



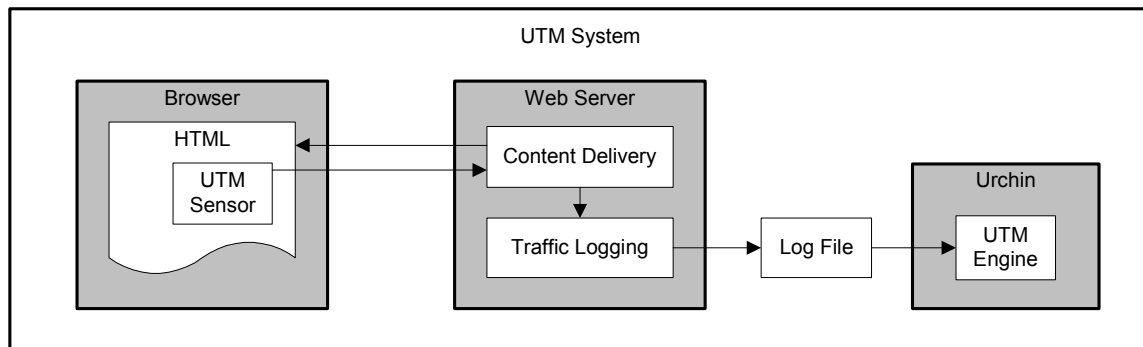
Urchin 4 Whitepaper: URCHIN TRACKING MODULE (UTM)
(©2002 Urchin Software Corporation - doc v1.1)

Overview

The patent-pending Urchin Tracking Module (UTM) available in Urchin 4 was specifically designed to provide the most accurate measurements of unique website visitors. For businesses looking to get a deeper understanding of their online visitor behavior, the UTM is an extremely valuable technology that combines the best of client and server side information while letting you control the data. Easy to install, this technology allows business owners to exactly identify unique visitors, click paths, and return loyalty metrics including: first time visitors, returning visitors, and frequency of use.

There are two components to the Urchin Tracking Module System: the UTM Sensor, which is a lightweight module installed into the content of the website; and the UTM Engine which is part of the log processing Urchin Engine. The UTM Sensor enables client-side data collection, which is then funneled back through the web server augmenting the normal logfile. The client-side information is combined with the existing server-side data by the UTM Engine to provide a more accurate and complete picture of website activity.

The UTM Sensor is a small amount of JavaScript code that accomplishes two important functions. First, the Sensor negates the effects of caching by forcing at least one hit to progress to the original web server for each pageview. The impact on the server is minimal, and the details about the additional hit are logged into the normal web logfiles resulting in a more complete data set. Secondly, the UTM Sensor uniquely identifies each visitor by using client-side cookies to keep track of first time and returning visitors. This cookie identifier is a communication tag only viewable to your web server in the same nature as session ids. It is *not* a third party cookie, which provides information outside your system, violating many privacy policies.

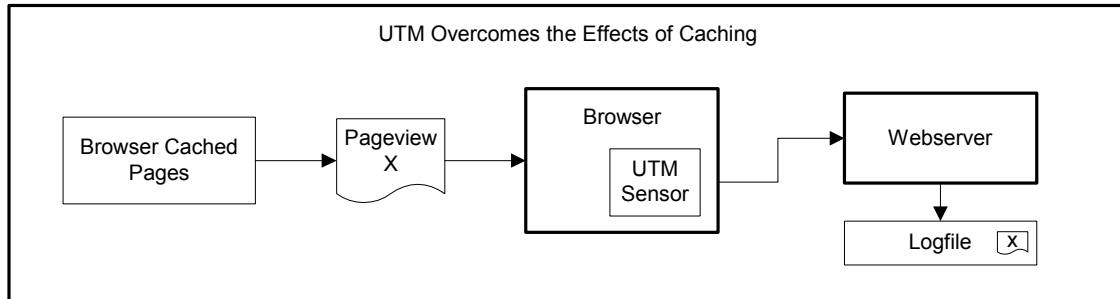


The above diagram illustrates the operation of the UTM System. The web server in the middle of the diagram provides two basic functions: content delivery and logging. The content of the website includes the UTM Sensor which is delivered to the user's browser, shown on the left. The UTM Sensor sets unique identifiers and sends an additional request back to the same web server. This additional request is logged into the normal log file along with all of the normal traffic. The UTM Engine, which is part of the Urchin log processing engine, understands this additional data and merges the two types of data together providing an accurate and more complete picture of visitor behavior.

UTM Sensor

The UTM Sensor increases the accuracy and completeness of logfile data by negating the effects of caching and proxying. The following example illustrates how the UTM Sensor handles caching.

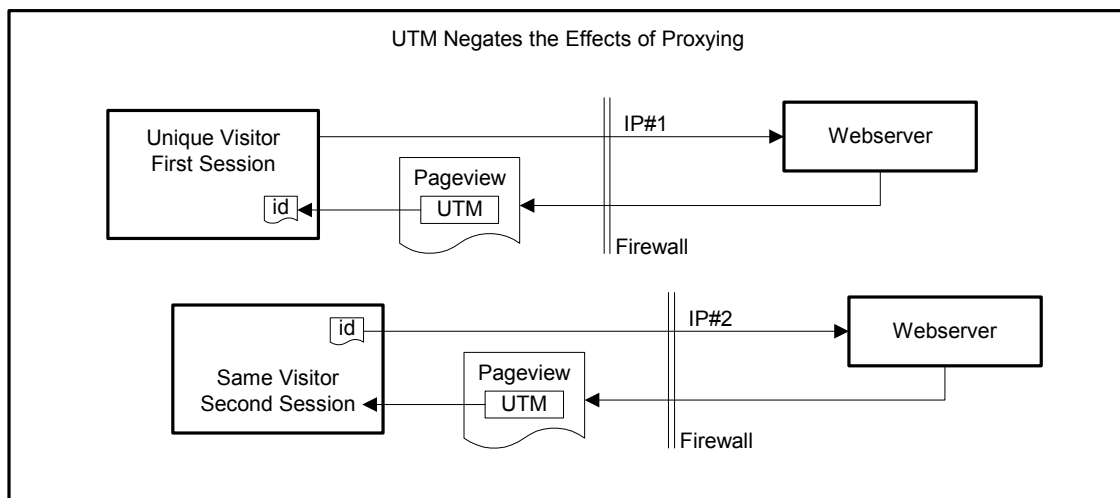
Shown in the figure below, the user receives the content of a pageview from the cached memory of the browser. This typically occurs when the user goes back to a previously viewed page. The same model applies if the caching is provided by a service provider. In the example, the content for page “X” is not delivered from the web server, but from the cached memory of the browser. At this point, there is no knowledge of the pageview as it is not seen by the web server. However, the UTM Sensor activates an additional unique hit that forces at least one small record back to the original web server. This information is logged in the normal logfile, which now has knowledge of the originating “X” pageview.



The second important function of the UTM Sensor is to uniquely identify both sessions and unique visitors. Through a patent-pending combination of browser cookies, the Sensor detects and initializes the unique visitor and session identifiers allowing exact monitoring of new and returning visitors regardless of service provider proxy behavior. Most service providers take advantage of proxying by recycling IP addresses and clustering users behind firewalls. This can cause problems with normal logfile tracking, which typically utilizes the IP address as an identifier of the user.

In the example shown in the figure below, the UTM Sensor is able to pierce the veil of the proxy by utilizing the cookie identifiers instead of the IP addresses. In the figure, a first time unique visitor accesses the website through a firewall with IP address #1. The delivered Pageview contains the UTM Sensor, which sets the identifier on the visitor’s browser.

On the return visit by the same visitor shown in the bottom of the figure, the unique id is passed to the web server along with each request. So even if the user is now assigned a second IP address, the UTM technology properly identifies the visitor with the original id. In addition to negating the effects of complex proxying techniques, this also tracks visitors who travel and may use their laptops from several locations and through several providers.

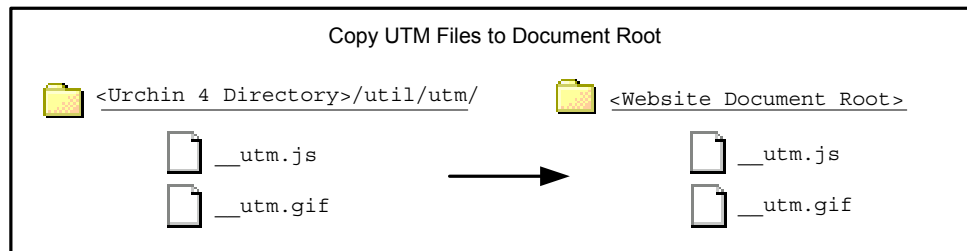


Once the additional UTM data is recorded in the normal web server log , the UTM Engine will recognize and process these additional hits in order to create an exact analysis of each click of the user. During installation, it is important that the logging format is checked for both referral and cookie logging to be present so that all of the appropriate data is stored.

Installation

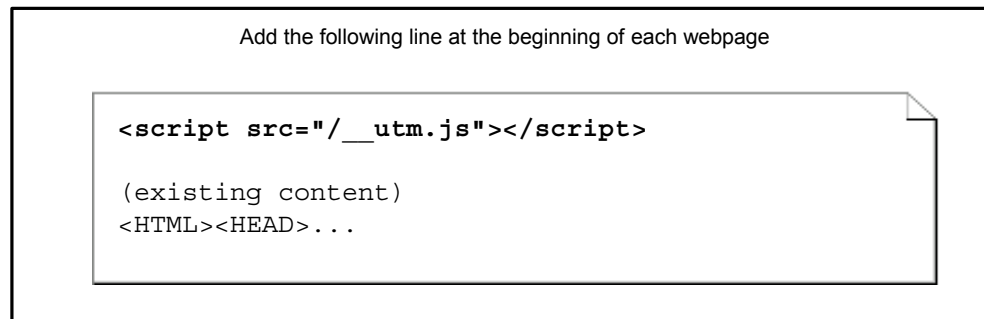
There are three steps to installing the UTM system, which can be accomplished in a very short amount of time. Complex sites may be able to take advantage of existing includes or centralized delivery methods to shorten the installation. During installation, you will need access and permissions to modify the content of the website. You may also need to modify the logging of the web server, which may require a different set of permissions. The following three steps do not necessarily need to be performed in order.

1. Install UTM Sensor into HTML content: The first step in installing the UTM is to include the JavaScript and GIF components of the UTM Sensor in the HTML content of the site. The two pieces necessary for completing this step are included in the `util/utm/` folder within the Urchin distribution. It is important that the names of these two files are not changed and that they are copied to the *document root* directory of the website. Either drag and drop, upload, or copy the `__utm.js` and the `__utm.gif` files into the main directory of your website.



Once these files are in place, you will need to include the `__utm.js` file at the *beginning* of each webpage in the website. If your site utilizes server side includes and you use a header include for each file, it is possible to include the UTM in the beginning of this include file only. It will then automatically be a part of each webpage. For static HTML sites that do not use includes, you will need to modify and add the UTM entry to each page individually. For dynamic sites that use a content generation engine, the UTM can be included at the beginning of the template that is delivered to the customer.

In any case, the following line of code should be included in the beginning of each HTML page that is delivered to the end user. For static sites, edit each webpage and add the bolded line below before the rest of the HTML content.



For sites that undergo regular maintenance or have multiple authors, be sure to build the addition of this line into the your internal website authoring procedures, guidelines, and QA processes.

2. Activate cookies in the logging: The second step to installing the UTM System is to verify and potentially modify the logging format of your web server. For the UTM to function properly, it is required that both referral and cookie information is logged. You will need access to the configuration of the web server. The following general guidelines should work for most IIS and Apache users, however you should check with your system administrator to ensure proper formats.

For Apache Users:

Apache servers typically use a configuration file called “httpd.conf.” Within this file, configuration directives determine the format and location of logfiles. By default, most Apache configurations will log in the NCSA Extended Combined format, which includes referrals and user-agents, but is missing the cookie information. Be sure that your logfiles contain the “%{Cookie}i” field specification. To modify your logging format from the default, a “special” LogFormat directive can be added and then the log files can reference this format using the CustomLog directive.

```
Apache Configuration (httpd.conf)

LogFormat "%h %v %u %t \"%r\" %>s %b \"%{Referer}i\" \"%{User-Agent}i\" \"%{Cookie}i\" " special
...

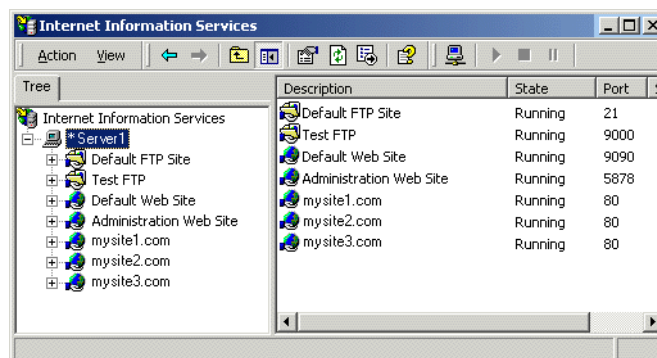
<VirtualHost 63.212.171.6>
ServerName www.mysite.com
DocumentRoot /www/www.mysite.com
CustomLog /var/logs/mysite.log special
</VirtualHost>
```

The above example is provided as a reference and does not apply to all possible Apache settings. Please refer to the Apache documentation and consult with your system administrator on the actual directives needed to activate cookie logging. The LogFormat directive specifies the specific format of the log file. The example shows the addition of the cookie information to the end of the log file. This format is then named “special” so that it can be identified in the virtual host configurations. The CustomLog directive in the virtual host specification identifies the location of the log file and the format to use. The example uses the “special” format as defined previously.

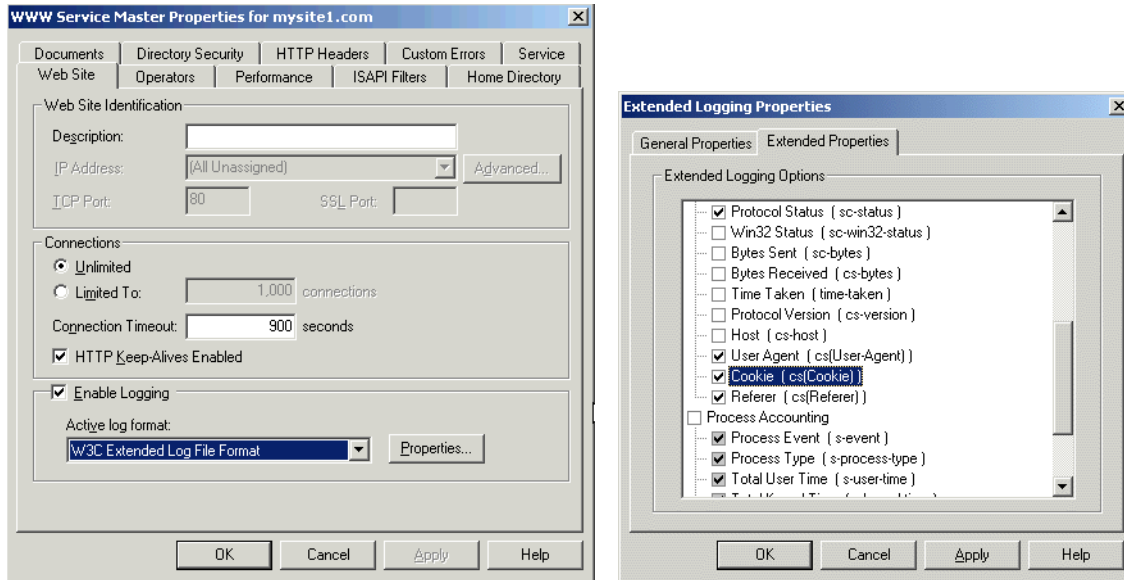
For Microsoft IIS users:

The Internet Services Manager provides a point-and-click interface for adjusting the web server configuration. To access this manager, you will need to login to the web server with the appropriate administrator privileges.

To access the Internet Services Manager, click on the “Start” menu → Settings → Control Panel, and then double click on the Administrative Tools Folder and then on the Internet Services Manager icon to open the manager.



Modifications to each website can either be made individually or the entire server can be modified. In the left window either right-click on the server name to modify the entire server, or right-click on the website name such as “mysite1.com.” Select the “Properties” option to open the properties dialog box. For the entire server, click on the “Edit...” button with “WWW Services” selected in the menu to bring up the Properties dialog box shown on the left below.

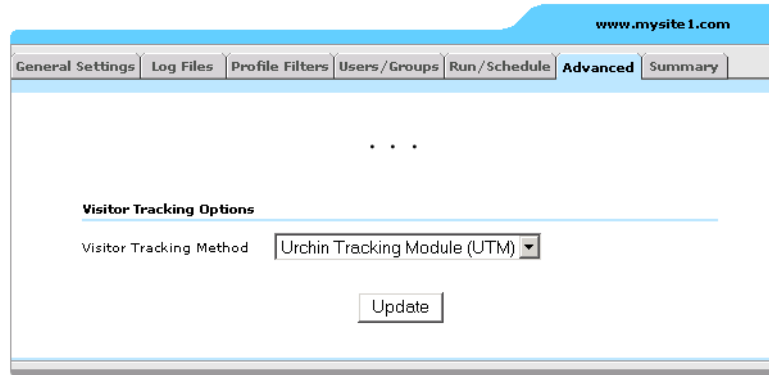


Shown in the bottom half in the above figure on the left, be sure that logging is enabled and set to the “W3C Extended Log File Format.” Then select the “Properties...” button to configure the log file format specifics. The window shown to the right will appear. Click on the “Extended Properties” tab, scroll down and make sure both the “Cookie” and “Referrer” boxes are checked. If not, check these boxes and “Apply” the changes to the site.

Whether you use IIS, Apache or another web server, please refer to your server documentation for more information on configuring logfile formats. All major web servers support the logging of cookies and are easily modified to activate this feature.

3. Set Urchin configuration to UTM: The final step in configuring the UTM for your site is to enable the UTM tracking in the Urchin Configuration. Open the Urchin Configuration either directly on the machine or by logging in remotely as the “admin” user. Your installation instructions will provide more details on how to access the configuration.

Once open, Click on the “Configuration” icon to the left to provide a list of the existing Profiles in the configuration. To enable UTM tracking for a particular profile, click the “Edit” to the right of the profile name. (Note: if you have not already added the profile, do so now using the “Add” button). After clicking on the “Edit” button, click on the “Advanced” tab to bring up the Advanced Settings Window.



Under the “Visitor Tracking Options” section, use the menu to select “Urchin Tracking Module (UTM).” Click the “Update” button to save your settings. That’s it. The installation is complete, and future traffic will contain and benefit from the the UTM System.