HTTPS comme critère de classement : que faut-il en penser ?

Retour au sommaire de la lettre

Domaine :	Recherche	**Référencement**
Niveau :	Pour tous	**Avancé**

Le mois dernier, Google a annoncé que les sites sécurisés (protocole SSL/TLS) recevraient dorénavant un "coup de boost" dans l'algorithme de pertinence du moteur. Si cet avantage est aujourd'hui très minime, il pourrait croître dans les mois qui viennent. Alors, faut-il sauter le pas et faire passer son site en version sécurisée ? Quels sont les risques ? Les solutions ? Les coûts ? Voici un panorama complet des réponses à toutes ces questions...

Le 6 août 2014, Google a annoncé sur son blog Webmaster Central que le moteur de recherche avait commencé à intégrer le support du protocole SSL/TLS comme un critère de son algorithme de classement (http://www.abondance.com/actualites/20140807-14164-les-sites-securises-https-seront-mieux-positionnes-google.html). Google est donc censé accorder un "bonus" aux sites qui encryptent et authentifient la communication entre leur site et les navigateurs des internautes.

Pourquoi Google a-t-il pris cette décision ? Quelle est l'importance du bonus accordé par le moteur aux pages en https:// dans son algorithme ? Comment doit-on adapter son site pour faire face à cette nouvelle donne ? Quels sont les avantages et les inconvénients de basculer un site en https:// ? Nous allons nous efforcer dans l'article qui suit de répondre à toutes ces questions.

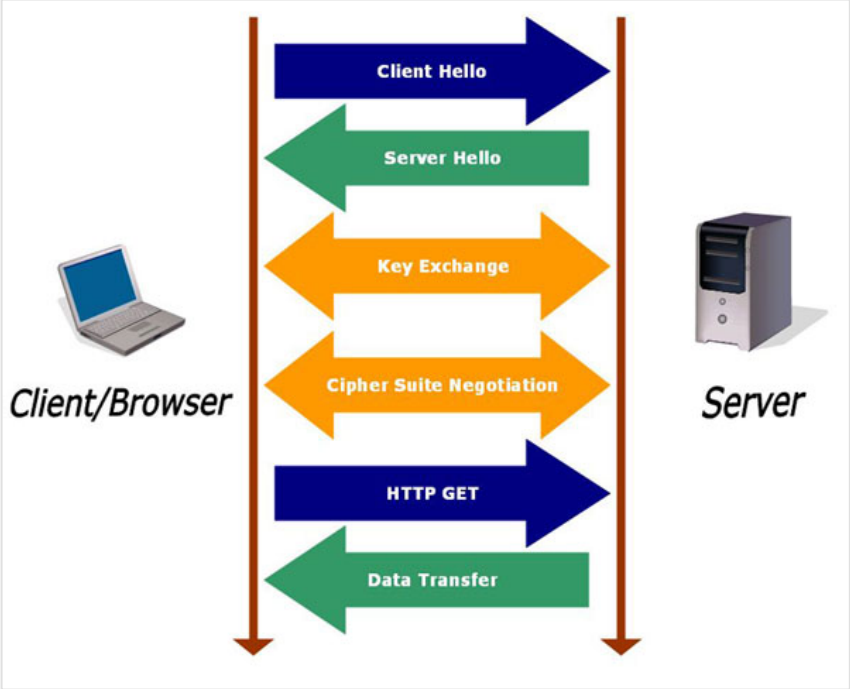
Qu'est-ce que le protocole SSL / TLS ?

Avant de commencer à parler des motivations qui expliquent cette nouvelle orientation de Google, il faut rappeler quelques informations sur le protocole SSL/TLS, qui reste finalement assez mal connu.

SSL/TLS est un protocole permettant de sécuriser les échanges de données entre le navigateur de l'internaute et un serveur web (et accessoirement un site web). Le protocole SSL (*secure sockets layer*) a été inventé dans un premier temps par Netscape. Mais suite au rachat des brevets de Netscape par l'IETF en 2001, le nom officiel de ce protocole est devenu TLS (*transport layer security*).

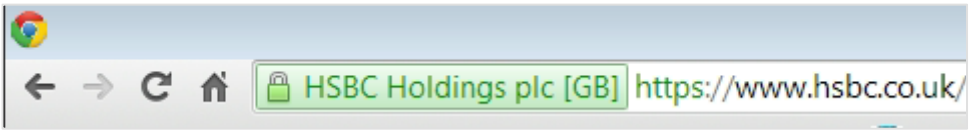
Ce qui signifie que très souvent, ce que l'on appelle "protocole sécurisé SSL" désigne en réalité le protocole TLS (mais ce nom a eu du mal à s'imposer).

La sécurisation apportée par SSL/TSL est obtenue par l'utilisation conjointe de deux approches :
- le chiffrement (le cryptage) des données avec une méthode asymétrique (basée sur une clé publique et une clé privée).
- l'authentification du serveur web auquel se connecte le navigateur de l'internaute.



Un schéma décrivant le déroulement d'une connexion sécurisée : elle commence par un « handshake » : le client (votre navigateur) demande une page (client hello) et le serveur lui répond (server hello) en communiquant son paramétrage. Le handshake se poursuit par l'échange des clés de chiffrement. Enfin, le serveur pourra répondre à un GET du client par l'envoi de la page en mode crypté.

On reconnait qu'une page web est sécurisée par le protocole http:// à l'aide de plusieurs indices :
- l'URL commence par https:// au lieu de http:// ;
- un petit cadenas fermé apparaît devant l'URL sur la plupart des navigateurs ;
- si le certificat appartient à une hiérarchie de certificats connue par le navigateur, un affichage supplémentaire permet d'identifier que la page consultée appartient à un domaine identifié :



Exemple de l'affichage obtenu sur chrome en mode sécurisé quand le certificat est reconnu par le navigateur.

Pourquoi Google veut-il utiliser ce critère dans son algorithme de classement ?

Google a évoqué pour la première fois sa volonté de promouvoir le protocole SSL/TLS au cours de la conférence Google I/O en juin 2014. Le principal argument évoqué était de rendre le web "plus sûr" en incitant les webmasters à utiliser ce protocole pour toutes les pages de leur site web. Nous verrons plus loin que TLS n'est pas la panacée universelle en matière de sécurité des échanges de données : c'est un pas dans la bonne direction, mais c'est loin d'être LA solution ultime.

Matt Cutts avait évoqué le sujet plus tôt lors du SMX WEST en mars 2014. En réponse à une question de Barry Schwartz du site Seroundtable.com, il avait révélé qu'il était en faveur de l'utilisation de ce critère mais que cela faisait débat au sein des équipes de Google (http://www.abondance.com/actualites/20140416-13817-google-privilegier-les-

sites-securises-https-algorithme.html) : "*At the end of the session, I [Barry Schwartz] asked Matt if this means Google is looking to give sites that enable SSL a ranking boost. Matt Cutts shrugged his shoulders and explained that if it was his choice, he would make it so. But he said, it is far from happening and there are people at Google that do not want this to happen. On one hand, if Google announced they would give a ranking boost to SSL sites, it would encourage a ton of sites to go SSL, which would be a good thing. On the other hand, some older sites are hard to make SSL and they would feel at a disadvantage.*"

Le débat interne a donc visiblement été tranché, et la décision a été prise de donner officiellement un bonus aux pages en https://.

Depuis l'annonce du 6 août dernier, il est plus clair que Google veut utiliser ce critère comme un signal pour son algorithme de classement. Et dans cette optique, l'objectif de cette campagne devient également plus clair. En effet, quand on se demande qui risque d'accepter de sécuriser son site et de payer un certificat à prix d'or, on identifie plutôt des sites disposant des moyens et de la motivation pour le faire. Ce qui sélectionne plutôt des sites "légitimes".

Et qui risque d'être gêné par ce changement ?

- Les sites illégaux, de contrefaçon, clandestins, etc., qui ont peu de chance d'obtenir un certificat de la part d'une autorité sérieuse, faute de pouvoir ou de vouloir fournir les documents nécessaires pour identifier les sites et leur propriétaires.

- Les sites de spam dont l'économie repose sur des tactiques de type "churn and burn", reposant sur la création de galaxie de sites qui vont se positionner un moment en tête des résultats, jusqu'à leur déclassement (par Penguin ou un autre filtre ou pénalité). Il suffit ensuite de recommencer avec un nouveau site et une nouvelle stratégie pour poser des backlinks. Evidemment, ici encore, le processus de certification pose problème, et le coût du certificat peut rendre ce genre de tactique black hat beaucoup moins tentante car bien moins rentable.

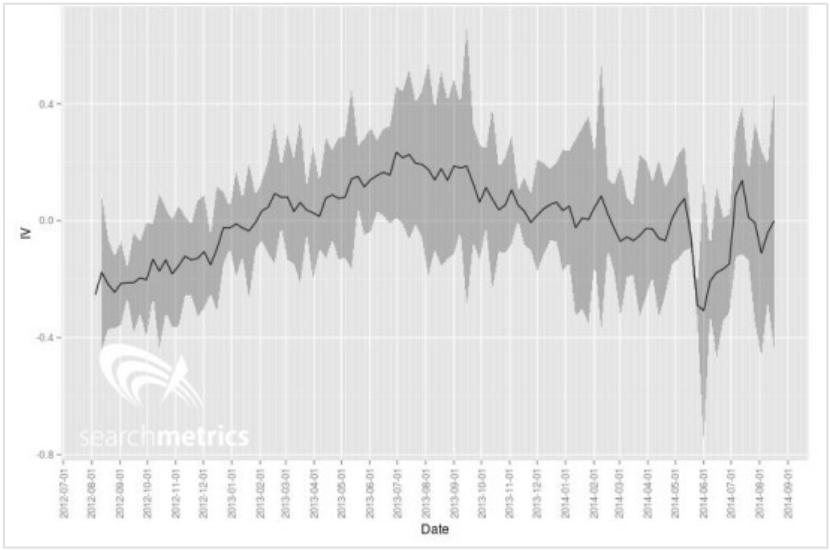
Le communiqué de Zineb Ait Bahajji et Gary Illyes précise qu'utiliser le "https://" comme un signal a été testé, et que ces tests ont été concluants. Cela signifie en clair que les résultats ont été améliorés par l'exploitation de ce critère : *"C'est pour cela qu'au cours des derniers mois, nous avons réalisé des tests en considérant l'utilisation de connexions sécurisées et chiffrées sur les sites en tant que signal dans nos algorithmes de classement. Nous avons pu observer des résultats positifs, et c'est pourquoi nous commençons à utiliser le protocole HTTPS en tant que facteur de positionnement."*

Quelle est l'importance du bonus accordé par Google aux sites sécurisés via SSL/TLS ?

Pour le moment, de l'aveu même des auteurs du communiqué d'annonce, ce critère a encore très peu de poids : *"Pour l'instant, cet indicateur a très peu de poids, et ce afin de laisser le temps aux webmasters de passer au protocole HTTPS. Il concerne moins de 1 % des requêtes mondiales, et il est moins important que d'autres indicateurs tels que le contenu de haute qualité. Mais au fil du temps, il est possible que nous décidions de lui donner une plus grande importance, car nous aimerions encourager tous les propriétaires de sites Web à passer du protocole HTTP au protocole HTTPS pour assurer la sécurité de tous les internautes sur le Web."*

La tactique utilisée ici est limpide : comme le niveau d'emploi du protocole https:// est encore faible, l'emploi de ce critère en tant que signal ne permet pas encore de l'utiliser comme un critère discriminant efficace. Google a donc décidé de communiquer dans un premier temps sur sa volonté d'utiliser le signal dans le futur, pour renforcer son utilisation. La même approche a été utilisée pour la vitesse de chargement des pages.

Dans ces conditions, il n'est pas étonnant qu'une étude de Marcus Tober, le fondateur de la société Searchmetrics, confirme qu'il semble impossible de détecter un impact de ce nouveau signal :
http://blog.searchmetrics.com/us/2014/08/29/https-vs-http-analysis-do-secure-sites-get-higher-rankings/



Une des courbes issues de l'analyse statistique de l'équipe de Marcus Töber : la corrélation entre ce critère et un gain de positions n'est pas concluante pour le moment.

Quels sont les avantages et les inconvénients d'un site en https:// ?

Utiliser un protocole d'échange crypté et sécurisé pour ses pages web représente, quoi qu'il arrive, un progrès pour les utilisateurs d'un site. Mais il faut souligner que la sécurité apportée par ce protocole est loin d'être suffisante pour préserver à 100% les utilisateurs d'un piratage. Rien n'empêche en particulier pour un bon hacker d'utiliser une attaque dite "man in the middle" pour intercepter les échanges entre un utilisateur et un site. Mais cela rend la tâche des hackers plus difficile (surtout si l'on active en plus le support du protocole HSTS).

Par contre, le protocole SSL/TLS pour les échanges en http:// présente quelques inconvénients qui ont empêché jusqu'ici son adoption par une majorité de webmasters. En règle générale, les sites limitent de plus l'emploi du protocole sécurisé aux pages de type "panier" ou "paiement", afin de rassurer leurs utilisateurs dans les phases de transaction.

Précisons au passage que Google annonce vouloir accorder un coup de pouce aux PAGES en https://, et non aux SITES comme cela a été dit par erreur dans plusieurs commentaires au sein de la communauté SEO. Il existe une confusion car certains webmasters pensent que sécuriser un site signifie employer https:// sur certaines zones du site (backoffice, pages de transaction, panier, compte personnel de l'utilisateur) mais pas sur la totalité des pages. Ce que cherche à obtenir Google ici, c'est l'emploi du protocole SSL/TLS sur toutes les pages du site, et non sur quelques pages dites "sensibles".

Parmi les reproches adressés au protocole, le plus fréquent est qu'il augmente le délai de téléchargement des pages.

En pratique, la phase de handshake, d'échange de clés et le cryptage crée un délai supplémentaire (et accessoirement, une charge serveur accrue). Mais ce délai supplémentaire ne dépasse pas dans la pratique quelques dizaines de millisecondes. Les webmasters inexpérimentés observent souvent une dégradation beaucoup plus sensible des performances. En général c'est dû au fait qu'ils ont oublié de changer les en-têtes de leurs pages pour permettre la mise en cache des pages en https:// . L'impossibilité de mise en cache des pages en https:// est une légende urbaine tenace : on peut "cacher" ces pages en https:// exactement comme des pages en http:, mais il faut utiliser des commandes "ad hoc".

Par ailleurs, le protocole est un peu plus exigeant quant à l'intégrité des données envoyées que le protocole http://, et cela se voit lorsque la bande passante vient à manquer ou sur des connexions lentes. Souvent, le navigateur arrivera à afficher une page incomplète en http://, mais pas en https://.

Le principal inconvénient : le problème des pages composites

L'un des reproches adressés aux webmasters au protocole https:// est l'envoi intempestif de messages par le navigateur sur des pages sécurisées avertissant l'utilisateur que la page contient "à la fois des éléments sécurisés et non sécurisés". Ce message désagréable s'affiche lorsqu'une page comporte des frames ou des iframes ou des appels externes à des pages non sécurisées. Ce qui ne manque jamais d'arriver quand un site affiche des flux externes ou - et c'est encore plus fréquent - des publicités affichés via des serveurs en cascades qui ne sont pas toujours sécurisés.

L'impact (très négatif) sur les revenus adsense

Côté Google Adsense, un choix drastique a été fait pour éviter ce problème : si la page affichant les adsense utilise le protocole http, la page a accès à l'inventaire intégral des publicités. Si la page est sécurisée, seules les publicités appelées via un protocole sécurisé sont sélectionnées. Résultat : basculer un site de http:// en https:// fait aujourd'hui chuter de manière très sensible les revenus adsense d'un site… Ce point est clairement expliqué dans l'aide de Google Adsense :
"Si vous décidez de convertir votre site HTTP en HTTPS, sachez que les annonces diffusées sur vos pages HTTPS risquent de générer des revenus moins élevés que celles diffusées sur vos pages HTTP. Cela s'explique par une pression plus faible au niveau des enchères, car les annonces non conformes à SSL ne participent plus à la mise en concurrence."

L'importance du bon choix du type de clé et du bon fournisseur de certificat

Mais le principal reproche adressé à SSL / TLS est le coût des certificats. En matière d'offres de certificats et de prix, c'est franchement la jungle et le novice est souvent

déconcerté de voir que les prix varient de quelques euros (voire la gratuité…) à des centaines de milliers de dollars, sans que la différence de service rendu ne saute aux yeux.

Dans un premier temps, techniquement, il existe plusieurs types de certificats, et il est important de choisir le bon en fonction des "hosts" que l'on souhaite sécuriser :
- **les certificats "simples"** (single domain) : ils ne sécurisent qu'un seul host (donc www.votresite.com et pas une adresse de type sousdomaine.votresite.com ;
- **les certificats "multidomaines"** : beaucoup plus chers, ils permettent de sécuriser plusieurs domaines avec le même certificat. Le coût est le plus souvent proportionnel au nombre de domaines à sécuriser ;
- **les certificats "wildcards"** : ils couvrent tous les hosts appartenant à un domaine donné (donc www.votresite.com, ssdomaine1.votresite.com, ssdomaine2.votresite.com…) Plus chers que les certificats "single domain", à l'usage ils sont plus économiques dès lors que l'on utilise plusieurs sous-domaines.

La longueur de la clé est également un critère important. Google recommande des clés de 2048 bits de longueur : c'est aujourd'hui le standard minimum pour les fournisseurs sérieux (plus la clé est longue, plus « casser » la clef demande du temps et des ressources machines). Il est possible de commander des clés plus longues mais il vous faudra remettre la main à la poche.

Quel fournisseur de certificat choisir ?

Les sociétés qui délivrent des certificats ont des pratiques extrêmement différentes. Certaines font preuve d'une rigueur extrême dans le processus de délivrance du certificat, d'autres beaucoup moins. Certaines s'engagent sur le niveau de sécurité, au point de vous indemniser si votre certificat est compromis. C'est ce qui explique le prix parfois exorbitant de certaines offres de certificat : il inclut une sorte "d'assurance dommages" pour le cas où le fournisseur de certificats serait pris en faute. Ce n'est pas une hypothèse d'école : l'une de ces sociétés les plus en vues a fait l'objet il y a quelques mois d'un piratage !

Nous vous conseillons de choisir un fournisseur connu, et d'éviter comme la peste les fournisseurs de certificats gratuits. En dessous de 100 dollars annuels, le sérieux du fournisseur de certificat peut-être mis en doute, et doit être soigneusement vérifié. Comptez 500 dollars minimum pour un certificat wildcard doté de sérieuses garanties et délivré par un fournisseur sérieux. Mais on peut prédire que l'annonce de Google va chambouler le marché et que le paysage concurrentiel et le prix des certificats vont bouger dans les mois qui viennent.

Néanmoins, il faut comprendre que ces prix sont en partie incompressibles, car ils rémunèrent un travail complexe et parfois impossible à automatiser : l'authentification du demandeur de certificat (en tant que personne) et l'authentification de l'entité destinataire du certificat. Les échanges administratifs et la vérification parfois manuelle des documents demandent du temps et du travail aux fournisseurs. Pour diminuer les coûts, il faut renoncer à des vérifications, c'est pourquoi certains fournisseurs délivrent des certificats… en chocolat, qui en réalité n'authentifient pas grand-chose.

Face à de telles différences entre les offres du marché, quelle va être la position de Google ? Pour l'instant, le discours officiel est neutre : la société de Mountain View n'évoque pas l'idée de faire une différence entre les certificats délivrés par des autorités reconnues, et ceux à prix cassés. Mais comme il est facile d'identifier le fournisseur d'un certificat, gageons que si nécessaire, Google pourrait à terme donner des coups de pouce différents pour les certificats "sérieux" et les autres. Evidemment, cela instaurerait une sélection par l'argent entre différents types d'éditeurs de sites… Mais c'est peut-être la volonté ultime de Google : renchérir de manière sensible le coût des actions de webspam. Rappelons que les certificats sont délivrés pour une durée limitée, qu'ils expirent, et que les coûts évoqués sont … annuels.

Comment basculer un site en https:// ?

Si votre site n'est pas encore sécurisé, il convient donc de se préparer à le faire. Mais attention, ce changement n'est pas anodin. Il est sans risques si la migration est bien maîtrisée, mais comme toute migration, ce changement demande beaucoup de préparation et une rigueur de tous les instants.

- Il faut bien sûr "installer" le certificat sur le ou les serveurs qui hébergent vos sites web. La tâche est relativement simple mais demande un bon niveau de technicité. Puis activer le support du SSL/TLS sur vos serveurs web.
Attention : à ce stade beaucoup de webmasters ont des déconvenues, parce qu'ils ont mal enregistré le "host" dans la phase de demande de certificat, et celui-ci peut se révéler inopérant sur le domaine ou le sous-domaine choisi ! Il faut être très rigoureux quand vous remplissez les formulaires du fournisseur.

- Il faut aussi s'assurer que votre site peut afficher des URL en https:// au lieu de http:. Il faut donc faire la chasse aux URL absolues "en dur" mentionnant "http://" et si possible les remplacer par des URL relatives sans mention du protocole.

- Toutes les ressources d'une page doivent être appelées via https:. C'est souvent là que réside le travail le plus compliqué à réaliser. Tous les fichiers images, css, js, json, xml doivent être appelés via https:// et le code doit être changé en ce sens.

- Il faut ensuite créer un plan de redirection 301 des urls en http:// vers https://. Cela permettra d'éviter un doublon entre les urls http et https. On peut doubler la sécurité en "canonicalisant" les pages en http: vers les pages en https:.

- Notez bien que l'on ne peut pas, à la date d'aujourd'hui, "migrer" les URL en http:// vers https:// à l'aide de l'outil de migration des webmaster tools. Ce n'est pas prévu.

- Pensez aussi à mettre à jour votre fichier robots.txt, votre compte Bing Webmaster Tools et Google Webmaster Tools pour suivre les URL en https, ainsi que les paramétrages de vos CDN si vous en utilisez.

- Les instructions de mise en cache dans les headers doivent être changées systématiquement pour éviter les problèmes de performance évoqués plus haut.

Attention aux effets de bord

Ce type de migration n'est jamais neutre : tous les outils de tracking sont potentiellement impactés par un tel changement, et vous serez obligés de les reparamétrer et/ou d'agir pour maintenir une série statistique intègre. Les effets de bord sont inévitables : remise à zéro de certains compteurs (chez Facebook par exemple), ou de certains indicateurs pour Google Adwords.

Pour les sites utilisant des technologies anciennes voire obsolètes, il peut s'avérer difficile de réaliser ce type de migration sans faire des changements importants (comme passer à une nouvelle version d'un CMS, renoncer à un plugin, refondre le code de zones entières du site).

Quel sont les risques liés à la bascule entre les URL en http:// et https:// ?

En dehors des points évoqués plus hauts, il n'y a pas de risques particuliers lié à ce genre de changement. Mais c'est une migration, et comme pour toutes les migrations, la moindre erreur peut avoir des conséquences importantes. On suivra donc les mêmes consignes de précautions que pour un changement de plafeforme logicielle accompagné

d'un changement de serveurs : tout doit être planifié, exécuté dans l'ordre avec rigueur, testé avant et après mise en production.

Dans quels délais le "bonus https://" deviendra-t-il significatif ?

Compte tenu de toutes les contraintes évoquées plus haut, il est douteux que l'on assiste à une adoption rapide et massive du protocole SSL/TLS par les sites. Pour beaucoup, il faudra des mois avant qu'une décision de bascule prise maintenant prenne effet. Les équipes de Google en sont probablement conscientes, et vont surveiller ce qui se passe avant de "pousser le curseur" pour accorder un bonus sensible aux pages en https://. Mais il est également probable que Google doive accorder un réel avantage aux pages sécurisées pour que l'effet incitatif apparaisse vraiment. Ce qui signifierait que Google peut très bien ne pas attendre que la majorité des sites aient basculé avant de changer son algorithme. Dans ce cas, un changement pourrait intervenir rapidement.

Google vous pousse à le faire : préparez-vous.

En conclusion, nous ne pouvons que vous recommander d'envisager dès maintenant la mise en place de ce protocole pour l'ensemble des pages de votre site web. Google accompagne avec ce geste une tendance lourde du web. Aujourd'hui, le "bonus" accordé par Google est symbolique, et il ne serait pas raisonnable de faire ce changement à marche forcée. Certains sites l'ont fait depuis le 6 août, et le moins que l'on puisse dire c'est qu'un certain nombre d'entre eux s'en mordent les doigts, faute d'avoir minutieusement préparé la migration.

Vous disposez probablement d'un délai de plusieurs mois pour opérer ce changement. Il est donc raisonnable de l'inclure dans votre roadmap de début d'année 2015, voire du premier semestre 2015. Mais il est urgent de commencer à planifier les travaux de préparation nécessaire !

Notez-bien que la logique voudrait que l'adoption du protocole https:// pour vos pages vous aide seulement à maintenir vos positions actuelles. Elle ne vous fera pas forcément gagner des positions dans les classements. Si tous vos concurrents ont basculé, tout le monde aura droit au même bonus !

Notez que le ratio coût/bénéfices de ce changement demandera à être évalué : n'oubliez pas, en particulier si vous percevez des revenus du programme Adsense, que le gain en trafic espéré peut être accompagné d'une perte notable de revenus. Dans tous les cas, ce sera une bonne idée de prévoir une phase d'évaluation et un scénario permettant un "roll back" (retour au protocole http) si les effets de bord et les pertes constatées sont trop importants.

Bibliographie

Le post officiel de Google sur le blog Webmaster Central : "HTTPS as a ranking signal"
http://googlewebmastercentral.blogspot.fr/2014/08/https-as-ranking-signal.html

Le même billet sur le blog en français :
http://googlewebmastercentral-fr.blogspot.fr/2014/08/le-protocole-https-en-tant-que-facteur.html

Le lien vers la conférence Google I/O au cours de laquelle la volonté de promouvoir le protocole SSL/TSL et l'emploi du https:// comme signal ont été évoqués pour la première fois :
https://www.google.com/events/io?utm_source=wmx_blog&utm_medium=referral&utm_campaign=tls_fr_post

Les bonnes pratiques fournies par Google :
https://support.google.com/webmasters/answer/6073543?utm_source=wmx_blog&utm_medium=referral&utm_campaign=tls_fr_post&hl=fr

L'article du blog de Searchmetrics résumant les conclusions de l'étude de Marcus Töber sur l'impact du signal https:// :
http://blog.searchmetrics.com/us/2014/08/29/https-vs-http-analysis-do-secure-sites-get-higher-rankings/

Outil de test de configuration SSL : Qualys :
https://www.ssllabs.com/ssltest/

Philippe YONNET , ***Directeur Général de l'agence Search-Foresight, groupe My Media (http://www.search-foresight.com).***