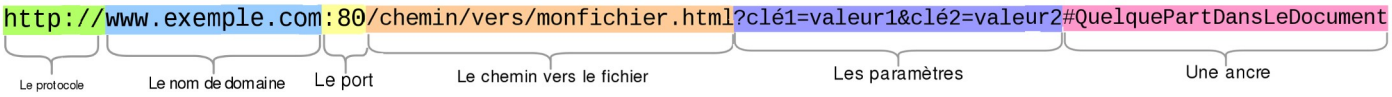
URL, HTTP(S)

manuel utilisé et mentionné dans cette fiche →

URL

Une URL (Uniform Resource Locator, littéralement *localisateur uniforme de ressource*) est une chaîne de caractères combinant les informations nécessaires pour indiquer à un logiciel comment accéder à une ressource Internet. Ces informations peuvent notamment comprendre le protocole de communication, un nom d'utilisateur, un mot de passe, une adresse IP ou un nom de domaine, un numéro de port TCP/IP, un chemin d'accès, une requête.



Le protocole

http:// correspond au protocole. Ce fragment indique au navigateur le protocole qui doit être utilisé pour récupérer le contenu. Généralement, ce protocole sera HTTP ou sa version sécurisée : HTTPS.
On peut aussi rencontrer d'autres protocoles comme mailto:// (qui permet d'ouvrir un client de messagerie électronique) ou ftp:// qui permet de transférer des fichiers.

file **t**ransfer **p**rotocol

Le nom de domaine

www.exemple.com correspond au nom de domaine. Il indique le serveur web auquel le navigateur s'adresse pour échanger le contenu. À la place du nom de domaine, on peut utiliser une adresse IP.

Le port

:80 correspond au port utilisé sur le serveur web. Il indique la « porte » technique à utiliser pour accéder aux ressources du serveur. Généralement, ce fragment est absent, car le navigateur utilise les ports standards associés aux protocoles (80 pour HTTP, 443 pour HTTPS). Si le port utilisé par le serveur n'est pas celui par défaut, il faudra l'indiquer.

Le chemin vers le fichier

/chemin/vers/monfichier.html est le chemin, sur le serveur web, vers la ressource.

Les paramètres

?clé1=valeur1&clé2=valeur2 sont des paramètres supplémentaires fournis au serveur web. Ces paramètres sont construits sous la forme d'une liste de paires de clé/valeur dont chaque élément est séparé par une esperluette (&). Le serveur web pourra utiliser ces paramètres pour effectuer des actions supplémentaires avant d'envoyer la ressource. Chaque serveur web possède ses propres règles quant aux paramètres. Afin de les connaître, le mieux est de demander au propriétaire du serveur.

Une ancre

#QuelquePartDansLeDocument correspond à une ancre, celle-ci désigne un endroit donné de la ressource. Une ancre représente, en quelque sorte, un marque-page à l'intérieur de la ressource. Ajouter une ancre à une URL permet au navigateur d'afficher la ressource à l'endroit de ce marque-page. Pour un document HTML, par exemple, le navigateur défilera la page jusqu'au niveau de l'ancre. Pour un document audio ou vidéo, le navigateur ira se placer à l'instant représenté par l'ancre.

Source : https://developer.mozilla.org/fr/docs/Apprendre/Comprendre_les_URL

Exemples d'URL : https://www.mathemathieu.fr/classes/terminaless/ts-dm#conj-collatz
ftp://ftp.rfc-editor.org/in-notes/rfc2396.txt
mailto://mathemathieu@free.fr

fichier ordinateur → file:///C:/Users/Public/Desktop/Microsoft%20Word.lnk

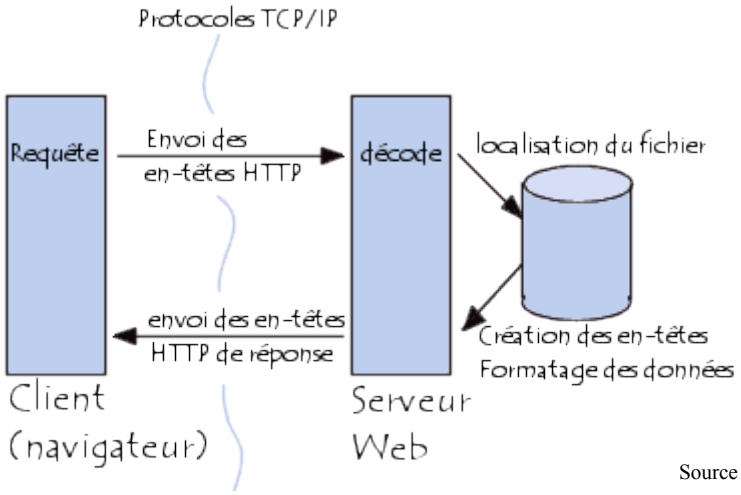
PROTOCOLE HTTP

HTTP (*Hypertext Transfer Protocol*) est un protocole de communication client-serveur développé pour le World Wide Web.

Les **clients HTTP** les plus connus sont les navigateurs web (Firefox, Chrome, Safari…) permettant à un utilisateur d'**accéder à un serveur** contenant les données. Il existe aussi des systèmes pour récupérer automatiquement le contenu d'un site tel que les « aspirateurs de site web » ou les « robots d'indexation ».

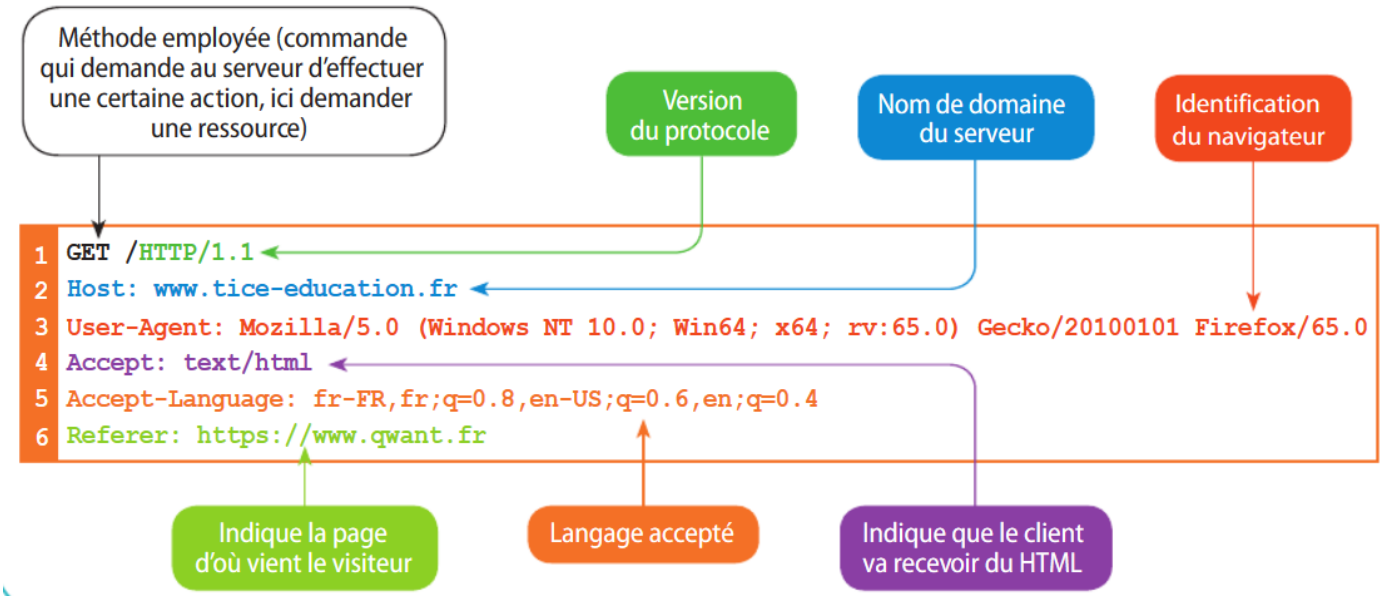
Une **requête HTTP** (*Hypertext Transfer Protocol*) est donc une demande effectuée par un navigateur web à un serveur HTTP.

La communication entre le navigateur et le serveur se fait en deux temps :

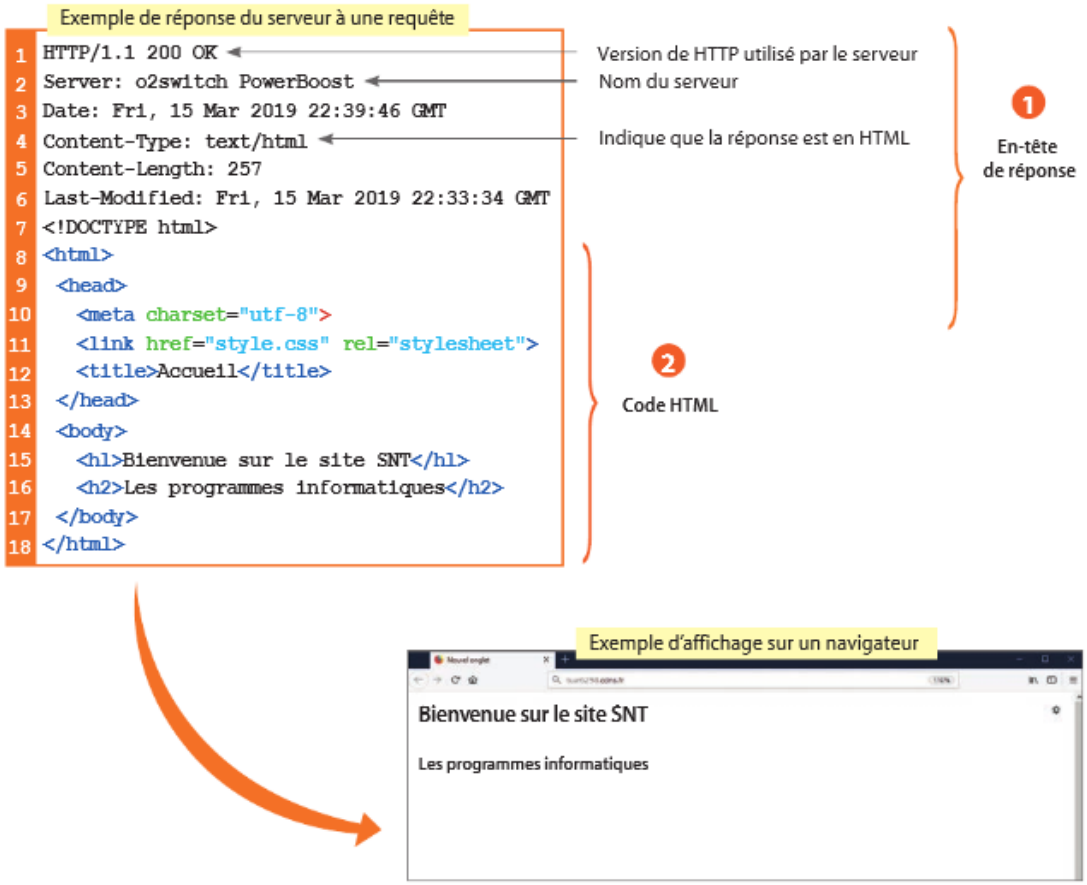


Source : https://www.commentcamarche.net

Un exemple de requête HTTP :



Source : SNT 2nde 2019, éd. Delagrave, ISBN 978-2-206-10338-9



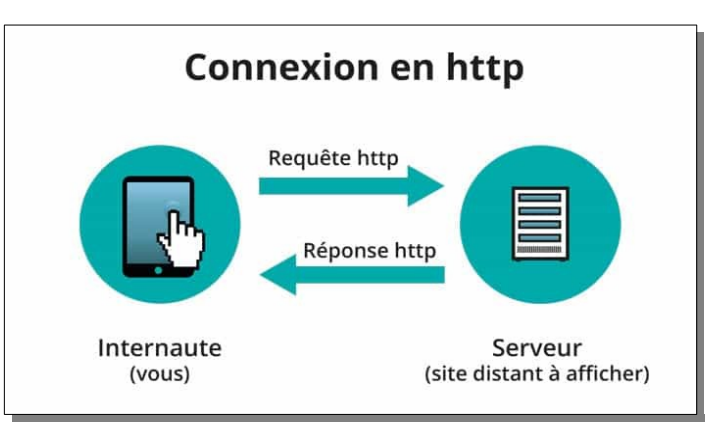
Source : SNT 2nde 2019, éd. Delagrave, ISBN 978-2-206-10338-9

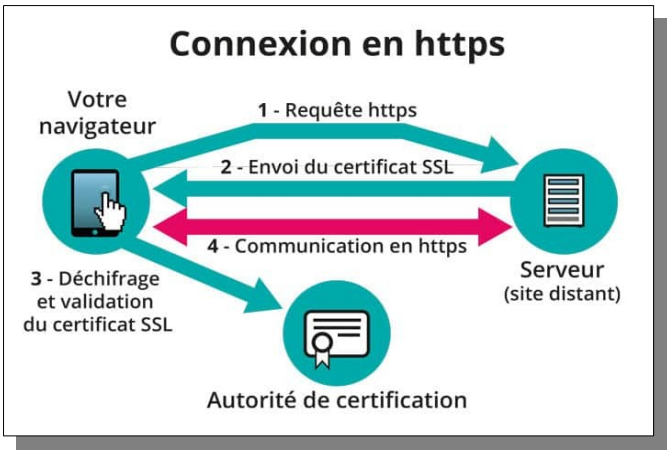
Bien sûr, il existe d'autres méthodes que GET. En voici quelques-unes (ce n'est pas à connaître) :

GET	demande une ressource
HEAD	ne demande que des informations sur la ressource, sans demander la ressource elle-même
POST	pour transmettre des données en vue d'un traitement à une ressource (le plus souvent depuis un formulaire HTML)
TRACE	demande au serveur de retourner ce qu'il a reçu, dans le but de tester et effectuer un diagnostic sur la connexion

PROTOCOLE HTTPS

HTTP**S** est une extension du protocole HTTP, sur lequel on rajoute un autre protocole de sécurité afin de chiffrer les informations entre le client et le serveur.
Ce protocole de sécurité est, depuis 2014, le **protocole TLS** (*Transport Layer Security* = sécurité de la couche de transport), successeur du **protocole SSL** (*Secure Sockets Layer*).





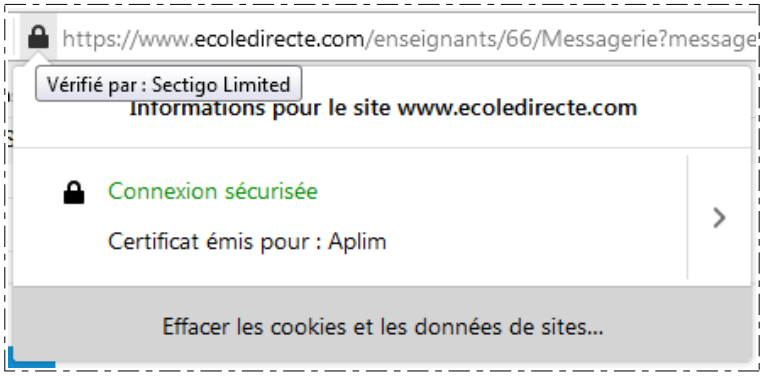
Source : https://clecomweb.fr

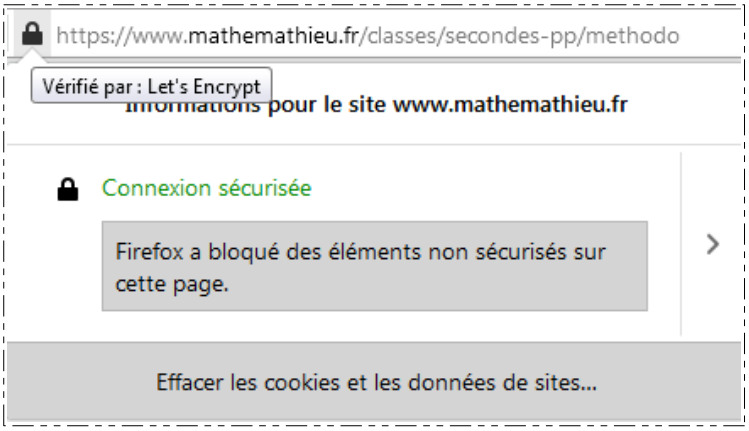
Sur Firefox, le certificat SSL est visible par l'apparition du petit cadenas dans la barre d'adresse.

site « sécurisé » (HTTPS) → 

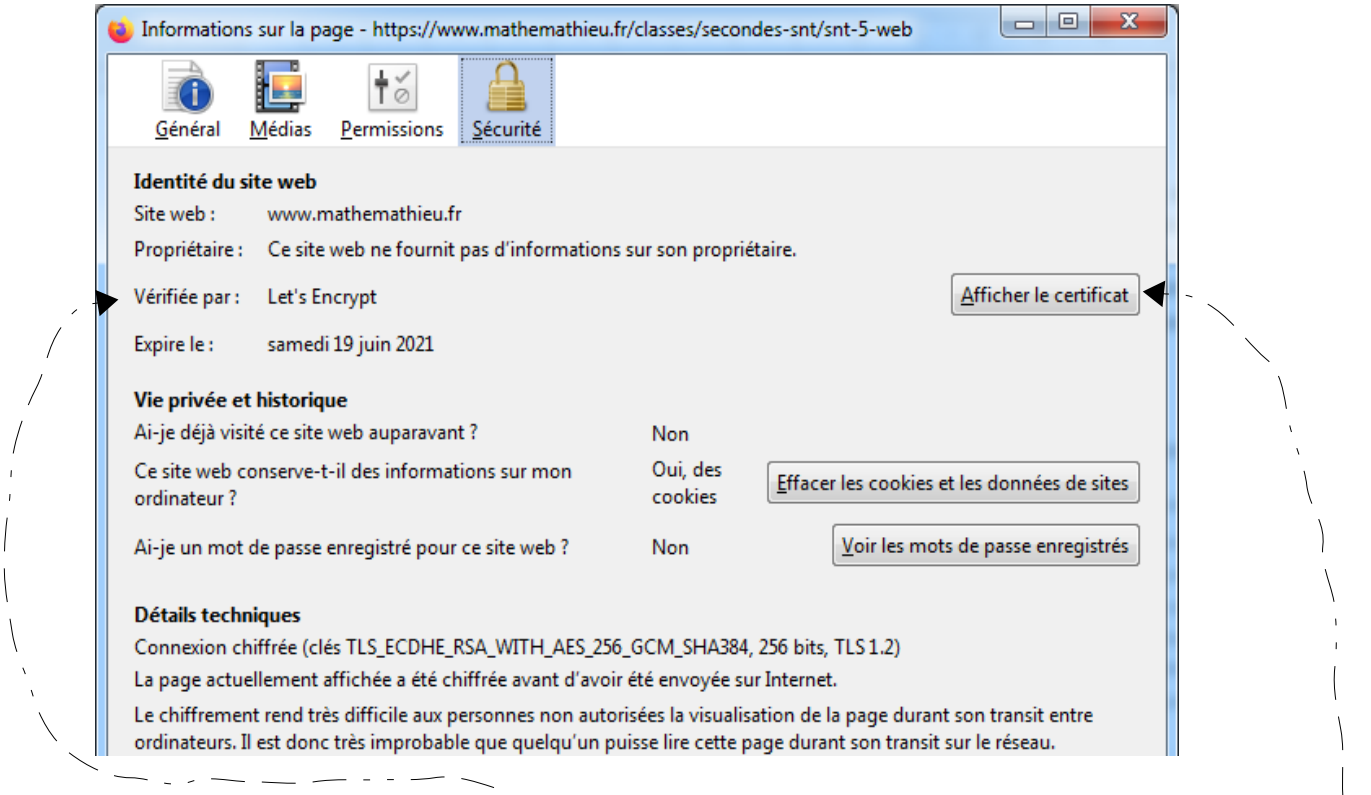
site non « sécurisé » → 

Si vous cliquez sur le cadenas, vous pouvez avoir plus d'informations :





Voici les détails obtenus avec le navigateur Firefox pour le site www.mathemathieu.fr le 4 mai 2021 :



On observe que le certificat est « vérifié par **Let's Encrypt** », une autorité de certification lancée en décembre 2015 qui **fournit des certificats gratuits**. Une **autorité de certification** (*Certificate Authority* = **CA**) est un **tiers de confiance permettant d'authentifier l'identité des correspondants**.

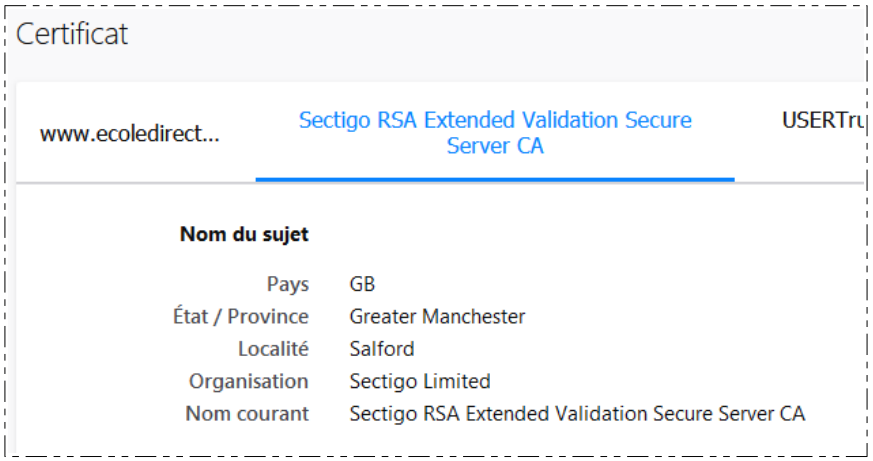
Ces tiers de confiance sont importants : une personne malveillante pourrait tout à fait créer un site avec le protocole HTTPS, créer lui-même son propre certificat afin de rassurer le client, et lui voler ses informations personnelles… En cliquant sur « Afficher le certificat », vous pouvez voir plus d'informations et vérifier si le certificat a un numéro de série, est déclaré par une organisation ou pas, etc.

Attention : même approuvés par une CA, les certificats SSL ne sont pas 100 % fiables puisque les CA ne sont pas tenues d'utiliser un processus spécifique pour authentifier les entités leur demandant de signer leurs certificats. Donc **même une entité malveillante peut obtenir un certificat SSL par une CA** !

En réalité, il existe différentes certifications SSL/TLS, plus ou moins fiables :

- **gratuit non approuvé** : gratuit, ce certificat est auto-signé par le site web et non-pré installé sur les navigateurs web. Il implique un degré de confiance réduit, raison pour laquelle il n'est pas recommandé, bien que la CA Let's Encrypt qui délivre des certificats SSL/TLS gratuits semble une solution fiable et reconnue par la plupart des navigateurs sérieux
- **approuvés** : signés par une CA et pré-installés sur les navigateurs web
- les ***Extended Validation*** (**EV**) : standard du e-commerce, le **certificat EV SSL/TLS** « nécessite une enquête plus approfondie de l'entité requérante par l'autorité de certification avant d'être délivré ». Ces certificats de validation étendus offrent les niveaux les plus élevés de confiance et d'authentification d'un site web. Ils ont été conçus pour renforcer la sécurité du commerce électronique et lutter contre les attaques d'hameçonnage (phishing).

Exemple d'un certificat fiable (pour ecoledirecte.com), avec une EV :



Remarque : dans le cas où un hacker aurait réussi à prendre le contrôle de votre machine (droits administrateur via un malware), il pourrait de toute façon se faire passer pour une autorité de certification dans votre navigateur, et vous faire croire, par exemple, que vous êtes sur (le vrai) amazon.fr, alors que l'URL pointe vers une adresse IP d'un serveur contenant un faux site… Il vous volera alors vos données : c'est du *phishing* très élaboré ! Si cela vous intéresse, vous pouvez allez regarder cette vidéo de Micode (≈ 5 min) qui résume aussi beaucoup de choses déjà vues en classe (DNS, etc.).

→

LE CONTENU MIXTE ET LES RISQUES ENCOURUS

Sur Firefox, il peut apparaître deux autres cadenas, car Firefox vous protège des attaques en bloquant le contenu potentiellement malveillant et non sécurisé des pages web censées être sûres.

Rappel : HTTP est un protocole de transmission des informations d'un serveur web vers votre navigateur.

Quand vous visitez une page entièrement transmise en utilisant HTTPS, comme sur le site de votre banque, vous voyez une icône de cadenas dans la barre d'adresse : cela signifie que votre connexion est authentifiée et chiffrée, et par conséquent protégée des écoutes.

Cependant, si la page HTTPS inclut des données HTTP, la portion HTTP peut être lue ou modifiée par des pirates, même si la page principale est servie avec HTTPS. Quand une page HTTPS a une partie de son contenu en HTTP, on appelle cela du contenu « mixte » : la page n'est qu'en partie chiffrée et, même si elle semble être sécurisée, elle ne l'est pas.

Quels sont les risques des contenus mixtes ? Un pirate peut remplacer les données HTTP de la page visitée afin de voler vos identifiants, s'emparer de votre compte, obtenir des données sensibles vous concernant ou tenter d'installer des logiciels malveillants sur votre ordinateur.

Vérifiez si l'icône représentant un cadenas est présente dans votre barre d'adresse pour déterminer si la page contient du contenu mixte :

Pas de contenu mixte : sûr

- Vous verrez un cadenas gris lorsque vous êtes sur une page complètement sécurisée (HTTPS). Pour savoir si Firefox a bloqué des parties de la page qui ne sont pas sûres, cliquez sur l'icône de cadenas gris.

Contenu mixte non bloqué : pas sûr

- Si vous voyez un cadenas barré de rouge, la page présente du contenu mixte actif et Firefox ne bloque pas ses éléments non sûrs. Cette page est vulnérable aux interceptions et aux attaques dans lesquelles vos données personnelles peuvent vous être volées à partir du site. Vous ne devriez pas voir cette icône sur une page sécurisée (HTTPS), sauf si vous avez débloqué le contenu mixte en suivant les instructions de la section suivante. **Note :** un cadenas barré de rouge est aussi affiché sur les sites web non chiffrés (HTTP ou FTP).

- Un cadenas gris avec un triangle orange ou jaune indique que Firefox ne bloque pas le contenu passif non sécurisé, comme des images. Par défaut, Firefox ne bloque pas le contenu mixte passif, vous verrez simplement un avertissement prévenant que la page n'est pas totalement sûre. Les pirates sont susceptibles de manipuler des parties de la page, comme afficher du contenu falsifié ou inapproprié, mais ils ne devraient pas être en mesure de voler vos données personnelles à partir de ce site.