YAHOO ! Inc.

Rapport d'expertise

ÉTUDE TECHNIQUE SUR LES POSSIBILITES DE FILTRAGE EN FONCTION DE LA PROVENANCE GEOGRAPHIQUE D'INTERNAUTES

Juillet 2000

Version 1.5

TABLE DES MATIERES

1. OBJET DU DOCUMENT

L'objectif de ce document est de présenter les mesures techniques de nature à identifier la provenance géographique des internautes afin de permettre le contrôle des accès aux informations consultables sur Yahoo.com et Yahoo.fr.

Le présent document décrit :

1. les composants et les techniques intervenant dans la communication entre un « Internaute » et le « site web » auquel il accède,
2. un inventaire des solutions techniques permettant :
   - l'identification de la nationalité de l'internaute
   - l'identification des informations illicites au travers des services rendus sur le Web,
   - de mettre en œuvre un contrôle d'accès aux informations illicites en fonction de la nationalité de l'internaute.

   Cet inventaire précise le niveau de confiance et les limites techniques et fonctionnelles de chacun des moyens techniques évoqués.
3. l'impact de mise en œuvre de ces solutions dans le cas des sites de Yahoo ! Inc. au travers du monde.

2. CONTEXTE

Monsieur le Président du Tribunal de grande instance de Paris a été saisi, en la forme des référés, par la Ligue Internationale Contre le Racisme et l'Antisémitisme et l'Union des Etudiants Juifs de France.

Ces deux associations demandaient à Monsieur le Président d'ordonner, sous astreinte, à la société Yahoo ! Inc. de détruire toutes les données et de cesser d'héberger certains de ses services et d'empêcher l'exhibition vente, d'objets nazis, sur son service de vente aux enchères.

Monsieur le Président du Tribunal de grande instance de Paris a ainsi rendu une ordonnance de référé en date du 22 mai 2000, aux termes de laquelle il a ordonné à la société Yahoo ! Inc. de prendre des mesures permettant de :

"dissuader et de rendre impossible toute consultation sur Yahoo.com du service de ventes aux enchères d'objets nazis et de tout autre site ou service qui constituent une apologie du nazisme ou une contestation des crimes nazis".

Par ailleurs, Monsieur le Président a ordonné la poursuite de l'instance à l'audience du 24 juillet 2000, au cours de laquelle la société Yahoo ! Inc. devra soumettre les mesures qu'elle compte prendre pour de conformer à cette ordonnance.

Yahoo ! Inc., dans cet optique, a décidé de s'adjoindre l'assistance d'un expert en Internet pour analyser l'ensemble des moyens techniques

envisageables permettant de répondre aux exigences de cette ordonnance.

3. CONTEXTE TECHNIQUE

« LE RESEAU DES RESEAUX PLANETAIRES »

Internet est devenu, notamment depuis l'apparition du World Wide Web en 1991, un phénomène planétaire.

En raison de sa nature décentralisée et de l'absence de fédération, il est difficile et souvent discutable d'établir des recensements et chiffrages portant sur Internet. Cependant, plusieurs études convergeaient, en Juin 2000, sur des chiffres fondamentaux :

- plus de 40 000 réseaux interconnectés (d'où son nom de «réseau des réseaux »)
- 364 millions de PC fin 1998, dont 112 millions en Europe (mais tous ne sont pas connectés),
- plus de 332 millions d'utilisateurs Internet dans le monde (plusieurs sources/méthodologies compilées par Nua Internet Surveys).
- dont plus de 7,7 millions d'internautes de 15 ans et plus en France au 1er trimestre 2000 sur le territoire français avec un taux de croissance annuelle de 46% (baromètre Internet de Médiamétrie),
- 236 pays connectés (TerraBay.com - International Internet Statistics),
- 10 millions de sites Internet de par le monde. Ce nombre a doublé en 1 an. (Benchmark Group).

« INTERNET UN ESPACE DE LIBERTES »

Cette affirmation était un principe, appliqué volontairement par les groupes d'utilisateurs qui ont influencé le développement de l'Internet. Aujourd'hui, il est techniquement impossible qu'il en soit autrement.

En effet, outre le fait que les bases techniques sur lesquelles repose l'Internet ont été volontairement définies pour offrir un espace de libertés sans contrôle central, la banalisation des moyens pour diffuser de l'information, leur faible coût et le volume d'informations véhiculées rendent illusoire toute opération de surveillance et de contrôle.

Ainsi, il est très difficile à un fournisseur de plate-forme d'hébergement de contrôler de manière efficace et sûr le contenu a priori des informations qu'il rend disponibles sur son site web.

« INTERNET N'A PAS DE FRONTIERE »

L'ouverture d'un site web sur Internet le rend automatiquement accessible depuis le monde entier.

Il est difficile pour un internaute de savoir si le site qu'il consulte est situé en France ou aux Etats-Unis. Tout au plus, l'internaute peut présumer qu'il a quitté l'espace français si le site consulté n'est pas en langue française, et si l'adresse du site consulté n'est pas dans le domaine de référencement «.fr». Des cas complexes peuvent apparaître : le nom du site est référencé dans un pays, les machines sont installées dans un second et le site est opéré par une société en provenance d'un troisième.

Pour qu'une communication puisse être établie entre un internaute et le site web qu'il souhaite consulter, il est nécessaire que les deux parties s'identifient mutuellement sur la *toile*. L'identification de l'internaute transmise au site web notamment au moment de la connexion contient une information qui, indirectement, est souvent liée à une localisation géographique. Néanmoins, dans la majorité des cas, c'est le fournisseur de l'accès Internet de l'utilisateur qui lui attribue cette identification sans garantir que la provenance géographique de l'internaute soit respectée.

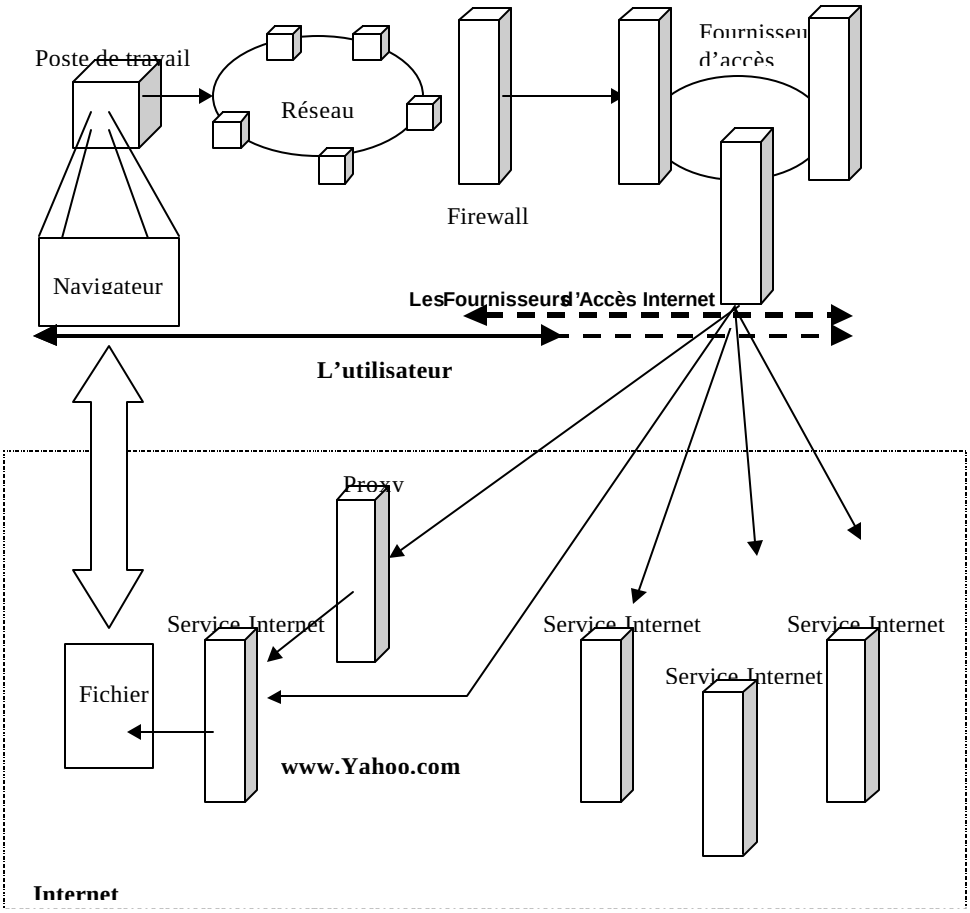
« L'USURPATION D'IDENTIFICATION »

Contrairement à d'autres protocoles de communication (par exemple : téléphone fixe ou mobile), le protocole IP ne fût pas formellement conçu pour identifier la provenance des données transportées.

Cette absence de contrôle facilite la masquarade, ce qui rend notamment l'identification de la source des données peu fiable.

4. LES COMMUNICATIONS ENTRE UN INTERNAUTE ET UN SITE WEB

4.1 LES COMPOSANTS D'INTERNET



Le **poste de travail** est le micro-ordinateur sur lequel l'internaute accède aux services offerts par les sites Internet grâce notamment à un logiciel appelé « **Navigateur** ». Plusieurs « navigateurs » existent. La majorité sont des logiciels « gratuits ». Les deux principaux sont « Netscape Navigator » et « Internet Explorer » et équipent presque la totalité des Internautes.

La machine et ses logiciels appartiennent à l'utilisateur ou lui ont été confiés par l'entreprise qui l'emploie.

L'ordinateur est connecté directement ou indirectement à un **Fournisseur d'Accès Internet**, société disposant d'équipements physiquement reliés au réseau Internet.

Dans une entreprise, le micro-ordinateur est en général connecté au réseau local qui dispose d'une connexion avec le fournisseur d'accès partagée par les utilisateurs de la société. Pour des raisons de sécurité, cette connexion est protégée par un **Firewall** dont l'une des fonctionnalités est de masquer vis à vis de l'extérieur l'identification des postes de travail (l'adresse IP des micro-ordinateurs).

Le fournisseur d'accès Internet dispose de son propre réseau local dont les équipements (les serveurs) peuvent être répartis dans le monde entier. Ainsi, certains d'entres eux ne disposent que de connexions physiques avec le réseau Internet dans leur pays d'origine et offrent leur service dans le monde entier. Ainsi, un français peut utiliser un fournisseur d'accès Internet qui fera passer toutes ses communications sur le réseau Internet via une connexion aux Etats Unis et qui pourra lui attribuer une adresse IP américaine. Cet utilisateur sera vu par les sites Internet comme provenant des États-Unis.

Sur l'Internet **les sites comme Yahoo.com et Yahoo.fr** peuvent être contacter directement en saisissant une chaîne de caractères que l'on appelle «l'URL», par exemple www.Yahoo.com. Il est également possible d'y accéder «indirectement». En effet, il existe des services appelés «**relais anonymes**» qui permettent de réaliser une passerelle entre le site à contacter et l'internaute. Ainsi, l'internaute se connecte au «relais anonyme», puis lui demande de se connecter au site www.Yahoo.com. Ces passerelles servent notamment à contacter un site web de manière totalement anonyme : toutes les informations pouvant être utilisées pour identifier l'internaute sont remplacées par ce relais. En particulier, le site web ne reçoit jamais les données d'identification de l'internaute tel que l'adresse IP.

4.2 L'ETABLISSEMENT D'UNE COMMUNICATION

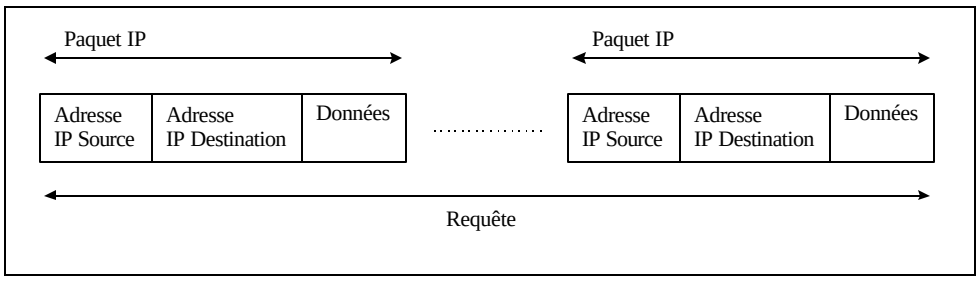
LES PAQUETS TCP/IP

Une communication Internet est constituée de données qui sont échangées entre l'appelant (l'internaute) et l'appelé (le site web).

Ces données sont véhiculées sur un support physique (réseau Ethernet) et doivent respecter une structure standard (protocole TCP/IP). Ce format standard impose notamment que les données soient transmises sous la forme de paquets (les paquets IP). Ces paquets sont générés par l'émetteur, dirigés et redirigés par les équipements réseau et transmis au destinataire.

Chaque paquet pouvant être véhiculé indépendamment les uns des autres, celui-ci contient l'identification du destinataire (l'adresse IP de destination) et l'identification de l'émetteur (l'adresse IP source).

Ces deux identifications dans chaque paquet permet au destinataire de trier les paquets qu'il reçoit et de retransmettre sa réponse au bon émetteur.



ÉTABLISSEMENT DE LA CONNEXION ENTRE UN INTERNAUTE ET UN SITE WEB

Lorsqu'un utilisateur saisi une adresse URL pour tenter d'accéder à une page Web, les étapes suivantes sont réalisées :

1 – Le micro-ordinateur utilise une connexion Internet déjà établie. S'il n'existe pas de connexion, il tente d'accéder à son fournisseur d'accès Internet via un modem ou une liaison spécialisée. Une fois la phase de négociation terminée entre le client et le fournisseur d'accès (demande d'identifiant, de mot de passe), le serveur attribue à l'internaute une adresse IP. Cette adresse IP est une adresse choisie par le fournisseur d'accès parmi les adresses IP qu'il a réservé sur l'Internet.

2 – Une fois l'adresse IP du client attribuée, il peut communiquer avec le monde Internet. Le navigateur installé sur le client, envoie donc sa requête (demande de connexion à l'URL saisie) via les serveurs du fournisseur d'accès Internet, à un serveur chargé de diriger sa demande vers le serveur contenant la page Web qu'il désire consulter. La requête peut ainsi rencontrer plusieurs serveurs qui vont la diriger et la rediriger jusqu'à arriver à destination.

3 – Une fois la destination atteinte, le serveur cible examine la requête et extrait des paquets IP la composant les informations concernant l'internaute, notamment l'adresse IP de l'appelant.

4 – Le serveur exécute la demande en recherchant dans ses fichiers le fichier demandé.

4bis – Si le site est conçu pour effectuer un contrôle des accès aux fichiers servis, alors il demande au client de s'authentifier. Le serveur vérifie alors que l'identité fournie dispose de droits suffisant pour accéder au fichier demandé.

5 – La page demandée est reçue par le demandeur et traitée par le navigateur pour en afficher le contenu.

4.3 L'ANALOGIE AVEC UNE COMMUNICATION TELEPHONIQUE

	Téléphone	**Réseau**
Composants	- L'appelant - L'appareil téléphonique - Numéro de l'appelant - Numéro composé par l'appelant - Indicatif régional et national - Appelle une opératrice - Demande de mise en relation avec un abonné TOTO - PABX d'entreprise - Centre d'exploitation France Télécom	- L'utilisateur - Le poste de travail - Adresse IP Source (modifiable par l'appelant plus ou moins difficilement) - Adresse IP Destination - Nom de domaine (Classes d'adresses TCP/IP) - Connexion à un DNS - Traduction du nom URL en adresse IP - Firewalll - Fournisseur d'accès Internet
Support	Lignes téléphoniques : normes de connexion électrique.	Réseau : standard Ethernet, …
Protocoles	- Procédures de connexion	- Couches basses du protocole

	techniques entre deux téléphones : générer une sonnerie, maintient de la communication, ... - Convention humaine : « Allo ? », « Qui est à l'appareil ? »,,...	TCP/IP - Couche applicative du protocole TCP/IP : HTTP, SMTP, FTP, ...

5. INVENTAIRE DES MESURES TECHNIQUES

Les « mesures techniques de nature à dissuader et rendre impossible toute consultation » de services litigieux en fonction de la nationalité de l'internaute conduisent à réaliser les trois objectifs suivants :

1. **Identification de l'appelant**

   Cette identification de l'appelant consiste à reconnaître la nationalité du demandeur avant toute consultation d'informations sur le site web.

2. **Identification des informations diffusées**

   Ce contrôle des informations consiste à qualifier les informations accessibles sur le site web en fonction de leur contenu.

3. **Filtrage des informations**

   Il s'agit de définir les actions à mener lorsqu'un internaute tente d'accéder à un service contenant des informations illicites vis à vis de la législation de son pays d'origine.

5.1 IDENTIFICATION DE L'APPELANT

Pour identifier la nationalité de l'internaute dans la chaîne de communication qui lui est transmise par le navigateur, cinq solutions sont envisageables :

1. A partir de l'adresse IP contenue dans les paquets émis par l'internaute et reçus par le site web.
2. A partir du nom de domaine associé à l'adresse IP
3. A partir de la version linguistique du navigateur utilisée sur le poste de l'utilisateur
4. A partir d'une déclaration de l'utilisateur.

5.1.1 A partir de l'adresse IP contenu dans les paquets reçus

Définition

Un mécanisme logiciel est activé à la réception d'un paquet afin d'en extraire l'adresse IP source.

Cette adresse est ensuite comparée à celles dont la provenance géographique est déjà identifiée. Si cette adresse ne fait pas partie de celles déjà identifiées, alors ce mécanisme logiciel consulte une base d'informations lui permettant de retrouver le pays correspondant à la plage d'adresses IP à laquelle appartient l'adresse obtenue.

De telles bases existent et peuvent être consultées sur l'Internet. Elles permettent en fournissant une adresse IP d'obtenir le pays dans laquelle elle a été déclarée.

Limites fonctionnelles

La discontinuité des adresses IP et leur modification fréquente engendrera une charge d'exploitation importante supplémentaire pour l'opérateur du site, notamment pour la mise à jour régulière de la table de correspondance entre les adresses IP et leur pays d'appartenance. Pour éviter cette surcharge l'opérateur du site peut utiliser des tables maintenues par les NIC (Network Information Center) qui donnent le pays d'origine à partir d'une adresse IP, ou des bases d'informations maintenues par des sociétés tierces.

Le principal problème réside dans la baisse importante des performances des serveurs, et par conséquent de la qualité de service pour tous les utilisateurs du monde. En effet, l'extraction de l'adresse IP dans un paquet est une opération simple, mais les traitements nécessaires à l'identification de la provenance géographique de l'appelant puis l'utilisation de cette information pour déterminer les règles d'accès à un service peuvent s'avérer très coûteux en temps.

Limites techniques

La qualité d'identification par ce moyen de filtrage bien qu'étant le meilleur n'est pas sûr au sens où :

- des utilisateurs français pourront être pris comme des internautes non français à leur insu : des fournisseurs d'accès Internet accessibles en France attribuent à leurs utilisateurs des adresses IP déclarées comme appartenant au pays où est implanté le siège de la société.
- des utilisateurs pourront volontairement usurper une nouvelle nationalité :
  - → Ils peuvent utiliser par le biais de manipulation simple des services de « relais anonymes » (aussi appelés « proxy »), accessibles à tout public, qui permettent de naviguer sur Internet en masquant l'adresse IP source des paquets transmis par l'appelant (l'adresse IP source réelle est remplacée par l'adresse IP du relais anonyme).
  - → Ils peuvent également utiliser des logiciels qui modifient leurs adresses IP, attribués par le fournisseur d'accès, et usurpent une nouvelle adresse non utilisée.

Conclusion

Cette solution est la plus précise disponible actuellement.

Cependant, elle n'est pas suffisamment fiable pour servir de base à un dispositif de contrôle de la provenance géographique d'un internaute dans le but de bloquer l'accès à des contenus litigieux vis-à-vis de la législation qui est applicable à l'internaute.

Il est à noter qu'une grande part des utilisateurs individuels disposant d'un accès Internet passe par des fournisseurs d'accès qui ne garantissent pas que l'adresse IP qu'ils attribuent à leur client est celle correspondant à leur pays d'origine.

5.1.2 A partir du nom de domaine

Définition

La solution consiste ici à transformer l'adresse IP en Nom de domaine. En effet, les sites web possèdent une fonction qui automatiquement retrouve le nom du domaine à partir de l'adresse IP source qu'ils reçoivent. Cette fonction s'apparente à l'utilisation d'un annuaire inversé : mise en correspondance du numéro IP avec le nom de domaine.

Le nom de domaine est constitué d'un mot clé (le 1er en partant de la droite) qui dans certains cas permet d'identifier le pays d'origine. Par exemple : fr pour la France, ch: pour la suisse, …

Limites fonctionnelles

L'analyse du nom de domaine ne permet pas toujours d'identifier le pays d'origine de l'utilisateur. Par exemple, le mot clé .COM très largement répandu identifie une société commerciale et non pas sa nationalité. Ainsi, un internaute utilisant le poste de travail mis à sa disposition par son employeur, et l'accès Internet de ce dernier, pourrait être identifié comme appartenant au domaine trans-nationale qu'est le « .com ». Il en existe ainsi plusieurs : « .org », « .net », etc.

Comme pour l'utilisation de l'adresse IP, le principal problème réside dans la baisse de performance des serveurs, et par conséquent de la qualité de service pour tous les utilisateurs du monde. En effet, l'extraction de l'adresse IP dans un paquet est une opération simple, mais les traitements nécessaires à l'obtention du nom de domaine associé puis l'utilisation de cette information pour déterminer les règles d'accès à un service web peuvent s'avérer très coûteux en temps.

Limites techniques

On retrouve les mêmes limitations que dans le cadre de l'utilisation des adresses IP sources :

- des utilisateurs français pourront être considéré comme des internautes non français à leur insu : des fournisseurs d'accès Internet accessibles en France associent à leurs adresses IP un nom de domaine dissocié du pays de provenance de leurs utilisateurs.
- des utilisateurs pourront volontairement usurper une nouvelle nationalité :
  - → Ils peuvent utiliser, par le biais de manipulations simples, des services de « relais anonymes » (aussi appelés « proxy »), accessibles à tout public, qui permettent de naviguer sur Internet en masquant l'adresse IP source des paquets transmis par l'appelant (l'adresse IP source réelle est remplacée par l'adresse IP du relais anonyme. C'est par conséquent le nom de domaine associé au relais anonyme qui sera obtenu après identification).
- des fournisseurs d'accès ne respectent pas l'obligation technique de fournir un annuaire inversé permettant d'associer un nom de domaine à une adresse IP.

Conclusion

Ce moyen d'identification présente des limites techniques importantes, mais il est simple d'utilisation.

Néanmoins, l'exploitation de cette information d'identification reste complexe à mettre en œuvre dans l'objectif de rendre impossible toute consultation d'informations illicites vis-à-vis de la législation du pays d'origine de l'internaute, et engendre une baisse notable des performances.

5.1.3 A partir de la version linguistique du navigateur utilisée sur le poste de l'utilisateur

Définition

L'objectif de cette solution est d'identifier le pays d'origine de l'internaute en analysant la version du navigateur qu'il utilise.

Cette information est automatiquement recueillie par les sites web au moment de la connexion de l'internaute. La version du navigateur permet dans la plupart des cas d'identifier la langue utilisée.

Limites fonctionnelles

L'analyse de la version du navigateur ne permet pas d'identifier le pays d'origine de l'utilisateur, mais uniquement la langue qu'il utilise.

Comme pour l'utilisation de l'adresse IP, le principal problème réside dans la baisse de performance des serveurs, et par conséquent de la qualité de service pour tous les utilisateurs du monde. En effet, l'extraction de la langue dans la version d'un navigateur est une opération simple, mais l'utilisation de cette information pour déterminer les règles d'accès à un service web peuvent s'avérer très coûteux en temps.

Limites techniques

Rien n'empêche un internaute français d'utiliser une version américaine de son navigateur, ou de modifier la langue utilisée par défaut par son navigateur.

Conclusion

Cette solution est techniquement la moins fiable.

5.1.4 A partir d'une déclaration de l'utilisateur

Définition

Cette solution consisterait à demander à l'utilisateur, au moment de sa première connexion, de remplir un questionnaire pour déclarer « sur l'honneur » le pays d'origine de sa connexion Internet.

Le serveur web renverrait sur le poste de l'utilisateur un élément d'identification de sa nationalité (par exemple, sous la forme d'un « Cookie »), élément que le site utiliserait par la suite pour identifier l'utilisateur à chacune de ces connexions.

Limites fonctionnelles

Comme pour l'utilisation de l'adresse IP, le principal problème réside dans la baisse de performance des serveurs, et par conséquent de la qualité de service pour tous les utilisateurs du monde. En effet, la récupération de l'information saisie par l'utilisateur est une opération simple, mais l'utilisation de cette information pour déterminer les règles d'accès à un service web peut s'avérer très coûteuse en temps.

De plus, contrairement aux moyens techniques précédemment décrits, cette obligation de déclaration n'est pas transparente pour l'utilisateur. Et rien ne pourrait interdire à l'utilisateur de mentir.

Limites techniques

Cette solution se base sur une déclaration de la part de l'utilisateur, elle n'est donc pas soumise à des faiblesses techniques.

Conclusion

Cette solution serait la plus efficace si les utilisateurs ne pouvaient pas mentir.

De plus, cela engendre pour les utilisateurs du monde entier la nécessité de remplir un formulaire avant l'utilisation du service.

5.1.5 Synthèse

Quelle que soit la méthode utilisée, il subsiste toujours un doute.

Actuellement, la seule et unique source fiable de l'identification de la provenance géographique de l'internaute est le fournisseur d'accès Internet. Cependant, cette information n'est pas techniquement disponible pour le site consulté.

5.2 IDENTIFICATION DES INFORMATIONS DIFFUSEES

5.2.1 Définition

L'objectif est ici d'identifier parmi les informations publiées sur un site Internet, celles qui sont interdites en France. Ces informations peuvent être du texte, des extraits sonores ou des images.

En fonction du volume d'informations et de la fréquence de mise à jour de ces informations, il est possible d'envisager d'appliquer des procédures manuelles visant à qualifier l'information avant sa mise à disposition sur l'Internet.

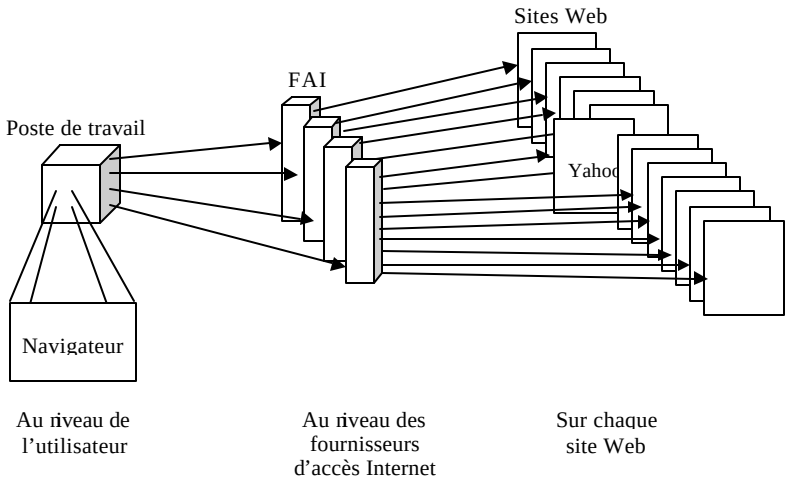
Outre, cette solution « manuelle », l'utilisation de logiciels automatiques de filtrage d'informations illicites est envisageable.

Ce filtre peut être appliqué selon trois méthodes :

- Par mots clés : le logiciel recherche dans la page HTML la présence de « mots clés » définis dans une liste préalablement renseignée par l'administrateur du logiciel.

- Par «URL» : le logiciel réalise un contrôle d'accès à des adresses «URL» préalablement définies. Des services de mises à jour de ces adresses sont proposés par certains éditeurs.
- Des organismes indépendants (PICS, cf. annexe 1) tentent de mettre en place un moyen de reconnaissance des sites web présentant un caractère immoral. Ils demandent aux éditeurs de pages web de volontairement ajouter aux pages concernées une identification de leur caractère immoral. Au niveau utilisateur, les navigateurs comme Internet Explorer disposent de règles (désactivées par défaut) permettant de filtrer les pages consultées grâce à cette identification.

Ce type d'identification des données peut être appliqué à trois niveaux :



- **Soit au niveau du moyen de consultation** (sur le poste de travail de l'utilisateur). Dans ce cas, les outils de filtre par mot clé ou par suivi d'une liste d'adresses URL peuvent être envisagés (cf. étude présentée en annexe).
- **Soit au niveau du moyen d'accès à l'Internet** (chez le fournisseur d'accès Internet). Comme pour le poste de travail, tous les outils de filtrage peuvent être utilisés. Ainsi, par exemple, EduNet propose un accès Internet filtré grâce à un logiciel qui garantit l'accès à un contenu sûr et approprié d'Internet.
- **Soit à la source** (sur tous les sites web). Dans ce cas, les outils de recherche par mot clé sont utilisés.

5.2.2 Synthèse

Quelle que soit la méthode d'identification utilisée, aucune n'est totalement fiable. La source la plus fiable d'identification de contenu illicite pour l'internaute reste l'internaute lui-même.

5.3 FILTRAGE DES INFORMATIONS

5.3.1 Actions

Lorsque l'internaute a été identifié comme français et que les informations auxquelles il tente d'accéder ont été reconnues comme illicites, deux types de réaction sont possibles :

1. **Affichage au moment de la connexion au site** (avant la consultation d'information) **d'un message d'avertissement** signalant à l'internaute que les informations diffusées dans ce site peuvent contenir des informations illicites.

   Cette solution permettrait aux utilisateurs, selon leur nationalité, de définir leur comportement : soit ils quittent le site, soit ils accèdent aux informations.
2. **Moyens visant à interdire le téléchargement** (et donc la consultation) de pages contenant des informations illicites.

Ces deux types de moyens peuvent être mis en œuvre à plusieurs niveaux :

1. Soit par le site web. Cette solution obligerait l'opérateur du site web à connaître l'ensemble des législations s'appliquant au contenu servi, en fonction de la provenance géographique de l'utilisateur.
2. Soit par le fournisseur d'accès Internet. Cette solution obligerait les fournisseurs d'accès français à trouver un consensus commun et à développer ces moyens de filtrage. Cette solution ne peut donc être envisageable dès maintenant.

   Il est à noter que cette solution est déjà utilisée dans plusieurs pays, notamment en Chine.
3. Soit par le micro-ordinateur de l'utilisateur. Cette solution est celle qui pourra être appliquée dans les meilleurs délais. Néanmoins, elle ne rendra pas impossible l'accès à des informations illicites si l'internaute le souhaite.

5.3.2 Synthèse

Parmi les différents niveaux auxquels il est possible d'agir, le plus pertinent, de part sa position dans le cheminement de l'information, et de part sa connaissance de la législation s'appliquant à l'internaute, reste le fournisseur d'accès Internet.

6. SOLUTIONS ENVISAGEABLES

Trois entités interviennent lors de la consultation d'informations diffusées sur un site Internet :

- Le micro-ordinateur de l'utilisateur,
- Le fournisseur d'accès Internet,
- Le site web « publiant » les informations.

6.1 IMPACTS POUR YAHOO ! INC.

6.1.1 Identification de l'appelant

6.1.1.1 A partir de l'adresse IP contenu dans les paquets reçus

Cette solution nécessiterait pour Yahoo ! Inc. une modification du fonctionnement de ses sites :

soit par un développement logiciel complémentaire de l'application qui est en charge de la gestion de la connexion de tous les internautes tentant d'accéder au site web, et une augmentation des ressources matérielles associées à cette application

soit par l'ajout d'un système matériel et logiciel en interposition entre l'Internet et les sites de Yahoo ! Inc.

La mise en place de cette solution demanderait donc une refonte du système dinformation des différents sites de Yahoo ! Inc. afin de rendre impossible toute consultation de données illicites vis-à-vis de la législation du pays d'origine de l'internaute.

De plus, cela entraînerait une importante dégradation des performances du site pour l'ensemble des utilisateurs du monde.

6.1.1.2 A partir du nom de domaine

Cette solution nécessiterait pour Yahoo ! Inc. de ré-introduire dans le logiciel en charge de son site web une fonction qui fût supprimée pour des raisons de performances.

La ré-introduction de cette fonction permettrait de disposer facilement de l'information de provenance géographique, mais l'utilisation de cette information nécessiterait pour Yahoo ! Inc. une augmentation notable des ressources matérielles associées à ses sites web.

6.1.1.3 A partir de la version linguistique du navigateur utilisée sur le poste de l'utilisateur

L'utilisation de cette information permettrait de disposer facilement d'une estimation de la nationalité d'un internaute, mais l'utilisation de cette information nécessiterait pour Yahoo ! Inc. de modifier le logiciel de gestion de ses sites, et une augmentation notable des ressources matérielles associées.

6.1.1.4 A partir d'une déclaration de l'utilisateur

L'utilisation de cette information permettrait de disposer facilement d'une estimation de la nationalité d'un internaute, mais l'utilisation de cette information nécessiterait pour Yahoo ! Inc. de modifier le logiciel de gestion de ses sites, et une augmentation notable des ressources matérielles associées.

6.1.2 Identification des informations diffusées

Compte tenu du volume d'informations et la fréquence très élevée de mise à jour de ces informations, il est impossible pour Yahoo ! Inc. d'envisager d'appliquer des procédures manuelles visant à qualifier l'information avant sa mise à disposition sur l'Internet au travers de ses différents sites dans le monde.

L'installation au niveau des sites de Yahoo ! Inc. d'un moyen automatique de filtrage des informations diffusées présente un coût important pour Yahoo! Inc. en terme de ressources humaines et matérielles, et en terme de perte de performance lors du référencement de ces informations.

6.1.3 Filtrage des informations

Le filtrage des informations au niveau du serveur web ne serait envisageable par Yahoo! Inc. que s'il était possible de s'assurer que l'interdiction ne s'applique qu'à des internautes français. Dans le cas contraire, Yahoo ! Inc. priverait les autres internautes du monde des informations publiées sur ses sites, ce qui n'est pas envisageable.

De plus, la mise en place d'un tel système demanderait une refonte du système d'information des différents sites web de Yahoo ! Inc. afin de rendre impossible toute consultation d'informations illicites vis-à-vis de la législation du pays d'origine de l'internaute, et engendrerait une importante dégradation des performances du site pour l'ensemble des utilisateurs

6.2 SOLUTIONS PROPOSEES

6.2.1 A court terme

Pour mettre en place à court terme un contrôle d'accès aux informations dites illicites en France, Yahoo ! Inc. propose de mettre en place **au niveau du micro-ordinateur de l'internaute un logiciel de filtrage**.

Les avantages de cette solution sont les suivants :

- Le contrôle du flux d'information sera réalisé sur le contenu des pages téléchargées depuis le micro-ordinateur. C'est à dire qu'il s'appliquera aux sites Yahoo.com et Yahoo.fr mais aussi à tout autre site.

- La mise en œuvre de tels logiciels est simple et permet de mettre en place une solution efficace dans des délais très courts. De plus, certains d'entre eux sont gratuits (Cf. annexe 2)
- Cette solution est compatible avec la philosophie d'Internet qui est de laisser la responsabilité à l'utilisateur de consulter les informations qu'il souhaite.
- L'identification de l'internaute par les sites de Yahoo ! Inc. n'est pas nécessaire. La fiabilité de la solution n'est donc pas dépendante de la qualité de l'identification.

Cette solution présente de très nombreux avantages (cf. ci-dessus). Néanmoins, elle nécessite une action volontaire de la part de l'utilisateur.

6.2.2 A long terme

Pour mettre en place à plus long terme un contrôle d'accès aux informations dites illicites en France, Yahoo ! Inc. propose de mettre en place **au niveau du fournisseur d'accès de l'internaute un système de filtrage**.

Les avantages de cette solution sont lessuivants :

- Le contrôle du flux d'information sera réalisé sur le contenu des pages téléchargées depuis le micro-ordinateur. C'est à dire qu'il s'appliquera aux sites Yahoo.com et Yahoo.fr, mais aussi à tout autre site.
- Une action volontaire ne sera pas nécessaire de la part de l'utilisateur, celui-ci ne pourra donc pas accéder aux informations identifiées comme illicites.

Cette solution présente de très nombreux avantages (cf. ci-dessus). Néanmoins, elle nécessite la mise en œuvre de systèmes complexes et coûteux au niveau des fournisseurs d'accès Internet. Et en particulier, d'une procédure d'identification des sites litigieux.

7. CONCLUSION

Il n'existe pas, dans l'état actuel des techniques présentées, de mesures, pouvant être mise en œuvre sur le site web, permettant de *"dissuader et rendre impossible toute consultation"* de certains services internet, sans détruire la qualité de fonctionnement des services proposés.

De plus, aucune des solutions techniques envisageables n'est incontournables, et certaines peuvent aboutir, soit à bloquer des internautes non français, soit permettre l'accès à des sites litigieux à des internautes français, soit encore à bloquer l'accès à des sites non litigieux.

En outre, la mise en œuvre de ces techniques, bien que non satisfaisantes au regard de l'objectif poursuivi, entraînerait un coût disproportionné et nécessiterait un temps d'étude, de validation puis de déploiement de plusieurs mois, qu'il est impossible d'évaluer avec précision.

La solution la plus fiable serait d'appliquer des moyens de filtrage au point d'initiation de la connexion, qui se trouve être à l'endroit même où la législation est applicable, c'est à dire au niveau de l'outil de consultation (sur le poste de l'internaute), ou au niveau du fournisseur d'accès Internet.

8. ANNEXE 1 : ORGANISMES DE CONTROLE DU CONTENU

8.1 PICS

« PICS (“Plateform for Internet Content Selection”) [www.w3.org/PICS]. Il s’agit d’un groupe de travail créé à l’initiative du World Wide Web Consortium (W3C) pour mettre au point un système de contrôle volontaire du contenu de médias interactifs tel Internet.
Le groupe de travail a établi un système d’étiquetage qui peut être appliqué sur trois niveaux :
- Self-rating” : permet aux fournisseurs de contenu d'étiqueter les contenus qu'ils proposent.
- Third-party rating” : permet à des services indépendants d’effectuer un étiquetage additionnel sur des contenus distribués par divers services.
- Ease-of-use” : permet aux parents et professeurs d'étiqueter des sites accessibles aux enfants et élèves.
La norme PICS a l’appui des producteurs de contenus aux États-Unis. Netscape et Microsoft ont accepté de l’inclure dans leurs navigateurs. De nombreuses organisations, en général dans le domaine scientifique, ont commencé de créer leurs propres systèmes de fiches et de labels. C'est le cas par exemple dans la santé avec l'initiative Health on the Net (http://www.hon.ch), reprise en France (Centrale Santé, cf. le site du CHU de Rouen (http://www.chu-rouen.fr)
D'une manière ou d'une autre, des normes devront s'établir entre sites administratifs et une concertation nationale et internationale aura lieu pour définir des noyaux communs de fiches et des méthodes de labellisation. »

8.2 ICRA

“[www.icra.org]

Welcome ...

Welcome to the Internet Content Rating Association website, home of the RSACi rating and filtering system. Thank you for your interest. Please find below detailed information about how to use the RSACi filtering system within both Microsoft's and Netscape's browsers, what the different levels represent, background information about ICRA and where to go for our latest news and product development.

ICRA is a unique organisation. Our dual aims are to protect children from potentially harmful material while also protecting free speech on the Internet. We are an independent charity with offices in the USA and Europe. We are backed by major hi-tech companies and foundations including AOL, BT, Cable & Wireless, IBM, Microsoft and the Bertelsmann Foundation. We have also just secured funding from the European Commission to make the system more internationally acceptable and to translate it into many of the major languages of the world. “

9. ANNEXE 2 : ANALYSE DES LOGICIELS DE FILTRAGE DU COMMERCE

Sommaire

9.1 PREAMBULE

Pour permettre à l'internaute de bloquer le contenu de certains sites, les logiciels de filtrage installés sur des ordinateurs clients utilisent deux approches distinctes et parfois complémentaires :

1. Utilisation d'une liste d'adresses Internet classées selon la nature de leur contenu (sexe, racisme, etc.). Cette liste peut être gérée par des organismes indépendants comme le RSAC (Recreational Software Advisory Council), le PICS (Plateform for Internet Content Selection) ou encore Surfwatch.
2. Utilisation d'une liste de MOTS CLES interdits pour bloquer l'affichage de la page web avant son chargement dans le navigateur si une occurrence est détectée.

Ces listes de mots-clefs ou d'adresses Internet sont paramétrables par l'utilisateur ou l'administrateur du système informatique.

Les internautes peuvent également s'abonner à un fournisseur d'accès Internet qui propose ce type de service (EduNet).

9.2 DEMARCHE

L'analyse des outils de filtrage a consisté à :

- Établir la liste aussi exhaustive que possible des logiciels disponibles sur le marché. Cette liste a été obtenue par recherche à travers le Web,
- Analyse des documentations commerciales fournies par les éditeurs,
- Premier tri des produits pouvant répondre a priori à la problématique exposée par Yahoo ! Inc,
- Installation des produits retenus,
- Réalisation d'une batterie de tests et analyse des résultats.

Les chapitres suivants présentent les produits testés et une synthèse des résultats des tests.

9.3 LISTE DES LOGICIELS

Les produits sont regroupés selon leur mode de diffusion : les freewares, les sharewares et les logiciels classiques.

FREEWARE

- Prowler
- Zeeks
- We-blocker
- Internet Security 2000

SHAREWARE

- CyberPatrol
- CyberSitter
- Netnanny
- SurfWatch

DIFFUSION CLASSIQUE

- Internet Security 2000

9.4 LES SHAREWARES

1 - CYBERSITTER 2000 - ÉDITEUR : SOLID OAK SOFTWARE

Ce logiciel simple d'utilisation permet de bloquer la consultation de sites illégaux.

En outre, ses fonctionnalités lui permettent d'empêcher l'utilisation de certains mots-clefs dans les moteurs de recherche.

Son contrôle se limite au Web. Il ne peut verrouiller l'ensemble des systèmes de messagerie actuellement disponibles.

Malgré le fait que le logiciel de filtrage demande un redémarrage de l'ordinateur après avoir modifié certains réglages, fil reste le meilleur rapport qualité/prix des logiciels sharewares étudiés.

2 - NETNANNY - ÉDITEUR : NETNANNY

Ce filtre est d'une installation aisée. Il a la possibilité de modification des listes de mots-clefs et d'urls du logiciel.

Néanmoins, les tests réalisés ont démontré une faible fiabilité lors du blocage des pages.

De plus son module de filtrage est inopérant pour le navigateur Netscape. Dans tous les cas son prix est excessif

3 - SURFWATCH - ÉDITEUR : SURFWATCH SOFTWARE

Ce logiciel de filtrage utilise une base de données référençant les sites douteux, il peut grâce aux spécifications de l'utilisateur bloquer certaines adresses d'Internet.

Le module résident est capable de filtrer la saisie des mots-clefs interdits dans les champs de certains moteurs de recherche.

Une contrainte réside dans le fait qu'il faut redémarrer le PC à chaque modification des paramètres de configuration. De plus, le navigateur Netscape n'est pas supporté. Son prix est assez élevé par rapport à ses concurrents.

4 - CYBER PATROL – ÉDITEUR : MATTEL

Ce Cyber-patrouilleur disponible en Français n'est pas une référence du genre. Son système de filtrage se limite aux blocages d'urls sensibles paramètrable.

Le blocage des sites web n'est pas très fiable, parfois certains moteurs de recherche se retrouvent bloqué sans raison apparente. Le coût très élevé de ce logiciel ne se justifie pas dans le contexte de Yahoo ! Inc.

9.5 FREEWARES

1 - PROWLER - ÉDITEUR : AMSD / WEBKEYS

Ce logiciel gratuit, très puissant permet un paramétrage pointu du comportement du navigateur au cours de la consultation des pages Internet.

Lors d'une navigation sur un site interdit, le module résident affiche une page web spécifique invitant l'utilisateur à ne pas continuer son chemin. Le choix lui revient : de saisir mot de passe pour passer outre cette recommandation ou de stopper la consultation.

La gestion multi-utilisateurs des droits permet une flexibilité accrue. Il un très bon choix pour les utilisateurs d'Internet explorer car pour l'instant il n'est pas compatible avec Netscape.

2 - ZEEKS - ÉDITEUR : ZEEKS SOFTWARE

Comme la grande majorité des logiciels testés, cette solution utilise à la fois un contrôle sur les mots clefs et un filtrage des urls interdites.

Lors de son utilisation un bandeau de publicitaire s'intègre à Internet explorer et à chaque tentative de connexion à une page interdite, l'internaute est redirigé sur le site de l'éditeur.

Ce produit n'est pas compatible avec le navigateur de Netscape.

3 - WE-BLOCKER – ÉDITEUR : WE-WEB CORPORATION

Ce patrouilleur d'un nouveau genre permet un contrôle perspicace d'Internet. Chaque internaute peut enrichir la base de restriction de la communauté « We-Blocker », en inscrivant les urls prohibées dans la base de donnée de la We-Web corporation. Ainsi l'internaute participe au contrôle actif d'Internet, car cette adresse sera alors immédiatement inaccessible pour tous les utilisateurs du logiciel We-Blocker.

Le contrôle par mots-clefs quant à lui est n'utilise qu'une liste de mots enregistrés localement sur le poste de travail.

D'une installation et d'une utilisation simples, il reste le challenger de cette étude.

9.6 LOGICIEL CLASSIQUE

1 - INTERNET SECURITY 2000 – ÉDITEUR : SYMANTEC

Ce logiciel très complet permet un contrôle étendu de la connexion Internet, il intègre un firewall personnel, des systèmes de contrôle du trafic http.

Il se limite cependant à un filtrage d'Internet basé uniquement sur une liste d'url interdite.

Etant donnée le nombre de fonctions disponibles, il est difficile d'évaluer le coup réel de ce produit face à la concurrence.

9.7 CONTOURNEMENT DES LOGICIELS DE FILTRAGE

PAR L'UTILISATEUR

Chaque logiciel de filtrage utilise des méthodes propriétaires pour ce protéger du monde extérieur. Les utilitaires d'autocensure comme Cyber Patrol possèdent par exemple une base chiffrée permettant de contrôler l'intégrité et la confidentialité des sites bloqués. Cependant des manipulations très simples ou des outils spécifiques peuvent en venir à bout.

Un site web est devenu une ressource inépuisable de cet art (http://www.peacefire.org), il est curieux de constater que ce site est bloqué par la totalité des logiciels de filtrage.

PAR LE WEBMASTER

Aucun logiciel de filtrage ne peu garantir un blocage parfait de tous les contenus illégaux présents sur Internet, l'utilisation des dernières technologies de mise en page, la présence d'image contenant du texte immoral ou le chiffrement des documents disponibles en ligne peu permettre de passer à travers la totalité des logiciels du marché.

9.8 MISE A JOUR DES FILTRES

Il faut distinguer deux méthodes de mise à jour spécifique payante ou gratuite des logiciels de filtrage :

1. La mise à jour de la liste d'urls interdites, gratuite pour la plus part mais payante pour Cyber patrol et Netnanny.
2. La mise à jour du logiciel est souvent payante sauf pour les produits gratuits. Généralement, ils possèdent un module spécifique permettant d'effectuer cette opération.

9.9 SYNTHESE

SHAREWARES

Cybersitters 2000, répond parfaitement aux besoins de flexibilités et de filtrages nécessaires aux blocages de contenus illégaux.

Son interface simple, son choix d'options étendu et son prix le moins élevé du marché en font le meilleur candidat dans le contexte de Yahoo ! Inc.

FREEWARES

We-Blocker, est une solution très performante, son système de base de données contributif en font un outil de rêve pour les utilisateurs exigeants.

Il est à noter que plus le nombre des utilisateurs de ce logiciel sera important, plus son système de filtrage sera fin.

Prowler, est une très bonne alternative si l'on utilise uniquement Internet explorer. Il est exemplaire en matière de filtrage. L'installation ou l'utilisation de ce logiciel et très simple presque intuitive. Hélas, un tel logiciel n'existe pas sur Mac.

LOGICIELS A DIFFUSION CLASSIQUE

Le seul logiciel existant à l'heure actuelle dans ce mode de diffusion est Internet Security 2000 de Symantec. Son objectif premier étant la sécurisation d'un micro-ordinateur personnel connecté à Internet son champ d'action déborde très largement des fonctionnalités étudiées.

Il est une bonne solution de sécurité Internet pour un usage privée.

9.10 CONCLUSION

Certains logiciels se sont trouvés être assez décevant comme **Cyber patrol** qui ne permet pas de filtrage des mots-clefs dans une page web ou ceux qui restent d'une utilisation lourde pour le grand public comme **Surfwatch**, **Zeeks** ou Les solutions de filtrage natives d'Internet explorer et de Netscape communicator se limitent uniquement aux listes d'adresses Internet classées par le RSAC, le PICS ou encore Surfwatch et donc sont totalement inutilisable pour répondre aux objectifs de Yahoo..

La meilleure solution disponible à l'heure actuelle :

- **We-blocker** pour les plate-formes **Windows**
- **Surfwatch** pour les systèmes **Macintosh**

Restent les solutions de filtrage natives d'**Internet Explorer** et de **Netscape communicator** qui se limitent uniquement au tri selon une liste d'adresses Internet classées par les organismes tel que RSAC, le PICS ou encore Surfwatch. Cette solution ne semble donc pas pouvoir répondre complètement au objectifs de Yahoo! Inc.

Tableau récapitulatif

	Cyber Patrol	CYBERSitter	Net Nanny	SurfWatch
Caractéristique générales				
Prix (en dollar Us)	59.99	39.95	49.95	49.95
Version	4.04.004	2000.0.5.25	3.1	3.0
Type de licence	Shareware	Shareware	Shareware	Shareware
Langues	Francais	Anglais	Anglais	Anglais
Systéme d'exploitation				
Windows	3.1, 95, 98, NT	95, 98, 2000, NT	3.1, 95, 98	95, 98
Macintosh	System 7.1	Non	Non	< Mac Os 9
Navigateurs *				
Internet explorer	Oui	Oui	Oui	Oui
Netscape	Oui	Oui	Non	Non

Tableau récapitulatif

	Prowler	Zeeks	We-Blocker	Internet Security 2000
Caractéristique générales				
Prix (en dollar Us)	-	-	-	69.95
Version	4.01	2.0	1.6.2	2.0
Type de licence	Freeware	Freeware	Freeware	Classique
Langues	Anglais	Anglais	Anglais	Anglais
Systéme d'exploitation				
Windows	95, 98, NT	95, 98	95, 98, 2000, NT	95, 98, 2000, NT
Macintosh	Non	Non	Non	Non
Navigateurs *				
Internet explorer	Oui	Oui	Oui	Oui
Netscape	Non	Non	Oui	Oui

Lieu de téléchargement ou d'achat

Cyber Patrol	http://www.cyberpatrol.com
CYBERSitter	http://www.cybersitter.com
Net Nanny	http://www.netnanny.com
SurfWatch	http://www.surfwatch.com
Prowler	http://www.webkeys.com/download.htm
Zeeks	http://www.zeeks.com
We-Blocker	http://www.we-blocker.com
Internet Security 2000	http://www.symantec.com/sabu/nis/nis_pe/

* L'ensemble des tests furent réalisés sous Windows 98 àla date du 10/07/00