

Avis n° 2015-0832
de l’Autorité de régulation des communications électroniques et des postes
en date du 7 juillet 2015
portant sur la structure de l’usage de la bande passante des réseaux d’accès à internet
sur le territoire français

L’Autorité de régulation des communications électroniques et des postes (ci-après « l’ARCEP »),

Vu le code des postes et des communications électroniques, notamment son article L. 135 ;

Vu le courrier en date du 9 avril 2015 par lequel le ministre de l’économie, de l’industrie et du numérique a saisi l’ARCEP, pour avis, sur la structure actuelle de l’usage de la bande passante des réseaux d’accès à internet sur le territoire français, ainsi que sur les modalités de mesure de cette dernière ;

Après en avoir délibéré le 7 juillet 2015,

Formule l’avis suivant :

Par un courrier du 9 avril 2015, enregistré par l’ARCEP le 13 avril 2015, le ministre de l’économie, de l’industrie et du numérique a sollicité l’avis de l’ARCEP sur la « *structure actuelle de l’usage de la bande passante des réseaux d’accès à internet sur le territoire français, ainsi que sur la possibilité et les moyens de recueillir un ensemble d’informations permettant une connaissance précise de la nature et de l’origine de ces flux [...]* ».

L’Autorité est en particulier interrogée sur (i) les caractéristiques techniques pertinentes pour mesurer l’occupation de la bande passante, (ii) les données disponibles sur les services en ligne à destination de consommateurs établis sur le territoire français qui utilisent le plus de ressources en bande passante, (iii) les méthodes pouvant être mobilisées pour assurer le recueil et l’enrichissement de ces données et enfin, (iv) les stratégies pouvant être mises en œuvre par les fournisseurs de service en ligne pour compliquer ou dégrader la collecte de telles informations.

I Introduction

Le développement d’internet au cours des deux dernières décennies en a fait un trait d’union pour plusieurs milliards d’êtres humains et un facteur de développement social et économique. En tant que plateforme mondiale ouverte, internet constitue aujourd’hui un bien collectif structurant, dont le bon fonctionnement constitue un enjeu crucial.

En vue de préparer le présent avis, l’Autorité a mené un cycle de rencontres avec les acteurs du secteur l’ayant conduit à auditionner huit acteurs clefs d’internet en France¹. Un questionnaire a également été envoyé à plusieurs acteurs visant, notamment, à obtenir des renseignements techniques et des éléments quantitatifs susceptibles d’enrichir l’avis.

Cet avis constitue donc une contribution technico-économique de l’ARCEP aux réflexions du Gouvernement sur la possibilité d’établir une analyse fiable et précise des flux acheminés au sein des réseaux d’accès à internet en France. Il ne porte pas sur l’utilisation ultérieure éventuelle qui pourrait être faite des résultats de mesure.

L’ARCEP a organisé sa réflexion autour des deux caractéristiques essentielles – et dans une certaine mesure interdépendantes – d’un dispositif de mesure du trafic :

- **la profondeur d’analyse** – du simple comptage du volume écoulé à l’identification de l’origine ou de la nature du trafic ;
- **le point de mesure** – lieu du réseau où est réalisée la mesure.

De manière générale, l’Autorité estime que la pertinence d’un dispositif de mesure de l’usage de la bande passante des réseaux d’accès à internet doit être évaluée au regard de plusieurs critères : la faisabilité technique et juridique, le coût et la complexité de mise en œuvre, la fiabilité, la complétude, le caractère facilement certifiable des résultats obtenus.

Après quelques considérations préliminaires sur le fonctionnement d’internet (section II), sont présentés et évalués – analyse comparée avantages / inconvénients – les différents points du réseau où peut être réalisée la mesure de trafic et les différentes profondeurs d’analyse pouvant être envisagées (section III). Est ensuite présenté l’impact des différents modes d’acheminement du trafic sur internet et les mesures de contournement pouvant être mises en place par des acteurs ne souhaitant pas que leurs flux soient mesurés ou identifiés (section IV).

II Considérations préliminaires sur le fonctionnement d’internet

Dans cette section sont exposées des considérations techniques sur le fonctionnement d’internet pertinentes pour la mesure de l’occupation de la bande passante sur les réseaux d’accès à internet.

Pour plus d’informations concernant notamment les modalités d’interconnexion, les grandes catégories d’acteurs et les différents modes d’acheminement du trafic sur internet, l’Autorité renvoie à son rapport au Parlement et au Gouvernement sur la neutralité d’internet de 2012².

II.1 Trafic et tendances d’évolution

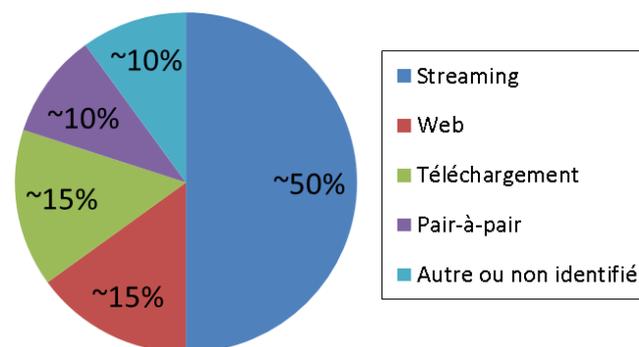
Avec le développement et la généralisation des réseaux d’accès à internet haut et très haut débit, des acteurs proposant des usages et des services innovants se sont multipliés. Ces nouveaux usages se sont accompagnés d’une augmentation, à un rythme soutenu, du trafic

¹ Deux fournisseurs d’accès à internet, deux fournisseurs de contenus et d’applications, un gestionnaire de point d’échange internet, un transitaire international, un hébergeur et une société spécialisée dans l’analyse des performances réseau.

² En particulier l’annexe 4, p.84 et l’annexe 6, p.93.

échangé qui a, à son tour, incité au développement de réseaux plus performants. Au niveau mondial, il est estimé que ce trafic devrait continuer de croître, chaque année au cours des prochaines années³, de près d'un cinquième pour les réseaux fixes et de moitié pour les réseaux mobiles. Les tendances en France devraient globalement suivre celles prévues au niveau mondial.

La nature des échanges qui se déroulent sur internet a également profondément changé. En proportion, la navigation web est devenue largement minoritaire, supplantée notamment par la diffusion de flux audiovisuels, très consommateurs de bande passante. Les flux vidéo connaissent en effet un essor tel qu'ils représentent désormais, selon les opérateurs, plus de la moitié du trafic acheminé et sont devenus le principal moteur de sa croissance.



Répartition des différents usages dans la totalité du trafic acheminé⁴

Cette évolution s'accompagne par ailleurs d'un profond changement de paradigme. En effet, alors qu'internet s'est initialement développé de manière décentralisée, sur la base d'échanges directs entre utilisateurs, la dernière décennie a vu se développer des usages, notamment dans le domaine audiovisuel, fondés sur un modèle client-serveur.

Par ailleurs, le trafic s'est progressivement concentré autour de quelques sources majeures, résultant de la concentration des principaux fournisseurs de contenus et d'applications (FCA) qui ont émergé à l'échelle mondiale. A titre d'exemple, les informations fournies à l'ARCEP dans le cadre de la préparation de cet avis révèlent que plus de la moitié du trafic acheminé sur les réseaux d'accès à internet en France est remis par seulement 5 systèmes autonomes⁵ (correspondant à des fournisseurs de services, hébergeurs ou intermédiaires techniques), sur les 60 000 qui composent internet.

Ces deux phénomènes de hausse constante et de concentration du trafic ont provoqué un changement majeur dans le profil du trafic, qui se dirige désormais majoritairement dans le sens descendant, des principaux FCA mondiaux vers les internautes.

³ Source : CISCO – *Visual Networking Index Forecast*, février 2015.

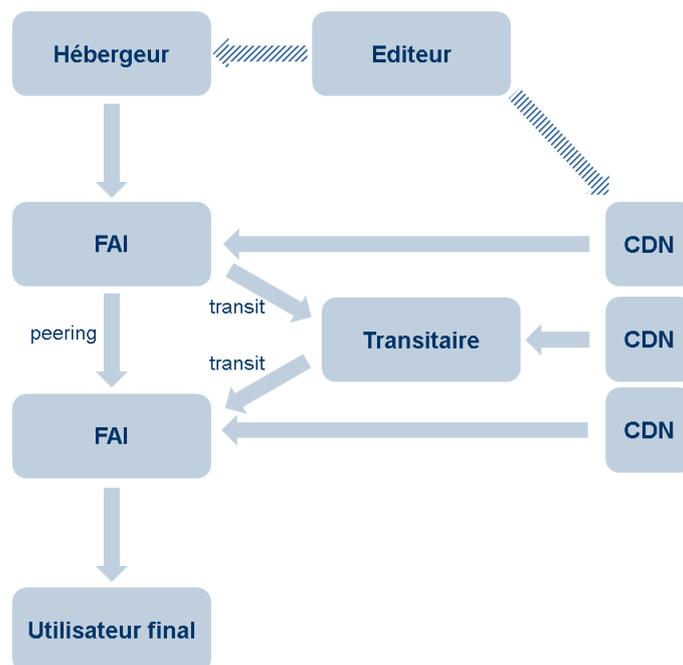
⁴ Source : réponses au questionnaire adressé par l'ARCEP.

⁵ Un système autonome (en anglais « *autonomous system* » ou AS) constitue la brique élémentaire sur internet et désigne un réseau, ou un ensemble d'adresses IP, sous le contrôle d'une entité unique.

II.2 Modes d'acheminement du trafic

Afin d'accompagner les mutations décrites *supra*, les modes d'acheminement du trafic sur internet se sont transformés. Sont notamment apparus différents types d'intermédiaires techniques dédiés à l'amélioration et à l'optimisation de la diffusion des contenus audiovisuels, ainsi que des solutions techniques visant à rapprocher les contenus de l'utilisateur final, au moyen de serveurs de cache ou de CDN⁶ (hébergeant une copie de ces contenus).

Ces modes d'acheminement peuvent se combiner, comme illustré sur le schéma suivant.



Modes d'acheminement du trafic sur internet (illustration)

II.3 Difficultés liées à l'identification des « fournisseurs de services en ligne »

Internet est caractérisé par une décorrélation forte entre, d'une part, les contenus, services et applications fournis et, d'autre part, les réseaux qui assurent leur diffusion de manière majoritairement indifférenciée.

A cet égard, il convient de souligner que la notion de « fournisseur de service en ligne » mentionnée par la lettre de saisine, ne fait pas, en tant que telle, l'objet d'une définition sur le plan juridique.

Par ailleurs, si, ainsi que le note le Conseil d'Etat dans son étude annuelle 2014, *Le numérique et les droits fondamentaux*, la loi pour la confiance dans l'économie numérique organise une *summa divisio* entre les éditeurs de site et les intermédiaires techniques tels que les hébergeurs, qui se fonde sur l'idée que ces derniers ont un rôle purement passif sans

⁶ Content delivery networks.

connaissance des informations qu'ils stockent, cette démarcation est de plus en plus difficile à réaliser dans la pratique.

A titre d'illustration, pour une même plateforme de vidéos en ligne, il peut être complexe d'identifier l'éditeur du contenu, qui peut être, selon les cas, l'exploitant de la plateforme ou un tiers.

Compte-tenu des incertitudes mentionnées ci-dessus et indépendamment de toute considération technique, il n'existe donc pas de méthode univoque, infaillible ou exhaustive permettant d'associer un paquet IP circulant sur internet à un fournisseur de service en ligne, ou respectivement, d'identifier l'ensemble des paquets IP associés à un fournisseur de service en ligne donné.

Par ailleurs, sur le plan technique, une analyse approfondie des flux acheminés et de leur contenu (de type *Deep packet inspection* ou DPI, cf. *infra*) peut permettre une connaissance fine de la nature d'un service. Une analyse au cas par cas pourrait alors permettre de déterminer le ou les fournisseurs de services pertinents, en fonction de la définition retenue pour cette notion. Toutefois, le recours croissant au chiffrement des données – visant à garantir l'intégrité et la confidentialité des échanges – peut rendre cette analyse inopérante en pratique.

Or, selon les informations recueillies par l'ARCEP auprès de différents acteurs dans le cadre de la préparation de cet avis, la part du trafic chiffré sur internet, vue par les FAI⁷ en France, représenterait près de 50 % à mi-2015, contre 5 % environ en 2012⁸.

III Mesure et identification des caractéristiques du trafic sur internet

Dans cette section, après l'exposé de quelques éléments de métrologie (III.1), sont présentés et comparés les différents niveaux d'analyse selon lesquels des mesures peuvent être effectuées (III.2) puis les différents points du réseau au niveau desquels un dispositif de mesure peut être mis en place (III.3).

Les avantages et les limites de chaque méthode sont évalués au regard des critères énoncés en section I.

III.1 Eléments de métrologie

III.1.1 Les grandeurs pertinentes

Afin de dimensionner leurs réseaux (liens et équipements actifs), les opérateurs se fondent le plus souvent sur un indicateur de débit (généralement exprimé en Gbit/s) des flux acheminés sur un ou plusieurs liens donnés. Cet indicateur traduit la quantité de données écoulée sur un lien, ou sur le réseau tout entier par unité de temps. Il permet donc de tenir compte des possibles pics de trafic dus à une intensification des usages. En pratique, une augmentation de capacité est ainsi planifiée, puis réalisée, par l'opérateur dès lors que le débit observé atteint

⁷ Fournisseur d'accès à internet.

⁸ Source : Réponse au questionnaire adressé par l'ARCEP.

une certaine proportion de cette capacité sur une partie de son réseau. L'indicateur de débit reflète donc les investissements consentis par les fournisseurs d'accès.

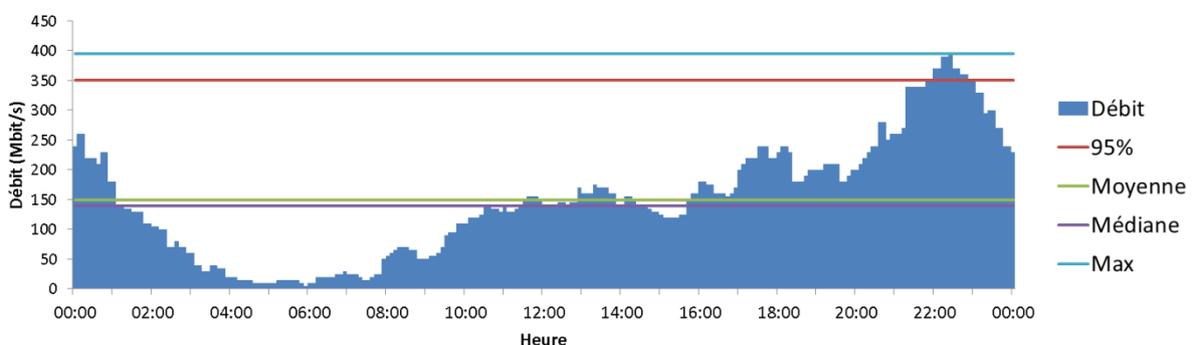
L'indicateur de volume, en téraoctets⁹ (To), traduit quant à lui la quantité totale de données ayant emprunté le réseau sur une période donnée. Bien qu'apportant un éclairage sur la consommation des utilisateurs finals, il semble moins pertinent pour la quantification de l'occupation de la bande passante. En effet, sans référence à la dimension temporelle, il ne permet pas à un opérateur de dimensionner correctement son réseau et n'est donc pas représentatif des coûts récurrents et des investissements consentis par ce dernier. Il n'est d'ailleurs en général pas suivi par les FAI sur les réseaux fixes.

III.1.2 Fréquence d'échantillonnage et valeur de référence

La fréquence d'échantillonnage du trafic correspond à l'intervalle de temps séparant deux prises de mesures. Plus la fréquence d'échantillonnage est élevée, plus les mesures permettent de représenter fidèlement l'allure du trafic acheminé. Des capacités de stockage et de traitement importantes seraient cependant nécessaires pour gérer le volume de données générées. Un échantillonnage peu fréquent – moins coûteux – risquerait quant à lui de manquer les pics ou les creux soudains caractéristiques du profil du trafic sur internet.

Afin d'être représentative de l'utilisation de la bande passante sur un réseau, la valeur de référence retenue dans le cadre d'un dispositif de mesure devrait inclure les périodes de la journée dites « chargées »¹⁰. En effet, retenir comme référence ces périodes au cours desquelles les ressources en bande passante sont sollicitées de manière intensive permet à un FAI de garantir une capacité d'acheminement suffisante y compris durant les périodes de plus forte intensité du trafic.

Selon les informations recueillies par l'ARCEP, la méthode la plus répandue chez les opérateurs, notamment à des fins de dimensionnement ou encore de facturation de certaines prestations d'acheminement, consiste à mesurer les volumes écoulés sur des intervalles de cinq minutes et d'en déduire le débit moyen sur ces intervalles. La valeur de référence journalière ou mensuelle retenue est alors le 95^{ème} centile de ces débits. Cette valeur est représentative de l'occupation de la bande passante d'un réseau en période de très forte charge, tout en écartant les mesures correspondantes aux pics les plus élevés, qui correspondent à des phénomènes rares et de courte durée.



⁹ 1 téraoctet = $8 \cdot 2^{40}$ bits $\approx 8 \cdot 10^{12}$ bits.

¹⁰ Également appelées heures « pleines » ou heures « de pointe », par opposition aux périodes dites « creuses », moins intenses en termes d'utilisation des ressources du réseau.

Illustration de l'évolution du trafic (débit) au cours d'une journée type

Pour plus de précision sur ces éléments de métrologie, il convient de se reporter, notamment, aux travaux de l'Union internationale des télécommunications (UIT) et l'Internet engineering task force (IETF) sur ces questions.

III.1.3 Asymétrie du trafic

Deux types de flux peuvent être distingués chez un FAI : les flux sortants, émis par ses clients, et les flux entrants, qui leur parviennent en retour. Le développement, au cours de ces dernières années, des services fondés sur le modèle client-serveur se traduit aujourd'hui par une asymétrie marquée entre ces flux entrants et sortants, qui s'accroît à mesure que des contenus plus riches et consommateurs en bande passante (par exemple les contenus audiovisuels, à haute définition) sont mis à disposition des internautes. Elle peut atteindre un rapport de un à vingt, voire d'avantage, entre les flux sortants et entrants au niveau des points d'interconnexion entre les FAI et les fournisseurs de contenus à plus forte audience¹¹.

Dans une logique de dimensionnement d'un réseau (ou dans les accords de transit, par exemple), les liens de transmission étant bidirectionnels, la valeur de référence retenue est la valeur la plus élevée entre les flux entrants et sortants. Aussi, dans le cadre de l'estimation de l'occupation de la bande passante d'un réseau d'accès à internet en France, il apparaît possible de négliger le trafic correspondant aux flux sortants par rapport au trafic correspondant aux flux entrants.

III.2 Caractérisation des flux sur internet : les différents niveaux d'analyse

La mesure de l'occupation de la bande passante des réseaux d'accès à internet peut être effectuée principalement selon trois niveaux d'analyse. Il s'agit, par ordre croissant de finesse et de complexité, de :

- la relève des volumes de données écoulées sur un lien grâce aux compteurs intégrés dans les équipements réseau et récupérés *via* le protocole SNMP. Une consultation de ces compteurs à intervalles réguliers et rapprochés permet de déduire le débit auquel s'écoule le trafic ;
- l'identification de la source et de la destination des flux, grâce aux outils de type Netflow/IPFix qui permettent de recueillir les adresses renseignées en en-tête des paquets IP et donc d'en déduire l'AS d'origine ;
- l'analyse du contenu des paquets transportés, grâce aux techniques et aux sondes de type DPI, qui permettent d'obtenir une information plus riche sur la nature des services transportés.

¹¹ Estimation réalisée à partir des données d'interconnexion recueillies périodiquement par l'ARCEP au titre de la décision n° 2012-0366 de l'ARCEP relative à la mise en place d'une collecte d'informations sur les conditions techniques et tarifaires de l'interconnexion et de l'acheminement de données, modifiée par la décision n° 2014-0433-RDPI.

Niveau d'analyse	Outils utilisés	Avantages	Limites
Relève du trafic écoulé	<ul style="list-style-type: none"> - Solutions logicielles qui collectent grâce au protocole SNMP les informations recueillies en continu par les différents équipements réseau (commutateurs, routeurs, etc...) 	<ul style="list-style-type: none"> - Données fiables. - Informations certifiables et exhaustives (à l'échelle de l'interface d'un équipement réseau). - Protocole SNMP normalisé (RFC 3411 et suivants) et couramment utilisé. - Outils peu coûteux et mise en œuvre aisée. 	<ul style="list-style-type: none"> - Aucune forme d'identification des flux (origine, destination, nature du service, etc...). - Identification de l'acteur interconnecté¹².
Identification de l'origine et de la destination (AS) du trafic	<ul style="list-style-type: none"> - Logiciels fondés sur les protocoles Netflow / IPFix permettant de récupérer et d'agréger les informations concernant les flux circulant par un routeur donné. 	<ul style="list-style-type: none"> - Permet d'identifier l'adresse IP à l'origine des flux et d'en déduire l'AS d'origine (idem pour la destination). - Permet de connaître les ports utilisés et d'en déduire l'utilisation de certaines applications répandues (navigation web, courriel, etc.). - Protocole IPFix standardisé (RFC 7011). - Compatible avec une grande variété de routeurs et de systèmes d'administration. - Utilisation répandue et relative simplicité de mise en œuvre. 	<ul style="list-style-type: none"> - Résultats statistiques établis sur un échantillonnage des flux (dans son utilisation actuelle consistant le plus souvent à un échantillonnage de 1:1000). - Coûts importants et risque de dégradation des performances des équipements en cas d'échantillonnage trop fréquent. - L'identification de l'AS d'origine des flux ne permet pas systématiquement d'identifier l'éditeur des contenus.
Identification de la nature et de l'éditeur des flux	<ul style="list-style-type: none"> - Cartes de service installables dans les routeurs. - Solutions logicielles déployables sur les équipements réseau. - Sondes matérielles installables en sortie des équipements réseau. 	<ul style="list-style-type: none"> - Permet d'identifier l'adresse IP à l'origine des flux et d'en déduire l'AS d'origine. - Permet d'identifier la nature des flux et donc le type d'usage associé. 	<ul style="list-style-type: none"> - Résultats statistiques établis sur un échantillonnage des flux - Risque de dégradation des performances des équipements en cas d'échantillonnage trop fréquent. - Solutions très coûteuses à mettre en place à l'échelle du réseau d'un opérateur : investissements et coûts d'exploitation importants en capacité de stockage et

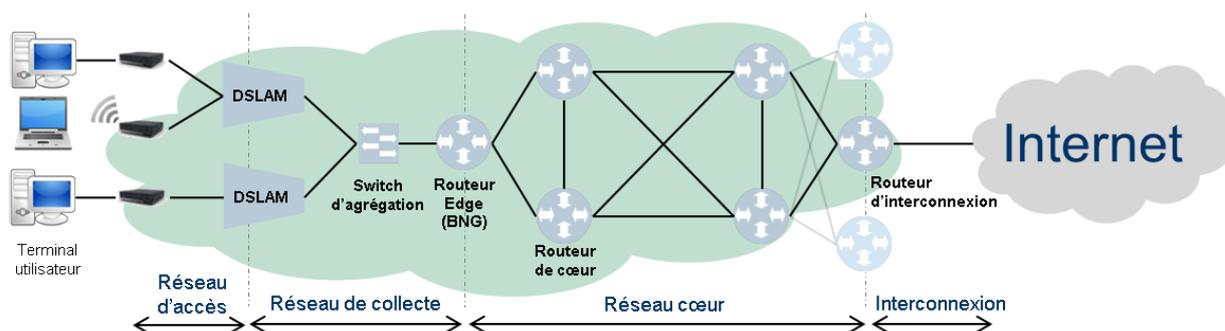
¹² Dans le cas d'un *peering* public au sein d'un point d'échange internet les données obtenues ne permettraient pas de préciser qui, parmi les membres du point d'échange, est à l'origine ou destinataire d'un paquet IP.

			<p>de traitement.</p> <ul style="list-style-type: none"> - Caractère intrusif de l'analyse effectuée, qui peut soulever des interrogations en matière notamment de protection des données personnelles. - Résultats très dégradés en cas de chiffrement des données.
--	--	--	--

III.3 Les différents points de mesure du trafic

Les flux acheminés par un FAI traversent différents niveaux de son réseau entre leur lieu de prise en charge et l'utilisateur final. Traditionnellement, quatre segments de réseau sont identifiés :

- le réseau d'accès (boucle locale), i.e. le tronçon situé entre l'abonné et le premier équipement actif d'accès au réseau¹³ ;
- le réseau de collecte (dit également réseau de transport) constitue le premier niveau d'agrégation du trafic issu des utilisateurs finals. Il a pour rôle d'acheminer le trafic entre les équipements d'accès et les premiers routeurs (BNG/BRAS¹⁴) qui gèrent l'authentification et l'accès au service et font l'interface avec le cœur de réseau ;
- le cœur de réseau, au niveau duquel le trafic est aiguillé à l'échelon national, voire international, selon les cas jusqu'en bordure du réseau. Il est constitué d'un petit nombre de routeurs de très forte capacité formant un maillage très résilient. Le cœur de réseau constitue l'épine dorsale, dite « *backbone* », du réseau d'un FAI ;
- l'interconnexion, qui est constituée des routeurs dits d'interconnexion et des faisceaux marquant la frontière du réseau du FAI avec le reste des réseaux constituant internet.



Architecture schématique d'un réseau d'accès à internet
(cas d'un réseau fixe avec boucle locale de cuivre)

Le trafic peut être mesuré à l'interface entre chacun de ces segments, laquelle est marquée par une catégorie d'équipements actifs pouvant accueillir un dispositif de mesure du trafic. Au

¹³ Par exemple : DSLAM (*Digital Subscriber Line Access Multiplexer*) pour les réseaux d'accès en cuivre, CMTS (*Cable modem termination system*) pour les réseaux d'accès câblés, OLT (*optical line terminal*) pour les réseaux d'accès en fibre optique, Node B pour les réseaux d'accès mobiles 3G et eNode B pour les réseaux d'accès mobiles 4G.

¹⁴ *Broadband remote access server / Broadband network gateway.*

niveau de chacun de ces points, il est, en dehors des exceptions listées ci-dessous, possible de recueillir des mesures selon les différents niveaux d'analyse et de finesse discutés en III.2.

	Avantages	Limites
Mesure au niveau des premiers équipements actifs (accès)	<ul style="list-style-type: none"> - Les mesures effectuées au niveau de l'accès sont exhaustives puisque tous les flux individuels empruntent nécessairement ce segment du réseau. 	<ul style="list-style-type: none"> - Le nombre de points de mesure est très élevé (de l'ordre de la dizaine de milliers) ce qui induit des coûts de mise en œuvre plus élevés, notamment en terme de nombre de licences pour les dispositifs de mesure. - Les équipements d'accès étant généralement dédiés à la transmission et au multiplexage, ils ne sont pas adaptés à une analyse fine du trafic. L'installation de matériel spécifique pourrait s'avérer nécessaire.
Mesure au niveau des routeurs de bordure (collecte)	<ul style="list-style-type: none"> - Le nombre de points de mesure est raisonnable à l'échelle d'un opérateur (de l'ordre de quelques centaines). - Les mesures sont effectuées au niveau des plateformes d'accès aux services (BRAS, BNG), qui agrègent et contrôlent le trafic en provenance des utilisateurs finals : elles sont donc empruntées par l'ensemble du trafic des abonnés. - C'est au niveau de la collecte qu'on observe la plus forte corrélation entre volume de trafic acheminé et coûts réseaux. 	<ul style="list-style-type: none"> - Certains flux acheminés de manière agrégée ne sont démultipliés en flux individuels qu'au niveau de l'accès et ne sont donc mesurés que de manière partielle au niveau de la collecte.
Mesure au niveau des routeurs d'interconnexion	<ul style="list-style-type: none"> - Le nombre de points de mesure est réduit (de l'ordre de quelques dizaines). - Le suivi du trafic acheminé par les routeurs d'interconnexion est une pratique répandue chez les opérateurs. - Chaque lien physique identifie un acteur interconnecté (opérateur de transit, point d'échange internet, acteur interconnecté au moyen d'un <i>peering</i> privé). 	<ul style="list-style-type: none"> - Certains types de flux qui sont acheminés de manière agrégée au niveau de l'interconnexion et démultipliés plus bas dans le réseau ne sont mesurés que de manière partielle. - Certains flux n'empruntent pas l'interconnexion et ne sont donc pas mesurés.

III.4 Conclusion partielle

Tout d'abord, l'Autorité note que les techniques dites de DPI, qui permettent d'analyser de manière fine la nature des flux, doivent être écartées, sans qu'il soit nécessaire d'examiner plus précisément leur faisabilité technique ou juridique, car, dans un contexte de recours de plus en plus généralisé aux solutions de chiffrement des échanges (comme exposé à la section II.3), elles ne permettent pas d'obtenir une vision suffisamment complète du trafic.

Une première méthode s'appuyant sur le protocole Netflow/IPFix et mise en œuvre au niveau des plateformes d'accès aux services situées en bordure du réseau de collecte pourrait être envisagée. Cette méthode permettrait d'englober plus de trafic qu'au niveau des points d'interconnexion et d'opérer un premier niveau d'identification en analysant l'origine et la destination des flux. Elle reste cependant fondée sur un échantillonnage statistique du trafic qui ne permet par une certification aisée des résultats et présente des risques de dégradation des performances des équipements réseau en cas d'utilisation généralisée et intensive.

Une deuxième méthode pourrait consister à relever les compteurs entretenus en continu par les équipements réseaux (grâce au protocole SNMP). Appliquée au niveau des points d'interconnexion, cette méthode, à défaut de pouvoir identifier l'origine, la destination et la nature des flux, permet de déterminer l'acteur qui l'a remis au FAI, grâce à la connaissance préalable qu'ont les FAI des partenaires interconnectés. Elle présente également plusieurs avantages en termes de fiabilité des résultats et surtout de faisabilité pour les FAI. L'Autorité considère donc que cette méthode présente le meilleur compromis au regard des critères d'analyse présentés en section I.

Il convient toutefois de noter que cette méthode ne permet pas de mesurer le trafic correspondant à des contenus hébergés au sein du réseau du FAI. Pour prendre en compte ce trafic, les mesures au niveau de l'interconnexion devraient être complétées par des mesures effectuées, en utilisant les mêmes outils, au niveau des serveurs internes au réseau du FAI. Des travaux complémentaires seraient toutefois nécessaires pour vérifier la faisabilité de ces mesures et préciser les serveurs à prendre en compte.

Enfin, il est important noter que les deux méthodes présentées ci-dessus ne sont pas complémentaires. En effet, chacun de ces dispositifs opérant au niveau d'un point de mesure différent dans le réseau, leurs résultats respectifs seraient difficilement conciliables et risqueraient d'entraîner un comptage multiple des mêmes flux.

IV Limites et possibilités de contournement d'un dispositif de mesure

Plusieurs modes d'acheminement existent dont il convient de tenir compte dans le cadre de la mise en place d'un dispositif de mesure et de caractérisation du trafic sur internet. En effet, certains de ces modes – déjà largement intégrés aux stratégies des opérateurs et FCA pour l'acheminement de leur trafic – peuvent aboutir à des mesures partielles des flux concernés en fonction des points de mesure retenus.

Par ailleurs, certaines pratiques sur internet visant à optimiser l'usage de la bande passante ou à sécuriser les échanges peuvent réduire de manière significative la capacité d'un dispositif de mesure à identifier l'origine, la destination ou la nature des flux acheminés. Un acteur qui ne souhaiterait pas que son trafic soit mesuré ou identifié pourrait donc les employer comme solutions de contournement.

Dans le cadre de cet avis, l'Autorité souhaite souligner plusieurs pratiques pouvant restreindre la portée utile d'un dispositif de mesure :

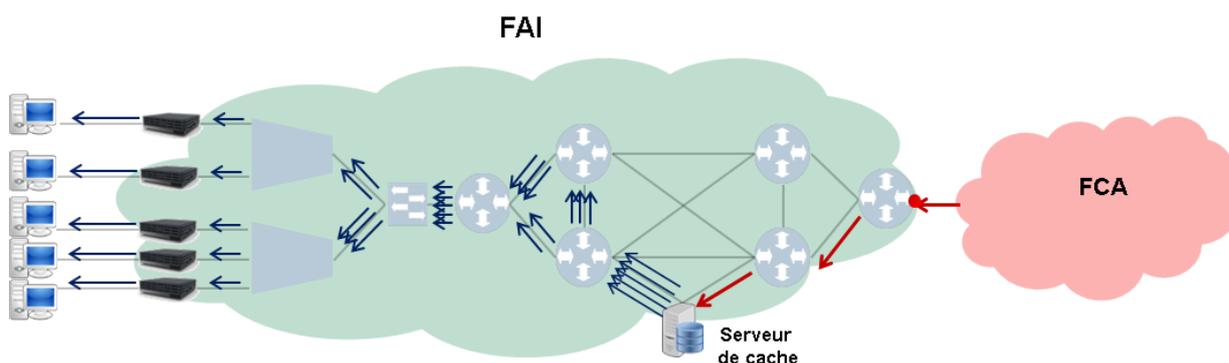
- l'hébergement des contenus au sein du réseau du FAI ;
- le développement des échanges de pair-à-pair (i.e. l'hébergement des contenus par les utilisateurs finals) ;
- la localisation des interconnexions en dehors du territoire national ;
- la diffusion en *multicast*.

Dans cette section sera donc étudié l'impact des divers modes d'acheminement et des possibles techniques de contournement sur la capacité d'un dispositif à obtenir des mesures exhaustives, fiables et à identifier l'origine et la nature des flux.

IV.1 L'hébergement des contenus au sein du réseau du FAI

Les serveurs internes tels que les serveurs de cache ou les CDN dits « *on-net* » sont des plateformes localisées au sein du réseau d'un FAI servant à stocker localement des copies de certains contenus disponibles sur internet. Ils visent à rapprocher les contenus des utilisateurs, dans l'optique d'optimiser les coûts et les performances : leur utilisation réduit donc le recours aux interconnexions. Dans ce mode d'acheminement, les contenus à plus forte audience ne sont acheminés qu'une seule fois à travers l'interconnexion pour alimenter¹⁵ ces serveurs qui se chargeront par la suite de desservir les utilisateurs autant de fois que les contenus seront sollicités.

Selon les informations recueillies par l'ARCEP dans le cadre de la préparation de son avis, on peut considérer qu'un ratio standard entre trafic entrant et sortant pour un serveur de cache ou un CDN est d'environ 1:5. Le recours à des serveurs de cache ou à des CDN internes permettrait donc de réduire de près de 80 % le recours aux interconnexions concernant les contenus pouvant y être stockés.



Fonctionnement des serveurs de cache interne

Un dispositif de mesure opérant au niveau des points d'interconnexion du FAI ne comptabiliserait les flux associés à chaque contenu qu'une seule fois, sans être en mesure de tenir compte de la multitude de flux individuels générés en aval dans le réseau. Il en va de même des contenus sollicités par les abonnés d'un FAI qui sont hébergés par le FAI lui-même, pour lesquels le trafic associé ne transitera pas non plus par l'interconnexion.

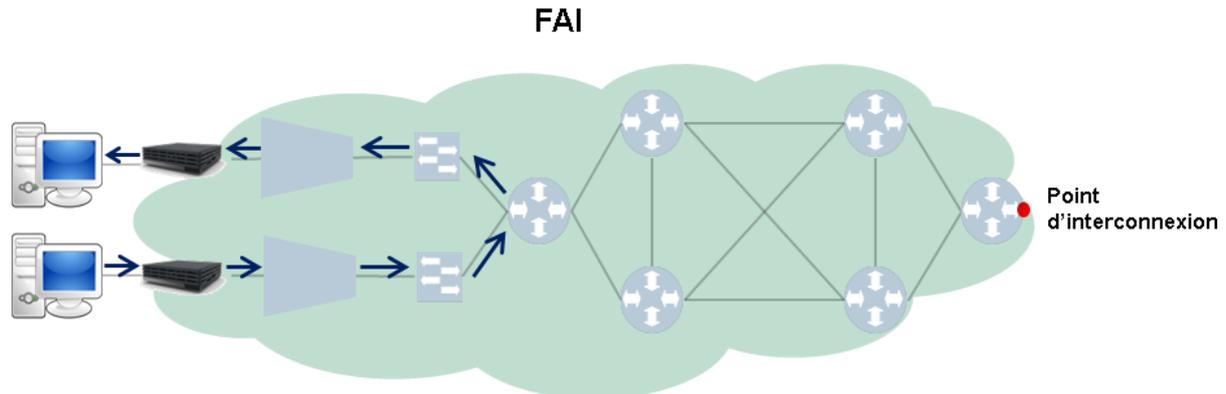
Dans le cas des contenus hébergés dans le réseau même du FAI, il est donc nécessaire de se placer en aval du point de diffusion afin de garantir la complétude des mesures.

IV.2 L'hébergement des contenus par les utilisateurs finals : cas du pair-à-pair

Le pair-à-pair (en anglais « *peer-to-peer* ») ou *P2P* est une technique permettant d'échanger des données entre internautes, par opposition au modèle client-serveur – désormais le plus répandu sur internet. Concrètement, chaque utilisateur devient un fournisseur des contenus

¹⁵ Opération de « *mise en cache* ».

qu'il met à disposition des autres utilisateurs du service. Ce mode de transmission est utilisé dans le cadre d'applications de partage de fichiers ou encore de calculs distribués et apporte une solution aux questions de disponibilité et de charge inhérentes aux modèles centralisés.



Acheminement du trafic pair-à-pair (entre abonnés d'un même FAI)

Les applications de type pair-à-pair, notamment celles permettant le partage de fichiers connaissent, depuis quelques années, un regain d'intérêt de la part des éditeurs de services. En effet, dans un contexte d'optimisation de l'usage de la bande passante et de réduction des coûts d'acheminement, plusieurs fournisseurs de services optent de manière croissante pour cette architecture.

S'agissant d'échanges entre utilisateurs finals, le trafic associé aux échanges entre les abonnés d'un même FAI reste interne au réseau de ce dernier. Un dispositif de mesure opérant au niveau de l'interconnexion n'aura donc aucune visibilité sur ce trafic.

De plus, dans le cas d'échanges de pair-à-pair, un dispositif visant à identifier les flux désignerait le FAI du client hébergeant le contenu mis à disposition comme étant à l'origine du trafic associé.

IV.3 La localisation des interconnexions en dehors du territoire national

Dans le cas de la mise en place d'un dispositif de mesure opérant au niveau des points d'interconnexion, il est important de veiller à ce que la base juridique sur laquelle reposerait cet outil permette effectivement de recueillir les informations issues des interconnexions situées en dehors du territoire national.

Par défaut, certains acteurs pourraient être incités à déplacer leurs interconnexions ou à héberger leurs contenus à l'étranger, ce qui aurait un impact défavorable sur les investissements et le déploiement des infrastructures en France. Une telle délocalisation des réseaux pourrait également dégrader la qualité de service ressentie par les utilisateurs localisés en France, ce point étant toutefois à apprécier au regard de la nature des services.

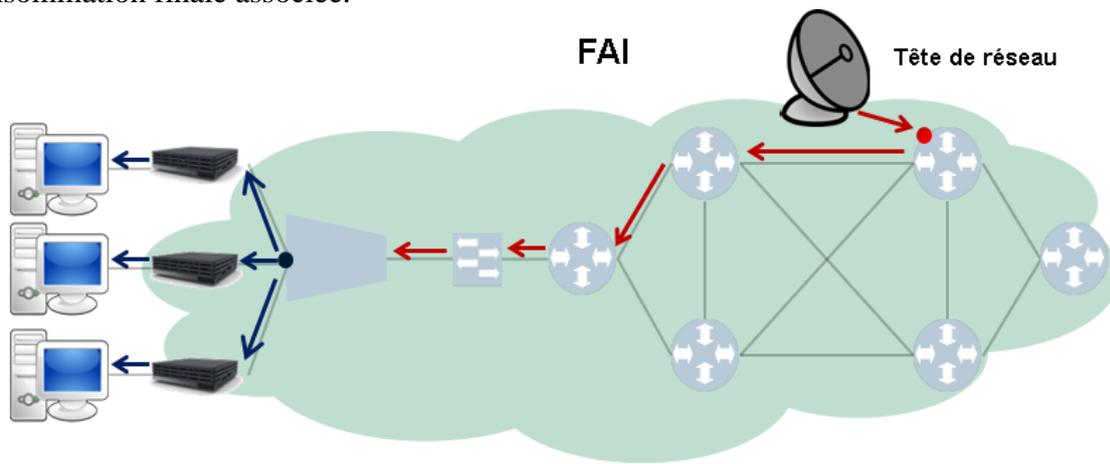
IV.4 La diffusion en multicast

La diffusion *multicast* est utilisée pour la diffusion d'une partie des services de télévision (linéaire) sur IP, dite « *TVIP* ». Elle permet à un opérateur, après récupération et transformation des signaux au niveau de la tête de réseau, de les acheminer de manière agrégée le plus longtemps possible, généralement jusqu'au dernier équipement d'accès avant

la ligne de l'abonné. La démultiplication en flux individuels est réalisée seulement à ce niveau du réseau, en fonction des sollicitations des abonnés.

Ce mode de diffusion vise à optimiser l'utilisation de la bande passante sur les réseaux en évitant de démultiplier inutilement des flux individuels identiques (visionnage et diffusion simultanés), comme c'est par exemple le cas pour les services de vidéo non linéaires (visionnage et diffusion individuels), dont la diffusion n'est réalisée en général qu'en *unicast* (un flux de bout en bout par abonné).

Dans le cadre de la diffusion en *multicast*, les flux étant émis à partir de la tête de réseau de l'opérateur, les points d'interconnexion en périphérie du réseau ne sont pas empruntés et un dispositif de mesure opérant à ce niveau ne sera donc pas à même de quantifier la consommation finale associée.



Diffusion multicast

De même, un dispositif qui opèrerait cette fois au niveau de la collecte ne mesurera que le flux agrégé sans être en mesure de tenir compte de la démultiplication des flux qui s'opère en aval dans le réseau¹⁶.

Or, à partir des données recueillies dans le cadre de son questionnaire, l'Autorité estime que les flux *multicast* agrégés acheminés depuis la tête de réseau jusqu'à l'accès, en passant par la collecte, ne représenteraient qu'un débit de 1 à 2Gbit/s contre quelques Tbit/s en consolidant tous les flux individuels consommés par les abonnés. Ce débit représenterait plus de la moitié de la bande passante disponible au niveau des accès des utilisateurs finals.

¹⁶ Hormis pour les réseaux câblés, dans lesquels l'injection des flux audiovisuels se fait en aval de la collecte et ne sont donc pas mesurés à ce niveau.

V Conclusion

Mesurer le trafic sur internet est une tâche rendue complexe par la diversité des modalités d'acheminement de ce trafic. Pour parvenir jusqu'à l'utilisateur final, les paquets IP peuvent en effet emprunter plusieurs routes différentes, plus ou moins directes, au cours d'un même échange. Ces routes traversent de multiples réseaux et peuvent reposer sur des serveurs cache (où une version du contenu sollicité est stockée) plus ou moins proches de l'utilisateur final.

Il n'existe par ailleurs pas de méthode univoque, infaillible ou exhaustive permettant d'associer un trafic à un service. La nature du trafic est en effet multiple – vidéo (linéaire ou à la demande), téléphonie, fichiers, sites internet, etc. – et internet est caractérisé par une décorrélation forte entre les contenus, services et applications, d'une part, et les réseaux qui en assurent la diffusion de manière majoritairement indifférenciée, d'autre part. L'identification du type de service transporté nécessite donc une analyse en profondeur des flux, qui peut non seulement soulever des questions en matière de protection des données personnelles, mais également être rendue impossible par le chiffrement des données émises.

Ces limites sont intrinsèques au fonctionnement d'internet, mais elles représentent aussi sa force puisqu'elles assurent à la fois sa résilience et son caractère ouvert.

C'est dans ce contexte que l'Autorité a évalué différentes options de mesures de la bande passante sur internet. Ces options ont été évaluées au regard des critères suivants : faisabilité technique, coût et complexité raisonnable de mise en œuvre, fiabilité, complétude, caractère certifiable des résultats obtenus et ce, dans le respect des principes de neutralité d'internet et du secret des correspondances.

En termes de niveau d'analyse, l'Autorité estime que la relève du volume de données écoulées sur un lien donné grâce au protocole SNMP permet d'obtenir des résultats fiables et certifiables, avec une relative simplicité de mise en œuvre, en comparaison des analyses de type Netflow/IPFix et *Deep packet inspection* (DPI).

En termes de point de mesure dans le réseau, l'Autorité considère qu'appliquer cette méthode (relève du volume grâce au protocole SNMP) au niveau des interconnexions serait raisonnable sur un plan technique et économique pour les fournisseurs d'accès à internet (FAI).

Ce type de mesure présenterait néanmoins deux limites :

- d'une part, s'agissant d'un simple comptage des paquets écoulés, le flux est identifié comme provenant du partenaire d'interconnexion. Il s'agit donc dans le cas général d'un intermédiaire technique (transitaire, *Content delivery network*, gestionnaire d'un point d'échange internet, etc.) et non d'un éditeur de service ;
- d'autre part, la mesure étant effectuée aux points d'interconnexion, elle ne permet pas de tenir compte des volumes de trafic acheminé dans des conditions particulières telles que la diffusion *multicast*, l'échange pair-à-pair entre deux abonnés d'un même FAI et la transmission depuis un centre d'hébergement en propre ou un serveur interne au réseau du FAI.

Il ressort de ce qui précède qu'il existe un risque que la mise en place éventuelle d'un dispositif de mesure incite certains acteurs à mettre en œuvre des techniques de contournement. Hormis pour le trafic pair-à-pair, qui ne saurait être associé à un service ou à

un éditeur particulier, une voie à explorer pour limiter ces contournements consisterait à compléter la mesure au niveau de l'interconnexion par des mesures effectuées par les FAI sur des liens spécifiques dans le réseau de collecte, notamment en ce qui concerne les serveurs de cache internes.

Fait à Paris, le 7 juillet 2015

Le Président

Sébastien SORIANO