HTTPS Migration For Enterprises

Services, Approach & FAQ's For Large Sites & Multi-domain Estates

Httpschecker.net

Helping thousands of users migrate worldwide

Authored and supported by Wildfire Internet

www.wildfireinternet.co.uk

info@wildfireinternet.co.uk

0845 643 1290 (UK)
+44 1905 770 277 (International)

Introduction

This document is to help larger organisations embarking on their HTTPS migration program:

- For very large sites (i.e. millions of pages)
- Across a significant number of domains and sub-domains (typically more than 50)
- Where content is dynamically generated on pages (which crawlers won't detect)
- Assigning work to multiple developers or content owners
- Identifying new errors creeping in afterwards
- Checking SSL certificates

Capability

Our tools have been used to scan over 25,000 websites across the globe for mixed content issues caused when moving to HTTPS. One of the largest sites it was used on had 8,000,000 pages. Users include Sony, Shutterstock, Symantec, PWC, Zendesk and AVG amongst thousands of other users.

Accompanying Guides

The following guides provide useful information on HTTPS which are worth reading:

Why HTTPS Matters

How To Fix Mixed Content

Buying Licences

HTTPS Checker can crawl up to 250,000 pages to look for issues on a standard Enterprise licence. For large sites over 250,000 pages then we recommend the Platinum licence for unlimited pages at $249/month (price correct at time of going to press, may

vary). HTTPS Reporter can monitor up to 50 sites per $99/month subscription, so if you had 135 sites for example then you would purchase 3 x $99/month subscriptions.

Overview

Mixed content issues occur on HTTPS sites due to links within the content using HTTP rather than HTTPS in the asset's URL. This breaks the HTTPS security padlock on the page that is affected. Most content on sites can be picked up during a crawl, but this can miss some dynamically generated content because the crawl cannot discover it.

In order to find both types of content, we have two applications to use:

- **HTTPS Checker - Run This proactive scan first**
  Proactively crawl the website to find issues on pages, either prior to a migration (to find where all links reside) or once you have added your SSL certificate. This does not pick up links embedded in dynamically generated content. Check your log files after use.

- **HTTPS Reporter - Capture miscellaneous issues in real-time**
  Identify all issues including those in dynamically-generated content in real-time. Uses CSP violation catching, which is then recorded in the app. This typically happens on a live site and is therefore reactive. However, the number of issues should be relatively small as you will already have crawled the site with HTTPS Checker. This also catches new issues that creep in when content is changed, added or data is imported.

How Does HTTPS Checker Work & What Does It Look For?

HTTPS Checker is a downloadable desktop app which identifies pages to crawl via links on your site, then identifies and reports on mixed content issues that are found. These

issues are stored in local storage on your machine and summarised at the end of the crawl on-screen plus you can download the issues in CSV or PDF formats. The following types of issues are detected:

- Active mixed content
- Passive mixed content
- Insecure form submissions
- Insecure pages asking for passwords or credit card details (Chrome "Not Secure" Warning)
- Insecure redirects within the site
- Canonical links to non-HTTPS pages
- Inconsistent linking from HTTPS to HTTP URLs on your site
- Social links to non-HTTPS pages
- Sitemap contains HTTP links
- Check for the presence of modern security headers
- SSL Certificate is valid
- SSL Certificate has not expired

Installation and use of HTTPS Checker

1. Download the free app (crawls up to 500 pages)
2. Double click / Install the downloaded file
3. Run the app to ensure it works on your computer
4. Request an Enterprise evaluation licence to unlock all features for 7 days
5. Fix issues then re-run scan

Log files are recorded to help you with troubleshooting- see section “Crawl & Error Logs (including Referrer URL’s To Track The Path)”.

Mode Control

It is possible to select a mode for scanning when you enter your website URL:

All Checks (normal)	Run all checks
All Except Hyperlink Checks	All checks except hyperlinks
Mixed Content Checks Only	Only find mixed content issues
Chrome Not Secure Warning Only	Find HTTP pages that collect passwords or credit cards which Google Chrome marks as non-secure

How Does HTTPS Reporter Work & What Does It Record?

HTTPS Reporter captures Content Security Policy violations from your site which are triggered when a page is visited that has a mixed content issue. You need to generate and add a CSP to your website header. You then log in to your HTTPS Reporter dashboard to see these violations and mark them as fixed.

Installation and use of HTTPS Reporter

1. Subscribe to one of the paid plans.
2. Generate the Content Security Policy header (examples are given in the HTTPS Reporter dashboard).
3. Add the CSP header to your website.
4. HTTPS Reporter will check the site to see if it has appeared.
5. If your site is not publicly visible or a dev site, ask us to verify the site for you
6. When it has been verified, CSP violations will appear in HTTPS Reporter & you'll receive daily emails of these.

7. Manage up to 1,000 violations in the dashboard, marking fixed issues as resolved. A violation is triggered whenever a user visits a page with an issue present. For example, 1,000 visitors to a page with 1 issue created 1,000 violation reports.
8. Repeat these steps for each website to be tracked.

Approach To Migrating Large Sites

Planning Tips For Large Sites Or Estates

Complexity - Large sites often have lots of technology to build pages or show ads, you need to identify and handle this array of technology plus front and back-end integrations. If you don't have an asset record of what's been implemented on your site, start with a search on Builtwith which will show your technology profile and then make sure each technology provider has an equivalent HTTPS service.

Timescales - It is going to take a lot of time and effort to migrate a site or multiple sites, so plan your resourcing and treat it as a high-level project. Start with your Dev site.

Bite-size Chunks - Handle the migration in sections or audiences rather than trying to do the whole thing in one go.

Learn As You Go - Measure and record the issues as you complete each section, then apply what you learn to the next section. HTTPS Checker allows you to download a PDF or CSV of crawl results and HTTPS Reporter stores and displays issues in your dashboard.

Tackling Widespread Common Issues First

It is a good idea to first look at common generated pages such as category or product pages, to see if there is a program issue likely to affect all of them and to resolve it first

and then progress to the main crawl. On a common page, consider using the browser's developer tool to spot the obvious issue and fix it first.

Little & Often Approach By Single Developer

HTTPS Checker is a downloadable desktop app which crawls pages to identify and report on issues found. These issues are stored in local storage on your machine and summarised at the end of the crawl on-screen plus you can download the issues in CSV or PDF formats. This works well when you have a single developer working on it.

Crawling By Section Or URL Lists For Many Developers

You may wish to split the crawl down into sections that individual developers work on. There are a few ways to achieve this:

1. Set the start point of your crawl with the section you wish to cover e.g. www.mydomain.com/menswear and allow the crawl to progress from there.
2. Upload a list of URL's to crawl so you can exactly target which are included.
3. Set additional robots.txt rules to keep the crawl within specified directories of the site.

Once the scan is complete you can then download the CSV's of issues and pass them to your developers to resolve in parallel.

HACK: To get a list of crawled URL's on your site without having thousands of potential errors, run HTTPS Checker in **"Chrome Not Secure Warning Only"** Mode, as there are only likely to be a small number of pages with this type of issue .

Crawling Dev or UAT Sites

HTTPS Checker

HTTPS Checker runs off your desktop so it can crawl sites you access which are not necessarily available on the web, such as development sites. As standard, it follows your robots.txt rules, so if there is an issue getting your crawl started because of this then go into Advanced Options > General and untick “Use Robots.txt File Rules “. In Advanced Options you can also set up a proxy to run your requests through, if required.

HTTPS Reporter

HTTPS Reporter attempts to verify a site to make it active, which means the site should be publicly available. However, we can manually override the verify for Dev and UAT sites upon request via our contact page on httpsreporter.net

Speeding Up Your Crawl Using Advanced Options

The main way to speed up the crawl is via the settings in the app and running it on a good spec machine (as it will not run multi-threaded). You can adjust these in Advanced Options > Limits.

Queue Interval	Faster crawl = shorter crawl time
Queue Concurrency	More concurrency = shorter crawl time
Crawl Timeout	Wait less time for a page response before abandoning it
Crawl Bailout:	The time that the crawler will wait during crawl inactivity, before stopping.

Remember: the faster the crawl, the more load it places on the server and this can itself cause timeout issues so exercise caution and check your error log files for any pages

which have not been crawled (other issues may also be the cause so you need to check your error log file anyway to ensure that all pages end up being crawled).

Crawl & Error Logs (including Referrer URL's To Track The Path)

HTTPS Checker saves log files onto the machine that is performing the scan. The crawl log shows the URL's crawled and, the immediate referrer URL so you can track the path to the issue. The error log shows URL's which couldn't be crawled with a reason recorded. You can then fix the underlying issues affecting these pages and then scan them again with HTTPS Checker (hint: upload them as a list of URL's to crawl to save time).

Location of Log Files - These can be accessed in the app via the **Crawl Logs** button

Linux	~/.config/httpschecker/
OS X	~/Library/Application Support/httpschecker/Logs
Windows	%USERPROFILE%\AppData\Roaming\httpschecker\

Other Advanced Options

General	Control whether to crawl HTTP links, set your User Agent to appear as a different browser, ignore robots.txt settings and enter your own rules.
Limits	Set crawl parameters and place a cap on issues found and pages crawled
Proxy	Run the requests through a proxy
Session	Run the app on your private pages or admin centre by entering your logged-in session cookie.

Monitoring & Fixing Real-time Issues in HTTPS Reporter

Once you have set up your CSP headers correctly, you will start to receive real-time CSP violation reports into your HTTPS Reporter admin console.

Handling Reports Generated About The Same Issue

A report is generated for each asset on a page that fails to comply with your CSP policy, which means if 10,000 people visited a page with a single issue then 10,000 identical reports would be generated.

HTTPS Reporter recognises they are all the same problem and therefore only records it as one issue. Dependent upon your subscription level, you can view up to 1,000 *latest* unique issues that have not been marked as resolved.

NB See the **Terms & Reasonable** Use section regarding our fair usage policy.

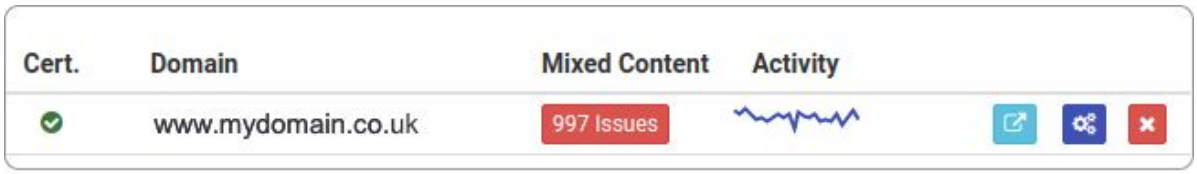
Marking CSP Violation Reports As Fixed

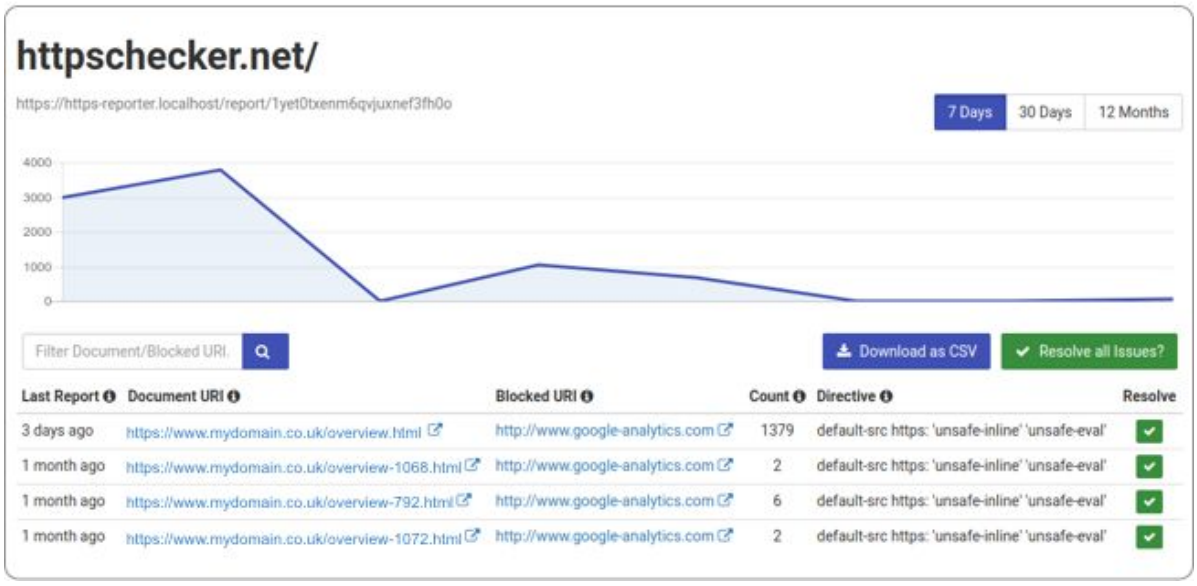
Once you have fixed a CSP violation issue, in Admin Console simply press the Resolve button on the report line and it will be removed, but still counted within the graph history.

Using In-Built Graphs To See Spikes In Activity

Use the graphs in admin console to quickly see which domains have the most activity at the domain level measured over 24 hours, and activity within individual domains over either a 7, 30 or 360 day period which you can select, as well as giving a list of up to 1,000 of the latest issues so you can prioritise what to fix first.

NB Graph data is collected and shows *after 2 hours,* if there are sufficient issues reported.





Blocking Irrelevant URI's

Over time you may see URI's appearing in your console which don't relate to your site or scripts you have placed on your pages. This is because a CSP violation is generated from each user's browser which can mean you receive reports due to other software installed with the browser, such as plug-ins, which aren't relevant to your website and don't stop the padlock being displayed. You can block these irrelevant URI's using the "Ignore Rules" function in the Admin Console.

SSL Certificate Validation & Expiry Notifications

HTTPS Reporter performs checks on the domain's SSL certificate to ensure it is valid and reports this against the domain line in the **Cert** column. Expiry notifications will be emailed out daily when the certificate is within 14 days of its expiry date.

Support Services & Team Sharing For Enterprises

Enterprise clients often have more complex requirements and require greater assistance with their migrations and understanding how our tools help. We are pleased to offer support to enterprises as follows:

Pre-sales Consultation	Free 30 minute telephone call to discuss your migration and how the app works. Made by appointment, UK hours.
Evaluation Trial	Free 7 day evaluation Enterprise licence. Download the free version of the tool off the HTTPSChecker.net site and then request the evaluation credentials.
Ad-hoc Advice	Remote consultancy is available at the rate of £125+tax / hour which can include: • Overview and advice • Determining best approaches to migrate using the tools • Participating in telephone conference calls

General Usage & Guides

Information is provided on the httpschecker.net website which contains guides, installation, use and FAQ's.

System Issues & Availability

System issues can be reported to us via the Contact page or live chat on the httpschecker.net website and we will endeavour to fix and release them as soon as possible.

HTTPS Reporter issues are reported on our Twitter status feeds:
https://twitter.com/HTTPSReporterUp

Hours Of Cover

Support and consultancy is provided in UK office hours on a reasonable endeavours basis.

Team Sharing

Enterprise users can invite other team members to login to HTTPS Reporter (via one of the available social logins) to read or resolve issues or administrate other users.

Software Enhancement Requests, Releases & IPR

From time-to-time we release enhancements, which are automatically installed on the HTTPS Checker app when you restart it. HTTPS Reporter is an online service so changes are applied to it automatically. If you have a specific feature request that is needed quickly, then we can quote to introduce it into the product as a priority.

Terms & Reasonable Use

When you use the HTTPS Checker app you are required to accept the terms given.
When you use HTTPS Reporter, you are accepting the terms given on the website.
NB Excessive sending of CSP violations may result in the HTTPS Reporter service being suspended until you have reduced the issues to a sustainable level, or through agreeing extra charges to build a bigger infrastructure for your purpose (see section 3.9 of the terms). Terms are not varied for individual customers.

Suggested Process Flow & Tips

PLAN

Estate Registry
Website Technology Discovery
3rd Parties & Tracking
SSL Certificate, Security Policies & CDN
Migration Targets & Content Types
Resourcing & Timescales
Dev & Test Environments
HTTPS Migration Tools
Workflow & Test Procedures
Post-migration Checks
Post-migration Monitoring

HTTPS Migrations can be complex and time-consuming so ensure you plan for this as a priority project that is fully resourced and follows a documented *living* plan which incorporates learning points as you progress through the program of work. Ensure your plan covers changes to analytics and post-migration checks.

SCAN PREPARATION & TUNING

Purchase Licences
Target Pages To Scan
Addressing Common Widespread Issues
Dev / UAT Site Setup
Test Scan On Static Content
Optimise Crawl Speed & Re-test

Decide how you are going to segregate large sites into manageable migration chunks. **Tip:** You can upload a CSV list of URL's to crawl into the HTTPS Checker app for precise targeting and also define sections of the site using robots.txt rules in Advanced Options. You can also set crawl parameters here to tune the speed of the crawl. **Tip:** Check the log files to make sure you haven't set it too fast.

SCAN

Run HTTPS Checker Scan On Target Pages
Monitor Server Load
Collect Issues Requiring Fixes
Check Log Files

Run the scan on your target pages and the HTTPS Checker app will summarise issues and provide PDF and CSV downloads of the issues found too. Check your log files to see which URL's were crawled and which gave errors, so you can investigate them further. Try slowing down the crawl if you are generating a lot of errors.

FIX & RE-SCAN

Assign List Of Issues To Fixers
Record Fixes Made
Re-run Scan On Fixed URL's
Check Log Files

Share the CSV of issues out to your fixers, maybe via Google Docs so it is in a central place, and assign the issues out. You can then check the fixes by importing a specific list of CSV's to crawl and then do a full scan as a double check when you believe everything is fixed.

CSP VIOLATION CAPTURE

Set up CSP Header
Set up & Verify HTTPS Reporter
Check Violations Being Captured

Set up your CSP header and start catching violations in HTTPS Reporter. This will pick up issues that come from dynamically generated content, which the HTTPS Checker app can't find, as well as any other mixed content issues that have crept back in.

FIX CSP VIOLATIONS

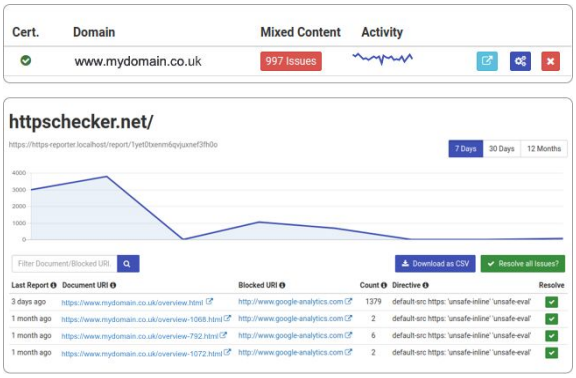
Assign List Of Issues To Fixers
Mark As Fixed In HTTPS Reporter

Mark fixed issues in HTTPS Reporter. A daily email is sent out if new violations are added to the database. NB There is a cap on how many issues are stored so it's important you clear down fixed issues.

CONTINUOUS MONITORING

Daily Email Alert Of CSP Violations
Assign List Of Issues To Fixers
Mark As Fixed In HTTPS Reporter

Continue to monitor your site(s) with HTTPS Reporter for any new mixed content issues that make their way onto the site(s) over time. Look to see if you can tighten up your website management and import tools to stop "http" content being added. You can consider using the HSTS web security policy, but do so with caution.



Look at the graphs in the admin console to quickly see which domains have the most activity at the top level over the last 24 hours, and then on individual domains by either 7, 30 or 360 day periods, as well as being able to view a list of reported issues so you can prioritise what to fix first.