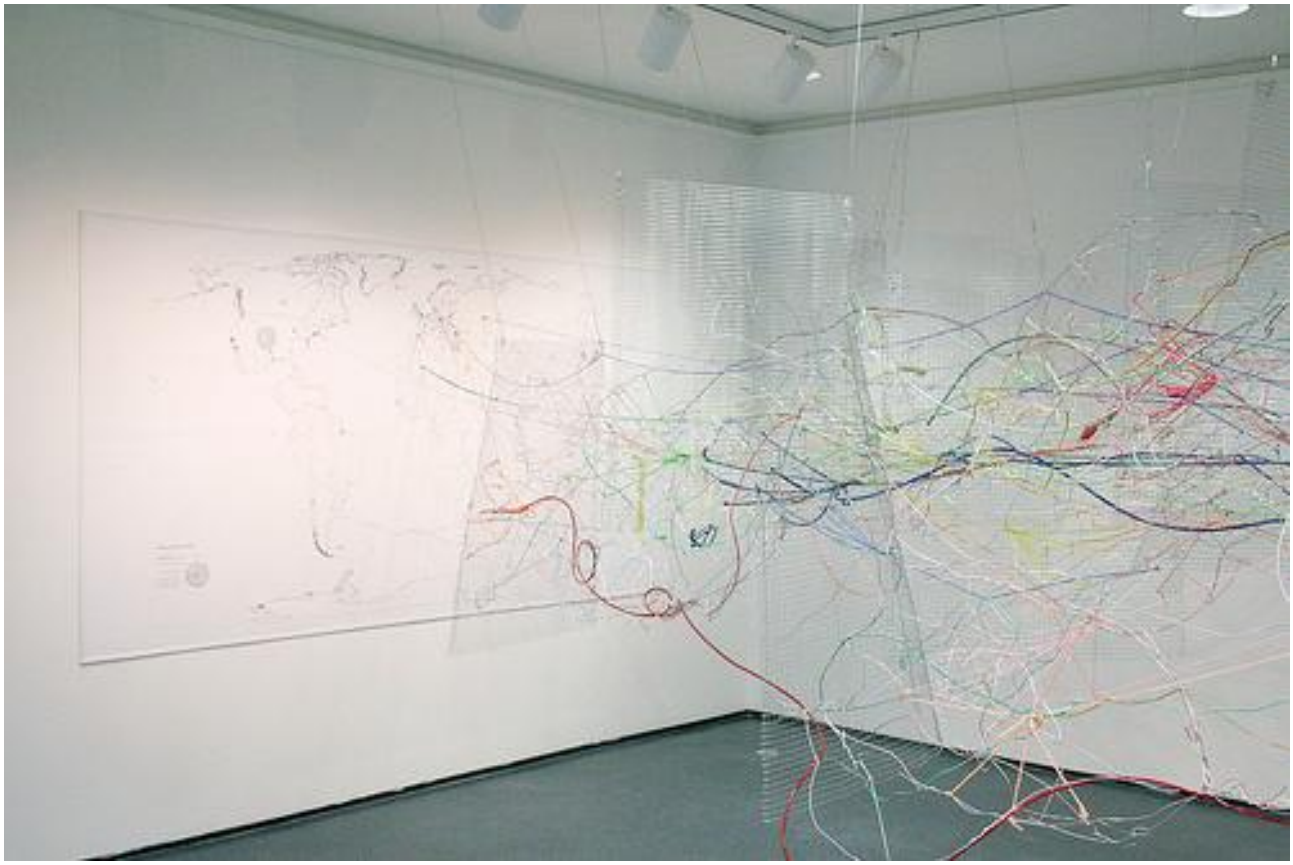
# Principe, intérêts, limites et risques du filtrage hybride à des fins de blocage de ressources pédopornographiques hébergées sur des serveurs étrangers



Auteur principal : Christophe Espern

Merci aux abonnés du canal #fdn et de la liste FRnOG qui ont contribué

Photo de couverture : Mapping the internet, by Fausto Fernós
http://www.flickr.com/photos/feastoffools/2126692786/

Pour discuter de cette note : #laquadrature sur freenode

Nota bene

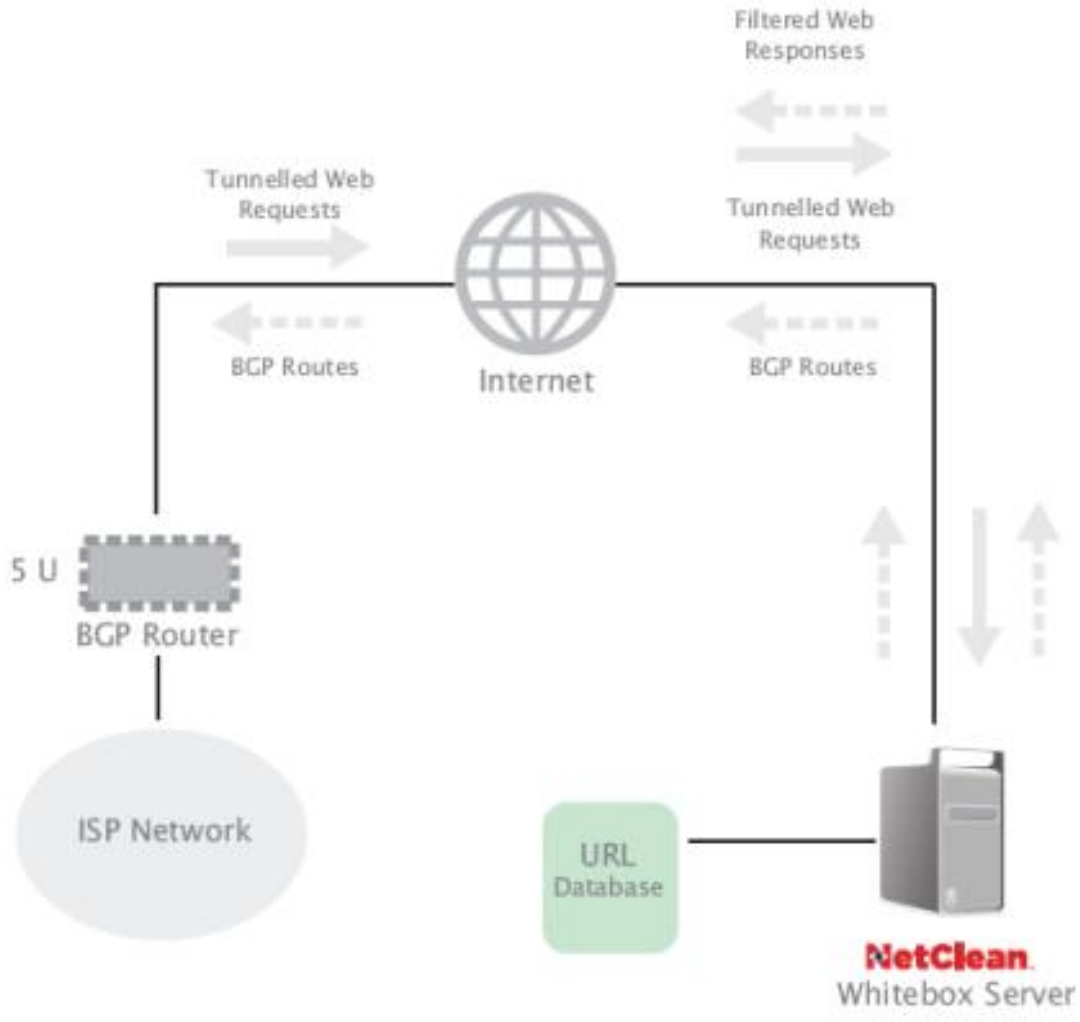
Ce document ne traite pas du filtrage par analyse des contenus, ni du filtrage de protocoles (sauf cas de dommages collatéraux), mais uniquement des technique de filtrage sur l'IP, le nom de domaine ou l'URL.

**Principe**

Au travers d'enquêtes ou sur signalement d'internautes, les services de police maintiennent une liste noire d'URL pointant sur des ressources pédopornographiques. Cette liste est communiquée aux Fournisseurs d'Accès à Internet (FAI) qui empêchent l'accès à ces ressources à leurs abonnés.

Concrètement, à partir de la liste noire, les FAI récupèrent la liste d'adresses IP correspondant aux noms de domaines où sont hébergées les ressources à bloquer. Puis, ils envoient une commande à leurs routeurs via le protocole Border Gateway Protocol (BGP) pour les reconfigurer, afin que toute demande d'accès à une des *IP suspectes* soit routée vers la plate-forme de filtrage, et non plus relayée directement au serveur demandé par l'utilisateur.

Ainsi, lorsqu'un abonné demande à accéder à une ressource hébergée sur un site dont l'adresse IP a été associée par un FAI à celui d'une URL fichée par la police, la requête est redirigée par les routeurs du FAI vers la plate-forme de filtrage qui bloque la communication si la ressource correspondante est dans la liste noire, et qui sinon relaie la communication.




L'architecture du système Netclean, typique d'un filtrage hybride

http://www.netclean.com/EN/documents/NetClean_Whitebox_Tech_EN.pdf

### Intérêts et limites

Cette technique est appelée filtrage hybride car elle combine plusieurs techniques pour répondre aux problèmes de surblocage inhérents au filtrage par IP ou par DNS, tout en évitant les coûts de déploiement des autres techniques de filtrage par URL.

Il est ainsi possible de ne bloquer qu'une photo d'une page web, mais les infrastructures nécessaires n'ont rien de comparable avec celles mises en œuvre dans des pays comme la Chine ou l'Arabie Saoudite : la plate-forme accueillant le trafic à filtrer ne traite qu'une partie du trafic grâce au tri préalable effectué par les routeurs sur les adresses IP.

Cette technique reste contournable par l'internaute en passant par des serveurs mandataires hébergés à l'étranger, ce qui peut se faire en quelques clics et être ensuite activé par défaut sur l'ordinateur. L'éditeur du site filtré peut aussi prendre des contre-mesures, par exemple, basculer sur le protocole https, rendant ainsi l'url complète indéchiffrable pour la plate-forme de filtrage. Il peut aussi générer des url uniques à la demande.

Une étude universitaire [Clayton, Cambridge, 2005] suggère, de plus, qu'au Royaume-Uni où de tels systèmes sont déployés, les fournisseurs de contenus filtrés utilisent déjà des techniques de leurres afin d'identifier les ordinateurs des services chargés de remplir la liste noire, pour pouvoir mieux leur masquer leurs sites par la suite.

### Risques

Le coût global dépend beaucoup du trafic à traiter, de l'architecture des opérateurs, et des effets en cas de surcharge, d'erreur de configuration, de détournement ou d'attaque du système, risques qui sont loin d'être théoriques comme cette partie entend le montrer.

Ces dommages peuvent impliquer la responsabilité de l'État français et perturber des activités économiques légitimes, et parfois critiques, de façon conséquente.

#### Risques d'engorgements

Le trafic dérivé vers les serveurs de filtrage doit pouvoir être absorbé, ce qui implique de le prévoir pour dimensionner la plate-forme Or le trafic de sites très fréquentés n'ayant rien à voir avec le site ciblé peut être dérivé, une adresse IP pouvant être partagée.

Une étude universitaire [Edelman,Harvard, 2003] soulignait que « *plus de 87% des noms de domaines actifs partagent leurs adresses ip (i.e. : les serveurs web) avec un ou plusieurs domaines additionnels, et plus des 2/3 des noms de domaines actifs partagent leurs adresses avec 50 domaines additionnels ou plus* » . Depuis, cette proportion n'a pu que croître.

L'estimation du trafic à supporter en fonctionnement normal est donc délicate, d'autant plus que les éditeurs de site pédopornographiques les déplacent d'adresse IP en adresse IP, comme le soulignent les vendeurs de filtres. Il s'agit de déjouer les systèmes déjà déployés, mais cela rend aussi de plus en plus coûteux le filtrage de premier niveau en forçant à une surveillance d'un nombre de plus en plus grand d'IP.

Le trafic à absorber peut par ailleurs subitement augmenter si l'une des IP suspectes est victime d'une attaque informatique visant à le saturer (*déni de service*) depuis le réseau filtré.

En plus d'avoir à supporter les attaques ciblant effectivement des sites filtrés ou partageant la même IP qu'un site filtré, le système de filtrage pourrait être visé directement par une organisation criminelle, en représailles. L'attaque ciblera des IP connues de l'attaquant comme présentes dans la liste, voire mises dans la liste via une dénonciation de circonstance.

### Risques liés à l'utilisation du protocole BGP

L'utilisation de commandes BGP pour redéfinir des routes à des fins de filtrage de contenus n'est pas une utilisation pour laquelle le protocole BGP, d'utilisation délicate, a été pensé.

Pour preuve, lorsque le Pakistan a ordonné le blocage de l'accès à des caricatures de Mahomet hébergées sur le service YouTube, un opérateur pakistanais a envoyé une commande BGP à des équipements mal paramétrés : ils ont propagé la demande aux réseaux d'opérateurs hors juridiction pakistanaise. L'accès à YouTube a alors été interdit pendant plusieurs heures dans plusieurs pays du monde. Cet événement a permis de mettre en évidence des risques pour la sécurité nationale, comme l'ont relevé des spécialistes réseaux.

*« Un petit groupe de personnes pourrait s'emparer d'une chaîne de routeurs compatible BGP qui auraient été piratés pour envoyer des préfixes BGP à tout l'internet. Le résultat ne ferait pas tomber Internet – mais il pourrait causer des perturbations à grande échelle – ce qui est exactement ce que vous recherchez si vous souhaitez que votre attaque terroriste ait plus d'impact. Autrement dit, la couverture presse sur cette faille du préfixe BGP met en lumière un vecteur d'attaque qui peut causer de sérieux dégâts pendant une période où les gens auront le plus besoin d'internet. »* (*YouTube Black Hole – What's the real point?* http://www.getit.org/wordpress/?p=82).

### Incompatibilité avec l'architecture technique et contractuelle

Le fait de demander aux opérateurs de modifier en permanence leur configuration de routage n'est pas compatible avec l'utilisation de techniques d'optimisation devenues standard comme l'agrégation de routes. C'est particulièrement vrai en France au regard du nombre d'accords de *peering* passés par les opérateurs entre eux, et dans lesquels les règles d'agrégation font l'objet de clauses contractuelles spécifiques.

Par ailleurs, le fait que les organisations criminelles utilisent une technique connue sous le nom de Fast Flux (http://en.wikipedia.org/wiki/Fast_flux) – visant à changer très régulièrement l'association nom de domaine-adresse IP – impliquera l'envoi fréquent de commandes aux routeurs pour reconfigurer les routes, multipliant d'autant les risques de dommages et la complexité des configurations à maintenir.

### Risques d'exposition de la liste noire

Une étude universitaire [Clayton, Cambridge, 2005] a montré que les systèmes de filtrage hybride en production au Royaume-Uni (CleanFeed, WebMinder) appliquent un traitement particulier aux communications électroniques des utilisateurs, ce qui fait que « *le système peut être utilisé comme un oracle pour trouver efficacement des sites web illégaux* ».

L'auteur a établi qu'il était possible pour un abonné anglais d'obtenir anonymement en 24h00 la liste de tous les sites russes filtrés à la demande de la police anglaise. [1]

En plus de présenter l'énorme risque que cette liste circule en clair sur internet, voire soit vendue avec une notice expliquant aux utilisateurs filtrés comment contourner le système, ou avec un logiciel permettant d'enrichir sa propre liste, cette faille peut être exploitée pour faciliter le contournement par les éditeurs de sites filtrés ou pour maximiser une attaque informatique, car rendant l'observation du système, et donc de ses défauts, plus simple.

Une des sociétés concernées par cette faille a annoncé après la publication de l'étude que le problème était résolu. L'auteur a alors fait une mise à jour montrant que ce n'était pas le cas.

## Conclusion

La mise en place d'un filtrage hybride, si elle apparaît séduisante sur le papier, présente des risques conséquents pour une efficacité limitée.

Son coût direct et indirect pourrait à l'usage exploser. Utilisateurs comme fournisseurs de contenus pédophiles pourront toujours la contourner facilement, mais aussi l'attaquer.

Sa mise en œuvre risque de durcir les techniques utilisées par les pédophiles et les fournisseurs de contenus pédopornographiques pour se cacher et entraver l'activité des enquêteurs. Elle présente en outre des risques de fuite de la liste noire.

Les spécialistes réseaux interrogés sont consternés que cette technique soit envisagée, vu ses failles et les risques qu'elle présente pour le réseau tout entier. Sa mise en œuvre constituerait pour eux une régression. Ils considèrent qu'il serait irresponsable que l'État encourage cette technique de filtrage et engage sa responsabilité si elle était utilisée par un opérateur.

*Post scriptum : plus largement, les acteurs techniques qui se sont exprimés considèrent que l'idée d'un filtrage « cœur de réseau » doit être définitivement abandonnée. D'une part, les autres techniques de ce type (voire Annexe I) impliquent soit un surblocage important ou très important et une efficacité très limitée (DNS) : soit un surblocage très important et une efficacité limitée (IP) ; soit un coût exorbitant, voire pharaonique, et une efficacité limitée (proxies généralisés, RST). Mais surtout, un tel filtrage irait à l'encontre de l'architecture même d'Internet et de son développement souhaitable, en recentralisant le trafic.*

*Au cours de la discussion qui a contribué à la rédaction de cette note, plusieurs abonnés à la liste FRnOG ont alors évoqué l'installation de dispositifs de filtrage par les FAI dans les boîtiers de connexion des abonnés (les boxes). Pour certains professionnels qui se sont exprimés, cette technique serait la seule envisageable sur le plan architectural.*

*Cette technqiue pose cependant des problèmes concrets, dont certains ont été abordés sur la liste FRnOG. Leur étude fera l'objet d'un développement ultérieur, comme l'examen des aspects juridiques et politiques inhérents à tout projet gouvernemental de filtrage d'internet.*

---

1 L'auteur de l'étude citée n'a pas cherché à obtenir une telle liste limitant ses investigations de chercheur en sécurité à la mise en évidence de la faille, notamment pour des raisons légales. Son estimation du temps nécessaire pour obtenir la liste des sites russes fichés se base donc sur un échantillon réduit de sites russes découvert via une exploration réseau volontairement interrompue, et sur les chiffres fournis par l'IWF qui maintient la liste noire (25% des sites fichés seraient russes).

# Annexe 1 : Ressources utilisées

## Études universitaires

**[Edelman, Harvard, 2003] - Filtrage sur l'adresse IP**

Edelman, B.: Web Sites Sharing IP Addresses: Prevalence and Significance.Berkman Center for Internet and Society at Harvard Law School, 2003.

http://cyber.law.harvard.edu/archived_content/people/edelman/ip-sharing/

**[Dornseif, Düsseldorf, 2003] – Filtrage sur le nom de domaine par DNS**

Dornseif, M.: Government mandated blocking of foreign Web content. In: von Knop, J., Haverkamp, W., Jessen, E. (eds.): Security, E-Learning, E-Services: Proceedings of the 17. DFN-Arbeitstagung Äuber Kommunikationsnetze, Dusseldorf, 2003.

http://md.hudora.de/publications/200306-gi-blocking/200306-gi-blocking.pdf

**[Clayton, Cambridge, 2005] - Filtrage hybride (Cleanfeed, WebMinder, NetClean)**

Clayton, Failures in a Hybrid Content Blocking System. University of Cambridge, Computer Laboratory, 2005

http://www.cl.cam.ac.uk/~rnc1/cleanfeed.pdf

**[Clayton, Cambridge, 2006] – Filtrage sur l'URL par injection de paquet RST**

Clayton, Murdoch, Watson : Ignoring the Great Firewall of China. University of Cambridge, Computer Laboratory, 2006

http://www.cl.cam.ac.uk/~rnc1/ignoring.pdf

## Autres ressources

**Vue d'ensemble**

The worst part of censorship is XXXXX : Investigating large-scale Internet content. 23C3, Berlin/Germany, December 29th, 2006

http://events.ccc.de/congress/2006/Fahrplan/events/1473.en.html

**Discussions entre acteurs techniques du réseau sur la liste FRnOG**

Charte sur la confiance en ligne" vers une division de l'inter-net ?

http://www.mail-archive.com/frnog@frnog.org/msg02883.html

Filtrage via BGP shunt : quelle faisabilité ?

http://www.mail-archive.com/frnog@frnog.org/msg02939.html

Ping: il n'y a plus personne ? (à propos du YouTube blackhole)

http://www.mail-archive.com/frnog@frnog.org/msg02441.html

# Annexe 2 : Autres techniques de filtrage par les FAI

### Le filtrage sur le nom de domaine (filtrage par DNS pour Domain Name Server)

Avec cette technique, ce n'est pas le contenu illégal qui est filtré, mais l'intégralité du domaine internet qui l'héberge (ex : geocities.com au lieu de geocities.com/siteperso/pedo.jpg).

Concrètement, l'intégralité d'un site hébergeant des millions de pages personnelles pourrait disparaître de la vue des internautes français pour une image non retirée dans les délais imposés par l'administration française (certains évoquent 24h00 de délai, d'autres aucun).

Cette technique peut également entraîner le blocage de sous-domaines (ex : pagesperso.free.fr) en fonction de la façon dont la requête est rédigée. Elle peut interdire des communications non visées par la requête, par exemple interdire l'envoi et la réception de courriels relatifs au domaine, et non plus seulement l'accès aux pages web hébergées.

Une étude universitaire [Dornseif, Düsseldorf, 2003] étudiant le cas du filtrage d'un site nazi ordonné par une autorité allemande a montré que sur les 27 FAI étudiés, tous ont fait au moins une erreur lorsqu'ils ont configuré leurs filtres DNS. Les FAI n'ont pas bloqué le site souhaité (sous-blocage), ont bloqué des sites ou des protocoles non visés par la requête (surblocage), ou étaient à la fois en situation de sous-blocage et de surblocage.

Ainsi, sur 27 fournisseurs d'accès, 45% étaient en situation de surblocage et de sous-blocage, 55% était "uniquement" en situation de surblocage, 16 FAI sur 27 (59%) bloquaient la communication courriel vers plusieurs domaines et tous bloquaient l'adresse de courriel par défaut du site ciblé, alors que tout cela n'était pas demandé par le juge.

L'étude soulignait aussi que *« le contenu web est très volatile. Les serveurs web sont réorganisés, les domaines ont de nouveaux propriétaires. Ceci a été très clairement démontré dans le cas des requêtes de blocage du site web www.front14.org : à l'automne 2001, ce site contenait un portail d'extrême-droite, mais au printemps 2002, il y avait un catalogue web à cette adresse, sans agenda politique. Ceci souligne la nécessité d'identifier les pages à bloquer pas seulement par leur emplacement, mais par leur contenu actuel. »*

Les opérations nécessaires au blocage par DNS sont relativement simples, bien que la complexité et la maintenance engendrées, et donc le coût global, dépendent là aussi des configurations actuelles des opérateurs. L'efficacité de cette technique est très limitée. Il suffit d'une manipulation triviale sur l'ordinateur de l'utilisateur pour définitivement passer outre. Les fournisseurs de contenus pédopornographiques n'ont qu'à proposer des liens utilisant des adresses IP à la place du nom du domaine pour déjouer le système.

Plus d'information : voir Annexe 1 - [Dornseif, Düsseldorf, 2003]

### Le filtrage sur l'adresse IP

Il s'agit pour les opérateurs de maintenir une liste d'adresses IP ou de blocs d'adresses IP pour lesquels leurs routeurs ne vont pas transmettre les paquets, mais simplement les ignorer. Ainsi, tout échange de données passant par un routeur appliquant ce filtrage est impossible.

Cette technique bloque tout accès à un serveur ou à un groupe de serveurs, et ne permet pas de traiter séparément des contenus différents ou des sites web différents sur une même machine. Le surblocage a de très grandes chances d'être conséquent, voire très conséquent.

Une étude universitaire [Edelman, Harvard, 2003] soulignait ainsi que : « *More than 87% of active domain names are found to share their IP addresses (i.e. their web servers) with one or more additional domains, and more than two third of active domain names share their addresses with fifty or more additional domains. While this IP sharing is typically transparent to ordinary users, it causes complications for those who seek to filter the Internet, restrict users' ability to access certain controversial content on the basis of the IP address used to host that content. With so many sites sharing IP addresses, IP-based filtering efforts are bound to produce "overblocking" -- accidental and often unanticipated denial of access to web sites that abide by the stated filtering rules.* » Depuis 2003, le nombre de sites web ayant explosé et les techniques de mutualisation d'adresses IP s'étant définitivement répandues, les risques de surblocage ont encore augmenté.

Cette technique est contournable par l'utilisateur à l'aide de serveurs mandataires situés à l'étranger, que l'utilisateur peut utiliser via *tunnelling* ou via des serveurs web dédiés. Ces outils ne peuvent pas être filtrés dans une démocratie, car offrant une fonctionnalité générique. Les fournisseurs de contenus contournent notamment le filtrage IP avec des automates qui réassignent leurs noms de domaines à de nouvelles IP à intervalles réguliers.

Plus d'informations : voir Annexe 1 - [Edelman, Harvard, 2003]

**Le filtrage sur l'URL via serveurs mandataires (proxies) généralisés**

Toutes les requêtes des internautes du pays passent par des serveurs de filtrage qui bloquent les communications relatives à une URL identifiée. Contrairement à un filtrage hybride, il n'y a donc pas de "tri" préalable sur l'adresse IP. Cette technique implique des plate-formes de filtrage conséquentes avec redondance des serveurs car tout le trafic web est traité.

Cette technique est celle choisie par l'opérateur national en Tunisie et en Arabie Saoudite. Le coût de mise en place de cette technique serait exorbitant dans un environnement concurrentiel comme la France où plusieurs opérateurs importants coexistent. Elle constituerait une régression d'un point de vue architectural. La société Noos utilisait il y a quelques années une tel technique sur son réseau à des fins de *cache*. Elle a été abandonnée, à cause du surblocage et du coût croissant au fil de l'extension du réseau. Cette technique de filtrage se contourne via l'utilisation de serveurs mandataires décrits précédemment.

**Le filtrage sur l'URL par injection de paquets RST**

Les URL des sites web visités sont analysées en regard d'une liste de mots-clés et d'une liste noire d'URL, et les routeurs par lesquels transite la connexion envoient au client et au serveur un paquet RST, ce qui a comme conséquence la clôture de la connexion TCP.

La connexion est close dès qu'elle est établie et reconnue comme à filtrer, aucun contenu ne peut être échangé. Cela nécessite que tout le trafic à contrôler passe par des infrastructures réseau maîtrisées par les autorités de contrôle. C'est une des techniques utilisées en Chine.

Pour plus d'informations : voir Annexe 1 - [Clayton, Cambridge, 2006]