MSS}{RTT} \frac{C}{\sqrt{p}}$$

où:

BW	correspond aux données transférées par unité de temps (durée de cycle)
MSS (<i>maximum segment size</i>)	est la taille de segment maximale pour le contrôle de flux TCP
RTT (<i>round trip time</i>)	est le temps d'aller-retour moyen
C	est une constante qui tient compte des effets de perte aléatoire/périodique et de la stratégie d'accusé de réception (ACK)
p	est la probabilité de perte de paquets

Cette équation doit être considérée comme imposant une limite supérieure au flux d'information, même pour les nouvelles versions améliorées du contrôle de flux TCP, SACK (accusé de réception sélectif) et TCP Reno. Elle pose en principe que la connexion TCP a une durée de vie suffisamment longue pour parvenir à l'équilibre dans l'état de prévention des encombrements. Lorsque le nombre total d'octets dans le transfert de données est faible, le contrôle de flux TCP ne peut parvenir à l'équilibre, et le mode de démarrage lent prévaut dans le calcul. On trouvera dans [Padhye] une relation plus souple en ce qui concerne la capacité TCP Reno ainsi que les effets de limitation de bande de la taille de fenêtre maximale.

$$B(p) \approx \min \left(\frac{W_{\max}}{RTT}, \frac{1}{RTT \sqrt{\frac{2bp}{3}} + T_0 \min \left(1, 3\sqrt{\frac{3bp}{8}} \right) p (1 + 32p^2)} \right)$$

où:

$B(p)$	est le modèle approché de débit TCP [en paquets/s]
W_{\max}	est la capacité maximale de mémoire tampon de fenêtre du récepteur [exprimée en paquets]
RTT	est le temps aller-retour [exprimé en secondes]
b	est le nombre de paquets pour lesquels un message d'accusé de réception (ACK) est reçu
p	est la probabilité qu'un paquet soit perdu

T_0 est la temporisation avant retransmission d'un paquet (perdu) n'ayant pas donné lieu à un accusé de réception [exprimée en secondes]

Le contrôle de flux TCP tente d'augmenter sa fréquence d'émission (taille de fenêtre) jusqu'à ce qu'il se heurte à un état d'encombrement (ou une liaison engorgée), état qui se manifeste par la perte de paquets. Ainsi, une certaine perte de paquets est inhérente à l'examen de l'état d'engorgement (*probing*). Le contrôle de flux de prévention des encombrements, qui consiste à diminuer de moitié la taille de la fenêtre en cas de perte d'un paquet, puis à agrandir progressivement la fenêtre à raison d'un paquet à la fois jusqu'à ce qu'un autre paquet soit perdu, a pour effet, dans la pratique, de limiter le débit à environ 75% de la taille de fenêtre/fréquence d'émission maximale. On peut modifier cette équation de manière à remédier à la perte de paquets inhérente à l'examen de l'état d'engorgement TCP.

$$\sqrt{p_{\text{Probing}}} \approx \frac{MSS}{RTT} \frac{C}{0,75 \times BW(\text{bottleneck_link})}$$

Le Tableau I.1 indique les pertes de paquets associées à un débit de transmission donné sur une liaison engorgée.

Tableau I.1/G.1030 – Pertes de paquets TCP dues à l'examen de l'état d'engorgement

C	MSS	RTT	BW (liaison)	75% BW	Délai de transfert*B	Fenêtre	p
0,866	12 000	0,08	10 000 000	7 500 000	800 000	66,666667	3,00E-04
0,866	12 000	0,08	1 536 000	1 152 000	122 880	10,24	1,27E-02
0,866	12 000	0,08	768 000	576 000	61 440	5,12	5,09E-02
0,866	12 000	0,08	384 000	288 000	30 720	2,56	2,03E-01
0,866	12 000	0,08	128 000	96 000	10 240	0,8533333	1,83E+00
0,866	12 000	0,08	64 000	48 000	5 120	0,4266667	7,32E+00
0,866	2 048	0,08	128 000	96 000	10 240	5	5,33E-02
0,866	2 048	0,08	64 000	48 000	5 120	2,5	2,13E-01

Le Tableau I.1 appelle plusieurs observations:

- la perte de paquets inhérente à l'examen de l'état d'engorgement est très faible lorsque la bande passante (BW) de la liaison engorgée est de 10 Mbit/s (bande passante de l'Ethernet ou supérieure). Un taux de perte de paquets du réseau de 10^{-4} ou supérieur aura généralement pour effet de réduire le débit proportionnellement au contrôle de flux TCP [TCPCon];
- pour les paramètres considérés (temps d'aller-retour (RTT) de 80 ms, taille de segment maximale (MSS) de 12 000 bits (1500 octets), etc.), environ un paquet sur 100 est perdu sur une liaison T_1 dans le cas où un seul flux de contrôle TCP est utilisé pour examiner l'état d'engorgement. Un taux de perte de paquets du réseau de 10^{-3} ou inférieur n'aurait pour ainsi dire aucune incidence sur le débit obtenu;
- en ce qui concerne les valeurs indiquées pour le produit du délai de transfert par la bande passante (délai de transfert \times BW) et la taille de fenêtre optimale (produit/MSS) pour le trajet, on relève deux valeurs de bande passante de liaison (128 kbit/s et 64 kbit/s) pour lesquelles la taille de la fenêtre est inférieure à 1 et le taux de perte de paquets calculé est insignifiant (> 1). Le réglage d'accord TCP étant garanti dans ces cas, on ramènera la taille de segment maximale (MSS) à 2048 bits ce qui permettra d'obtenir un résultat plus satisfaisant;

- notons que la réduction de la bande passante (BW) due à la redondance de la couche inférieure n'a pas été prise en considération ici.

La valeur obtenue pour le taux de perte de paquets du réseau sera associé à la valeur obtenue pour le taux de perte de paquets inhérent à l'examen de l'état d'engorgement comme indiqué par l'équation suivante:

$$p_{Total} = 1 - \{(1 - p_{Network}) \times (1 - p_{Probing})\}$$

Lorsque les valeurs obtenues pour ces deux taux de perte de paquets sont du même ordre de grandeur, il ne sert pas à grand-chose de les associer pour déterminer la bande passante (BW), étant donné qu'un certain nombre des pertes de paquets inhérentes à l'examen de l'état d'engorgement ne se produiront pas.

Enfin, il est à noter que notre propos ici n'est pas d'examiner de manière approfondie le réglage des paramètres TCP, mais simplement de signaler que ces paramètres peuvent limiter le débit dans certains cas. Si la taille de fenêtre maximale n'est pas assez grande pour occuper l'intégralité du trajet aller-retour, on limitera alors le débit au produit de la fenêtre par le rapport MSS/RTT (fenêtre \times MSS/RTT) [Padhye]. Bon nombre de connexions TCP ne subissent jamais aucune perte de paquets pendant toute leur durée de vie, grâce à l'utilisation de fenêtres de petites dimensions et au fait que le contrôle de flux TCP met généralement tous les paquets en file d'attente lorsqu'il est confronté à une situation d'engorgement (ce qui augmente le temps d'aller-retour (RTT) pour certains paquets – raison pour laquelle on utilise un temps d'aller-retour moyen).

1.5 Durée de la prise de contact

L'échange des paquets se déroule en deux phases distinctes:

- 1) la prise de contact;
- 2) le transfert des données (certaines applications ne comportent pas cette phase – les applications utilisant des terminaux pour transactions au point de vente par carte de crédit ou les applications de signalisation VoIP, par exemple. Pour la simplicité de notre propos, nous incluons dans cette phase le dernier échange de paquets visant à mettre fin à la connexion. Dans la matrice 3×3 traditionnelle, la déconnexion fait l'objet d'une phase distincte).

Si un paquet ou la réponse à ce paquet est perdu durant la phase initiale de prise de contact, l'émetteur attend généralement un laps de temps déterminé avant de retransmettre le message. Le temps d'attente, ou temporisation avant retransmission, qui peut être de l'ordre de 1 à 3 secondes, augmente sensiblement la durée de la phase initiale de prise de contact en cas de perte d'un paquet.

Si l'on distingue tous les temps liés au réseau (y compris ceux qui découlent de la perte d'un paquet) des temps de traitement par le serveur, on obtient deux mesures adaptées à des tâches administratives relevant d'entités séparées (réseau et équipement des locaux client (CPE)/serveurs).

Cette décomposition est relativement simple pour la phase initiale de prise de contact. Les temps de transfert des paquets et les temporisations dues à la perte de paquets sont imputés aux réseaux.

En ce qui concerne le temps de transfert des données, il pourra suffire d'indiquer ce temps et de préciser si le facteur restrictif était la perte de paquets par le réseau, le délai de transfert sur le réseau ou le traitement/les paramètres du serveur.

1.5.1 Effet de la perte de paquets pendant la prise de contact

La contribution du réseau IP à la durée de la phase initiale de prise de contact peut être déterminée comme indiqué ci-dessous. Examinons tout d'abord le cas où huit messages avec réponses doivent être échangés et où tous les transferts de paquets aboutissent. Les probabilités correspondantes à ce cas sont indiquées ci-dessous:

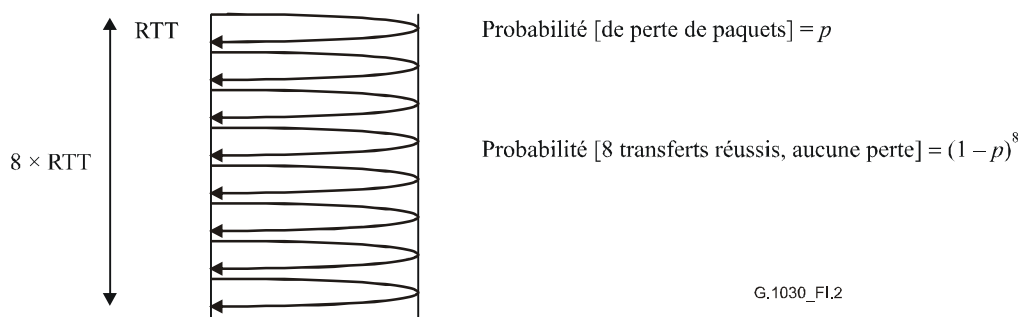


Figure I.2/G.1030 – Prise de contact avec 8 messages aller-retour et aucune perte de paquets

Il s'agit là d'un cas simple, mais qui rend bien compte de la contribution du réseau IP à la quasi-totalité des effets subis par le client lorsque la probabilité de perte de paquets aller-retour, p , est suffisamment faible. Notons que la probabilité de perte aller-retour est exprimée comme suit:

$$p = p_{RT} = 1 - \{(1 - p_{\text{dans un sens}}) \times (1 - p_{\text{dans l'autre sens}})\}$$

Le temps nécessaire pour mener à bien les huit échanges est égal à huit fois le temps d'aller-retour moyen (RTT), auquel il convient d'ajouter le temps de traitement par le serveur distant (HPT, *host processing time*), le temps total ainsi obtenu étant la durée de la prise de contact. Ces éléments peuvent être calculés séparément par l'équation suivante:

$$\text{Durée de la prise de contact (Handshake_Time)} = NCTT + \text{temps HPT total (Total_HPT)}$$

où NCTT est la contribution du réseau à la durée des transactions (NCTT, *network contribution to transaction time*), définie dans la Rec. UIT-T G.1040.

Notons que la prise de contact pour la transaction POP3 considérée ici comporte l'échange de 8 messages aller-retour (demande/réponse). Dans l'hypothèse où $p = 10^{-3}$, $RTT + HPT = 0,080$ s, RTO (temporisation avant retransmission) = 1 s et où les tentatives de transactions sont au nombre de 350 000, les probabilités pour chacun des cas de perte s'établissent comme dans le Tableau I.2.

Tableau I.2/G.1030 – Durée de la prise de contact pour les cas de perte de 0 à 3 paquets

Pertes	Durée de la prise de contact (en secondes)	Probabilité d'apparition	Transactions
0	0,64	0,99202794	347 210
1	1,64	0,00793622	2 778
2	2,64	3,5713E-05	12
3	3,64	1,1904E-07	0

Il ressort de cet exemple que le réseau IP contribue pour moins d'une seconde à la moyenne, au 95^e centile et au 99^e centile de la durée de la prise de contact pour une transaction POP3 avec échange de 8 messages aller-retour.

I.5.2 Effet de la perte de paquets pendant la phase de transfert des données

Tout paquet perdu devra être retransmis et, en cas d'utilisation du protocole TCP le contrôle de flux de prévention des encombrements réduit la fréquence d'émission, comme indiqué au § 1.4.

I.6 Exemple d'application: transaction HTTP

La Figure I.3 représente une transaction HTTP identifiant la durée de la prise de contact. L'échange des paquets restants constitue le temps de transfert des données du point de vue du client.

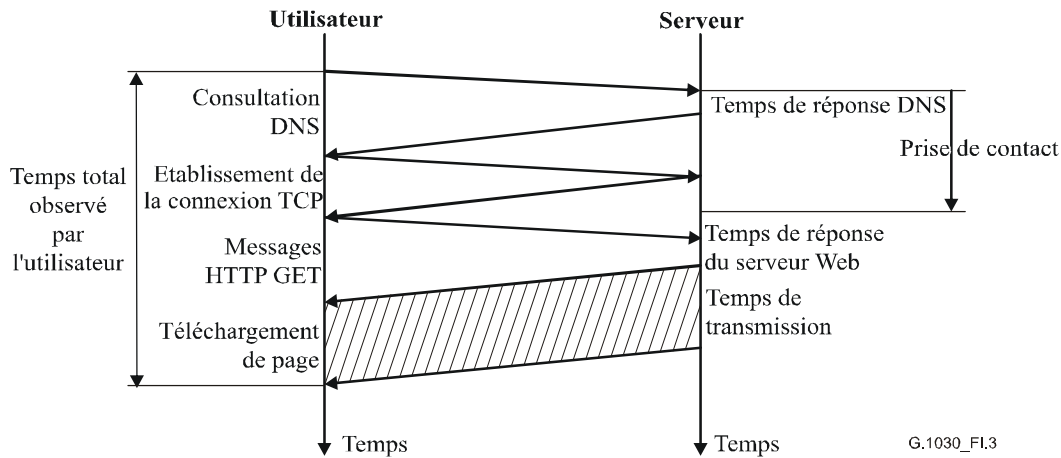


Figure I.3/G.1030 – Echange utilisant le protocole HTTP

On utilise ici la connexion de référence représentée à la Figure I.1, qui comporte deux sections de réseau, des traducteurs d'adresses de réseau (NAT), des pare-feu et un proxy faisant office de "boîtes intermédiaires" types. Il est à noter que le proxy ne met pas les objets demandés en antémémoire et qu'il adresse sa demande au serveur Web distant.

1.6.1 Durée de la prise de contact

Deux des messages aller-retour (demande/réponse) empruntent des trajets différents. Il s'agit du message de consultation du serveur DNS et du message d'établissement de la connexion TCP avec prise de contact en 3 temps (TCP 3-way handshake) (dans lequel les paquets SYN et SYN-ACK sont synchronisés, le dernier paquet ACK étant censé être transmis en même temps que le message HTTP GET, et n'augmentant pas sensiblement le temps de transmission). Bien que le proxy fractionne la connexion TCP, nous assimilons ce fractionnement, dans l'exemple considéré ici, à un délai supplémentaire sur une seule et même connexion.

L'accumulation consistant essentiellement en une simple addition ou multiplication, nous nous contenterons d'indiquer ici les valeurs et les totaux sous forme de tableau (plutôt que de reproduire intégralement les équations comportant de nombreux paramètres nouveaux).

Examinons tout d'abord le Tableau I.3, qui donne les éléments de la qualité de fonctionnement du serveur DNS.

Tableau I.3/G.1030 – Consultation du serveur DNS

Client vers DNS	RTT [ms] (réseau)	HPT [ms] (temps de traitement)	RT (1-p)	BW flux
LAN	2		0,99999	5,0E+6
NAT/Pare-feu		4	0,99999	
Liaison 1	10		0,9999	5,0E+6
NAT/Routeur périphérique			0,99999	
NAT vers liaison DNS	4		0,99999	10,0E+6
DNS		6	0,99999	
Totaux (perte inversée)	16	10	0,00015	

Examinons à présent le Tableau I.4, qui donne les éléments de l'établissement de la connexion TCP.

Tableau I.4/G.1030 – Trajet client-serveur pour la durée de la connexion TCP et le temps de transfert de données

Client vers serveur	RTT [ms] (réseau)	HPT [ms] (temps de traitement)	RT (1-p)	BW flux
LAN	2		0,99999	5,0E+6
NAT/pare-feu		4	0,99999	
Liaison UNI 1	10		0,9999	5,0E+6
NAT/Routeur périphérique			0,99999	
NAT vers liaison proxy	4		0,99999	10,0E+6
Proxy		4	0,99999	
Liaison NNI	20		0,99999	10,0E+6
Section de réseau 2	14		0,99999	10,0E+6
Liaison UNI 2	10		0,99999	3,0E+6
Pare-feu		2	0,99999	
LAN	2		0,99999	5,0E+6
Serveur (traitement TCP)		1	0,99999	
Totaux (perte inversée)	62	11	0,00021	

Pour une durée de consultation du serveur DNS de $16 + 10 = 26$ ms et une durée d'établissement de la connexion TCP de $62 + 11 = 73$ ms, la durée de la prise de contact est de 99 ms, ce qui semble constituer une durée valable pour plus de 99,9% des transactions HTTP.

I.6.2 Temps de transfert des données

Il convient dans un premier temps de déterminer si la perte client-serveur jouera un rôle dans le calcul de ce paramètre.

En se reportant une nouvelle fois au Tableau I.4, on constate que la valeur la plus faible de la bande passante (BW) par flux se produit sur la liaison UNI 2, à 3 Mbit/s. Le produit du délai de transfert par la bande passante (délai de transfert \times BW) est de $0,073 \times 3 \text{ M} = 219$ kbit/s, ce qui correspond à une taille de fenêtre optimale de 18,25 paquets pour une taille MSS de 12 000 bits (1500 octets). Le contrôle de flux TCP pouvant être réglé en fonction de ce trajet, aucune contrainte n'est associée à la taille de la fenêtre.

Pour la bande passante (BW) de la liaison engorgée, on calcule le taux de pertes de paquets inhérent au contrôle de flux TCP dues à l'examen de l'état d'engorgement comme suit:

$$\sqrt{p_{\text{Probing}}} \approx \frac{MSS}{RTT} \frac{C}{0,75 \times BW}, \text{ d'où il découle que } p = \left(\frac{12\,000}{0,073} \frac{0,866}{0,75 \times 3M} \right)^2 = 0,004003$$

La valeur calculée pour la perte du réseau (0,00021) équivalent environ à $p_{\text{Probing}}/20$, on en déduit que cette valeur est négligeable et que le débit du flux d'octets pour le contrôle de flux TCP sera de $0,75 \times 3 \text{ Mbit/s} = 2,25 \text{ Mbit/s}$.

Si, en revanche, la valeur calculée pour le taux de perte du réseau était du même ordre que la valeur calculée pour le taux de perte de paquets inhérent à l'examen de l'état d'engorgement, il ne servirait pas à grand chose dans ce cas d'associer ces valeurs pour recalculer la bande passante (BW):

$$p_{\text{Total}} = 1 - \{(1 - p_{\text{Network}}) \times (1 - p_{\text{Probing}})\}$$

Dans le cas d'une page Web comportant des graphiques complexes, le temps de transfert pour 1 Mbit est de 0,444 s, ce à quoi il convient d'ajouter le temps que met le serveur à traiter le message HTTP GET (0,01 s), ce qui porte le temps de transfert total à 0,454 s.

I.6.3 Temps total observé par l'utilisateur

Le temps total observé par l'utilisateur est la somme de la durée de la prise de contact et du temps de transfert des données, soit $0,099 + 0,454 = 0,553 \text{ s}$.

I.7 Résumé

Le présent appendice définit une méthode d'évaluation de la qualité de fonctionnement de bout en bout dans les réseaux IP. Les résultats obtenus sont à rapprocher des objectifs de qualité et de fonctionnement vus par l'utilisateur (de la Rec. UIT-T G.1010, par exemple) pour évaluer la satisfaction globale de l'utilisateur final à l'égard des différentes applications multimédias.

I.8 Références à l'Appendice I

- Recommandation UIT-T G.1010 (2001), *Catégories de qualité de service multimédia pour l'utilisateur final*.
- Recommandation UIT-T Y.1541 (2006), *Objectifs de qualité de fonctionnement pour les services en mode IP*.

[Padhye] PADHYE (J.), FIROIU (V.), TOWSLEY (D.), and KUROSE (J.), Modeling TCP Throughput: a Simple Model and its Empirical Validation, SIGCOMM 1998.
<ftp://gaia.cs.umass.edu/pub/Padhye-Firoiu98:TCP-throughput.ps.Z>

[TCPCong] MATHIS (M.), SEMKE (J.), MADAVI (J.), OTT (T.), The macroscopic behavior of TCP congestion avoidance algorithm, Computer communications review, Vol. 27, No. 3, juillet 1997, ISSN# 0146-4833.
http://www.psc.edu/networking/papers/model_ccr97.ps

SÉRIES DES RECOMMANDATIONS UIT-T

Série A	Organisation du travail de l'UIT-T
Série D	Principes généraux de tarification
Série E	Exploitation générale du réseau, service téléphonique, exploitation des services et facteurs humains
Série F	Services de télécommunication non téléphoniques
Série G	Systèmes et supports de transmission, systèmes et réseaux numériques
Série H	Systèmes audiovisuels et multimédias
Série I	Réseau numérique à intégration de services
Série J	Réseaux câblés et transmission des signaux radiophoniques, télévisuels et autres signaux multimédias
Série K	Protection contre les perturbations
Série L	Construction, installation et protection des câbles et autres éléments des installations extérieures
Série M	Gestion des télécommunications y compris le RGT et maintenance des réseaux
Série N	Maintenance: circuits internationaux de transmission radiophonique et télévisuelle
Série O	Spécifications des appareils de mesure
Série P	Qualité de transmission téléphonique, installations téléphoniques et réseaux locaux
Série Q	Commutation et signalisation
Série R	Transmission télégraphique
Série S	Equipements terminaux de télégraphie
Série T	Terminaux des services télématiques
Série U	Commutation télégraphique
Série V	Communications de données sur le réseau téléphonique
Série X	Réseaux de données, communication entre systèmes ouverts et sécurité
Série Y	Infrastructure mondiale de l'information, protocole Internet et réseaux de prochaine génération
Série Z	Langages et aspects généraux logiciels des systèmes de télécommunication