

Measurement, collection, analysis and reporting of web data

Web analytics is the measurement, collection, analysis, and reporting of web data to understand and optimize web usage.[1] Web analytics is not just a process for measuring web traffic but can be used as a tool for business and market research and assess and improve website effectiveness. Web analytics applications can also help companies measure the results of traditional print or broadcast advertising campaigns. It can be used to estimate how traffic to a website changes after launching a new advertising campaign. Web analytics provides information about the number of visitors to a website and the number of page views, or create user behavior profiles.[2] It helps gauge traffic and popularity trends, which is useful for market research.

Basic steps of the web analytics process [edit]

Basic Steps of Web Analytics Process

Most web analytics processes come down to four essential stages or steps,[3] which are:

Collection of data: This stage is the collection of the basic, elementary data. Usually, these data are counts of things. The objective of this stage is to gather the data.

Processing of data into information: This stage usually takes counts and makes them ratios, although there still may be some counts. The objective of this stage is to take the data and conform it into information, specifically metrics.

Developing KPI: This stage focuses on using the ratios (and counts) and infusing them with business strategies, referred to as key performance indicators (KPI). Many times, KPIs deal with conversion aspects, but not always. It depends on the organization.

Formulating online strategy: This stage is concerned with the online goals, objectives, and standards for the organization or business. These strategies are usually related to making money, saving money, or increasing market share.

Another essential function developed by the analysts for the optimization of the websites are the experiments

Experiments and testing: A/B testing is a controlled experiment with two variants, in online settings, such as web development.

The goal of A/B testing is to identify and suggest changes to web pages that increase or maximize the effect of a statistically tested result of interest.

Each stage impacts or can impact (i.e., drives) the stage preceding or following it. So, sometimes the data that is available for collection impacts the online strategy. Other times, the online strategy affects the data collected.

Web analytics technologies [edit]

There are at least two categories of web analytics, off-site and on-site web analytics.

Off-site web analytics refers to web measurement and analysis regardless of whether a person owns or maintains a website. It includes the measurement of a website's potential audience (opportunity), share of voice (visibility), and buzz (comments) that is happening on the Internet as a whole.

refers to web measurement and analysis regardless of whether a person owns or maintains a website. It includes the measurement of a website's audience (opportunity), share of voice (visibility), and buzz (comments) that is happening on the Internet as a whole. On-site web analytics, the more common of the two, measure a visitor's behavior once on a specific website. This includes its drivers and conversions; for example, the degree to which different landing pages are associated with online purchases. On-site web analytics measures the performance of a specific website in a commercial context. This data is typically compared against key performance indicators for performance and is used to improve a website or marketing campaign's audience response. Google Analytics and Adobe Analytics are the most widely used on-site web analytics service; although new tools are emerging that provide additional layers of information, including heat maps and session replay.

Historically, web analytics has been used to refer to on-site visitor measurement. However, this meaning has become blurred, mainly because vendors are producing tools that span both categories. Many different vendors provide on-site web analytics software and services. There are two main technical ways of collecting the data. The first and traditional method, server log file analysis, reads the logfiles in which the web server records file requests by browsers. The second method, page tagging, uses JavaScript embedded in the webpage to make image requests to a third-party analytics-dedicated server, whenever a webpage is rendered by a web browser or, if desired, when a mouse click occurs. Both collect data that can be processed to produce web traffic reports.

Web analytics data sources [edit]

The fundamental goal of web analytics is to collect and analyze data related to web traffic and usage patterns. The data mainly comes from four sources:[4]

Direct HTTP request data: directly comes from HTTP request messages (HTTP request headers). Network-level and server-generated data associated with HTTP requests: not part of an HTTP request, but it is required for successful request transmissions - for example, the IP address of a requester. Application-level data sent with HTTP requests: generated and processed by application-level programs (such as JavaScript, PHP, and ASP.Net), including sessions and referrals. These are usually captured by internal logs rather than public web analytics services. External data: can be combined with on-site data to help augment the website behavior data described above and interpret web usage. For example, IP addresses are usually associated with Geographic regions and internet service providers, e-mail open and click-through rates, direct mail campaign data, sales, lead history, or other data types as needed.

Web server log file analysis [edit]

Web servers record some of their transactions in a log file. It was soon realized that these log files could be read by a program to provide data on the popularity of the website. Thus arose web log analysis software.

In the early 1990s, website statistics consisted primarily of counting the number of client requests (or hits) made to the web server. This was a reasonable method initially since each website often consisted of a single HTML file. However, with the introduction of images in HTML, and websites that spanned multiple HTML files, this count became less useful. The first true commercial Log Analyzer was released by IPRO in 1994.[5]

Two units of measure were introduced in the mid-1990s to gauge more accurately the amount of human activity on web servers. These were page views and visits (or sessions). A page view was defined as a request made to the web server for a page, as opposed to a graphic, while a visit was defined as a sequence of requests from a uniquely identified client that expired after a certain amount of inactivity, usually 30 minutes.

The emergence of search engine spiders and robots in the late 1990s, along with web proxies and dynamically assigned IP addresses for large companies and ISPs, made it more difficult to identify unique human visitors to a website. Log analyzers responded by tracking visits by cookies, and by ignoring requests from known spiders.[citation needed]

The extensive use of web caches also presented a problem for log file analysis. If a person revisits a page, the second request will often be retrieved from the browser's cache, and so no request will be received by the web server. This means that the person's path through the site is lost. Caching can be defeated by configuring the web server, but

this can result in degraded performance for the visitor and bigger load on the servers.[6]

Page tagging [edit]

Concerns about the accuracy of log file analysis in the presence of caching, and the desire to be able to perform web analytics as an outsourced service, led to the second data collection method, page tagging or "web beacons".

In the mid-1990s, Web counters were commonly seen $\hat{\epsilon}$ " these were images included in a web page that showed the number of times the image had been requested, which was an estimate of the number of visits to that page. In the late 1990s, this concept evolved to include a small invisible image instead of a visible one, and, by using JavaScript, to pass along with the image request certain information about the page and the visitor. This information can then be processed remotely by a web analytics company, and extensive statistics generated.

The web analytics service also manages the process of assigning a cookie to the user, which can uniquely identify them during their visit and in subsequent visits. Cookie acceptance rates vary significantly between websites and may affect the quality of data collected and reported.

Collecting website data using a third-party data collection server (or even an in-house data collection server) requires an additional DNS lookup by the user's computer to determine the IP address of the collection server. On occasion, delays in completing successful or failed DNS lookups may result in data not being collected.

With the increasing popularity of Ajax-based solutions, an alternative to the use of an invisible image is to implement a call back to the server from the rendered page. In this case, when the page is rendered on the web browser, a piece of JavaScript code would call back to the server and pass information about the client that can then be aggregated by a web analytics company.

Logfile analysis vs page tagging [edit]

Both logfile analysis programs and page tagging solutions are readily available to companies that wish to perform web analytics. In some cases, the same web analytics company will offer both approaches. The question then arises of which method a company should choose. There are advantages and disadvantages to each approach.[7]

Advantages of logfile analysis [edit]

The main advantages of log file analysis over page tagging are as follows:

The web server normally already produces log files, so the raw data is already available. No changes to the website are required.

The data is on the company's servers and is in a standard, rather than a proprietary, format. This makes it easy for a company to switch programs later, use several different programs, and analyze historical data with a new program.

Log files contain information on visits from search engine spiders, which generally are excluded from the analytics tools using JavaScript tagging. (Some search engines might not even execute JavaScript on a page.) Although these should not be reported as part of human activity, it is useful information for search engine optimization.

Log files require no additional DNS lookups or TCP slow starts. Thus there are no external server calls that can slow page load speeds, or result in uncounted page views.

The web server reliably records every transaction it makes, e.g. serving PDF documents and content generated by scripts, and does not rely on the visitors' browsers cooperating.

Advantages of page tagging [edit]

The main advantages of page tagging over log file analysis are as follows:

Counting is activated by opening the page (given that the web client runs the tag scripts), not requesting it from the server. If a page is cached, it will not be counted by server-based log analysis. Cached pages can account for up to one-third of all page views, which can negatively impact many site metrics. [citation needed]

Data is gathered via a component ("tag") in the page, usually written in JavaScript. It is typically used in conjunction with a server-side scripting language (such as PHP) to manipulate and (usually) store it in a database.

The script may have access to additional information on the web client or on the user, not sent in the query, such as visitors' screen sizes and the price of the goods they purchased.

Page tagging can report on events that do not involve a request to the web server, such as interactions within Flash movies, partial form completion, mouse events such as onClick, onMouseOver, onFocus, onBlur, etc.

The page tagging service manages the process of assigning cookies to visitors; with log file analysis, the server has to be configured to do this.

Page tagging is available to companies who do not have access to their web servers.

Lately, page tagging has become a standard in web analytics.[8]

Economic factors [edit]

Logfile analysis is almost always performed in-house. Page tagging can be performed in-house, but it is more often provided as a third-party service. The economic difference between these two models can also be a consideration for a company deciding which to purchase.

Logfile analysis typically involves a one-off software purchase; however, some vendors are introducing maximum annual page views with additional costs to process additional information. [citation needed] In addition to commercial offerings, several open-source logfile analysis tools are available free of charge.

In addition to commercial offerings, several open-source logfile analysis tools are available free of charge. For Logfile analysis data must be stored and archived, which often grows large quickly. Although the cost of hardware to do this is minimal, the overhead for an IT department can be considerable.

For Logfile analysis software need to be maintained, including updates and security patches.

Complex page tagging vendors charge a monthly fee based on volume i.e. number of page views per month collected.

Which solution is cheaper to implement depends on the amount of technical expertise within the company, the vendor chosen, the amount of activity seen on the websites, the depth and type of information sought, and the number of distinct websites needing statistics.

Regardless of the vendor solution or data collection method employed, the cost of web visitor analysis and interpretation should also be included. That is, the cost of turning raw data into actionable information. This can be from the use of third party consultants, the hiring of an experienced web analyst, or the training of a suitable in-house person. A cost-benefit analysis can then be performed. For example, what revenue increase or cost savings can be gained by analyzing the web visitor data?

Hybrid methods [edit]

Some companies produce solutions that collect data through both log files and page tagging and can analyze both kinds. By using a hybrid method, they aim to produce more accurate statistics than either method on its own. An early hybrid solution was produced in 1998 by Rufus Evison.[citation needed]

Geolocation of visitors [edit]

With IP geolocation, it is possible to track visitors' locations. Using an IP geolocation database or API, visitors can be geolocated to city, region, or country level.[9]

IP Intelligence, or Internet Protocol (IP) Intelligence, is a technology that maps the Internet and categorizes IP addresses by parameters such as geographic location (country, region, state, city and postcode), connection type, Internet Service Provider (ISP), proxy information, and more. The first generation of IP Intelligence was referred to as geotargeting or geolocation technology. This information is used by businesses for online audience segmentation in applications such as online advertising, behavioral targeting, content localization (or website localization), digital rights management, personalization, online fraud detection, localized search, enhanced analytics, global traffic management, and content distribution.

Click analytics [edit]

Clickpath Analysis with referring pages on the left and arrows and rectangles differing in thickness and expanse to symbolize movement quantity.

Click analytics, also known as Clickstream is a special type of web analytics that gives special attention to clicks.

Commonly, click analytics focuses on on-site analytics. An editor of a website uses click analytics to determine the performance of his or her particular site, with regards to where the users of the site are clicking.

Also, click analytics may happen real-time or "unreal"-time, depending on the type of information sought. Typically, front-page editors on high-traffic news media sites will want to monitor their pages in real-time, to optimize the content. Editors, designers or other types of stakeholders may analyze clicks on a wider time frame to help them assess performance of writers, design elements or advertisements etc.

Data about clicks may be gathered in at least two ways. Ideally, a click is "logged" when it occurs, and this method requires some functionality that picks up relevant information when the event occurs. Alternatively, one may institute the assumption that a page view is a result of a click, and therefore log a simulated click that led to that page view.

Customer lifecycle analytics [edit]

Customer lifecycle analytics is a visitor-centric approach to measuring that falls under the umbrella of lifecycle marketing.[citation needed] Page views, clicks and other events (such as API calls, access to third-party services, etc.) are all tied to an individual visitor instead of being stored as separate data points. Customer lifecycle analytics attempts to connect all the data points into a marketing funnel that can offer insights into visitor behavior and website optimization.[citation needed]

Other methods [edit]

Other methods of data collection are sometimes used. Packet sniffing collects data by sniffing the network traffic passing between the web server and the outside world. Packet sniffing involves no changes to the web pages or web servers. Integrating web analytics into the webserver software itself is also possible.[10] Both these methods claim

to provide better real-time data than other methods.

On-site web analytics metrics [edit]

There are no globally agreed definitions within web analytics as the industry bodies have been trying to agree on definitions that are useful and definitive for some time. The main bodies who have had input in this area have been the IAB (Interactive Advertising Bureau), JICWEBS (The Joint Industry Committee for Web Standards in the UK and Ireland), and The DAA (Digital Analytics Association), formally known as the WAA (Web Analytics Association, US). However, many terms are used in consistent ways from one major analytics tool to another, so the following list, based on those conventions, can be a useful starting point:

Bounce rate - The percentage of visits that are single-page visits and without any other interactions (clicks) on that page. In other words, a single click in a particular session is called a bounce.

- The percentage of visits that are single-page visits and without any other interactions (clicks) on that page. In other words, a single click in a particular session is called a bounce. Click path - the chronological sequence of page views within a visit or session.

- the chronological sequence of page views within a visit or session. Hit - A request for a file from the webserver. Available only in log analysis. The number of hits received by a website is frequently cited to assert its popularity, but this number is extremely misleading and dramatically overestimates popularity. A single web-page typically consists of multiple (often dozens) of discrete files, each of which is counted as a hit as the page is downloaded, so the number of hits is really an arbitrary number more reflective of the complexity of individual pages on the website than the website's actual popularity. The total number of visits or page views provides a more realistic and accurate assessment of popularity.

- A request for a file from the webserver. Available only in log analysis. The number of hits received by a website is frequently cited to assert its popularity, but this number is extremely misleading and dramatically overestimates popularity. A single web-page typically consists of multiple (often dozens) of discrete files, each of which is counted as a hit as the page is downloaded, so the number of hits is really an arbitrary number more reflective of the complexity of individual pages on the website than the website's actual popularity. The total number of visits or page

views provides a more realistic and accurate assessment of popularity. Page view - A request for a file, or sometimes an event such as a mouse click, that is defined as a page in the setup of the web analytics tool. An occurrence of the script being run in page tagging. In log analysis, a single page view may generate multiple hits as all the resources required to view the page (images, .js and .css files) are also requested from the webserver.

- A request for a file, or sometimes an event such as a mouse click, that is defined as a page in the setup of the web analytics tool. An occurrence of the script being run in page tagging. In log analysis, a single page view may generate multiple hits as all the resources required to view the page (images, .js and .css files) are also requested from the webserver. Visitor/unique visitor/unique user - The uniquely identified client that is generating page views or hits within a defined period time (e.g. day, week or month). A uniquely identified client is usually a combination of a machine (one's desktop computer at work for example) and a browser (Firefox on that machine). The identification is usually via a persistent cookie that has been placed on the computer by the site page code. An older method, used in log file analysis, is the unique combination of the computer's IP address and the User-Agent (browser) information provided to the web server by the browser. It is important to understand that the "Visitor" is not the same as the human being sitting at the computer at the time of the visit, since an individual human can use different computers or, on the same computer, can use different browsers, and will be seen as a different visitor in each circumstance. Increasingly, but still, somewhat rarely, visitors are uniquely identified by Flash LSO's (Local Shared Objects), which are less susceptible to privacy enforcement.

- The uniquely identified client that is generating page views or hits within a defined period time (e.g. day, week or month). A uniquely identified client is usually a combination of a machine (one's desktop computer at work for example) and a browser (Firefox on that machine). The identification is usually via a persistent cookie that has been placed on the computer by the site page code. An older method, used in log file analysis, is the unique combination of the computer's IP address and the User-Agent (browser) information provided to the web server by the browser. It is important to understand that the "Visitor" is not the same as the human being sitting at the computer at the time of the visit, since an individual human can use different computers or, on the same computer, can use different browsers, and will be seen as a different visitor in each circumstance. Increasingly, but still, somewhat rarely, visitors are uniquely identified by Flash LSO's (Local Shared Objects), which are less susceptible to privacy enforcement. Visit/session - A visit or session is defined as a series of page requests or, in the case of tags, image requests from the same uniquely identified client. A unique client is commonly identified by an IP address or a unique ID that is placed in the browser cookie. A visit is considered ended when no requests have been recorded in some number of

elapsed minutes. A 30-minute limit ("time out") is used by many analytics tools but can, in some tools (such as Google Analytics), be changed to another number of minutes. Analytics data collectors and analysis tools have no reliable way of knowing if a visitor has looked at other sites between page views; a visit is considered one visit as long as the events (page views, clicks, whatever is being recorded) are 30 minutes or less close together. Note that a visit can consist of a one-page view or thousands. A unique visit session can also be extended if the time between page loads indicates that a visitor has been viewing the pages continuously.

- A visit or session is defined as a series of page requests or, in the case of tags, image requests from the same uniquely identified client. A unique client is commonly identified by an IP address or a unique ID that is placed in the browser cookie. A visit is considered ended when no requests have been recorded in some number of elapsed minutes. A 30-minute limit ("time out") is used by many analytics tools but can, in some tools (such as Google Analytics), be changed to another number of minutes. Analytics data collectors and analysis tools have no reliable way of knowing if a visitor has looked at other sites between page views; a visit is considered one visit as long as the events (page views, clicks, whatever is being recorded) are 30 minutes or less close together. Note that a visit can consist of a one-page view or thousands. A unique visit session can also be extended if the time between page loads indicates that a visitor has been viewing the pages continuously. Active time/engagement time - Average amount of time that visitors spend actually interacting with content on a web page, based on mouse moves, clicks, hovers, and scrolls. Unlike session duration and page view duration/time on page, this metric can accurately measure the length of engagement in the final page view, but it is not available in many analytics tools or data collection methods.

- Average amount of time that visitors spend actually interacting with content on a web page, based on mouse moves, clicks, hovers, and scrolls. Unlike session duration and page view duration/time on page, this metric accurately measure the length of engagement in the final page view, but it is not available in many analytics tools or data collection methods. Average page depth/page views per average session - Page depth is the approximate "size" of an average visit, calculated by dividing the total number of page views by the total number of visits.

- Page depth is the approximate "size" of an average visit, calculated by dividing the total number of page views by the total number of visits. Average page view duration - Average amount of time that visitors spend on an average page of the site.

- Average amount of time that visitors spend on an average page of the site. Click - "refers to a single instance of a

user following a hyperlink from one page in a site to another".

- "refers to a single instance of a user following a hyperlink from one page in a site to another". Event - A discrete action or class of actions that occur on a website. A page view is a type of event. Events also encapsulate clicks, form submissions, keypress events, and other client-side user actions.

- A discrete action or class of actions that occur on a website. A page view is a type of event. Events also encapsulate clicks, form submissions, keypress events, and other client-side user actions. Exit rate/% exit - A statistic applied to an individual page, not a web site. The percentage of visits seeing a page where that page is the final page viewed in the visit.

- A statistic applied to an individual page, not a web site. The percentage of visits seeing a page where that page is the final page viewed in the visit. First visit/first session - (also called 'Absolute Unique Visitor' in some tools) A visit from a uniquely identified client that has theoretically not made any previous visits. Since the only way of knowing whether the uniquely identified client has been to the site before is the presence of a persistent cookie or via digital fingerprinting that had been received on a previous visit, the First Visit label is not reliable if the site's cookies have been deleted since their previous visit.

- (also called 'Absolute Unique Visitor' in some tools) A visit from a uniquely identified client that has theoretically not made any previous visits. Since the only way of knowing whether the uniquely identified client has been to the site before is the presence of a persistent cookie or via digital fingerprinting that had been received on a previous visit, the label is not reliable if the site's cookies have been deleted since their previous visit. Frequency/session per unique - Frequency measures how often visitors come to a website in a given time period. It is calculated by dividing the total number of sessions (or visits) by the total number of unique visitors during a specified time period, such as a month or year. Sometimes it is used interchangeable with the term "loyalty."

- Frequency measures how often visitors come to a website in a given time period. It is calculated by dividing the total number of sessions (or visits) by the total number of unique visitors during a specified time period, such as a month or year. Sometimes it is used interchangeable with the term "loyalty." Impression - The most common definition of impression is an instance of an advertisement appearing on a viewed page. Note that an advertisement can be displayed on a viewed page below the area actually displayed on the screen, so most measures of impressions do not

necessarily mean an advertisement has been view-able.

- The most common definition of is an instance of an advertisement appearing on a viewed page. Note that an advertisement can be displayed on a viewed page below the area actually displayed on the screen, so most measures of impressions do not necessarily mean an advertisement has been view-able. New visitor - A visitor that has not made any previous visits. This definition creates a certain amount of confusion (see common confusions below), and is sometimes substituted with analysis of first visits.

- A visitor that has not made any previous visits. This definition creates a certain amount of confusion (see common confusions below), and is sometimes substituted with analysis of first visits. Page time viewed/page visibility time/page view duration - The time a single page (or a blog, ad banner) is on the screen, measured as the calculated difference between the time of the request for that page and the time of the next recorded request. If there is no next recorded request, then the viewing time of that instance of that page is not included in reports.

- The time a single page (or a blog, ad banner) is on the screen, measured as the calculated difference between the time of the request for that page and the time of the next recorded request. If there is no next recorded request, then the viewing time of that instance of that page is not included in reports. Repeat visitor - A visitor that has made at least one previous visit. The period between the last and current visit is called visitor recency and is measured in days.

- A visitor that has made at least one previous visit. The period between the last and current visit is called visitor recency and is measured in days. Return visitor - A unique visitor with activity consisting of a visit to a site during a reporting period and where the unique visitor visited the site prior to the reporting period. The individual is counted only once during the reporting period.

- A unique visitor with activity consisting of a visit to a site during a reporting period and where the unique visitor visited the site prior to the reporting period. The individual is counted only once during the reporting period. Session duration/visit duration - Average amount of time that visitors spend on the site each time they visit. It is calculated as the sum total of the duration of all the sessions divided by the total number of sessions. This metric can be complicated by the fact that analytics programs can not measure the length of the final page view. [citation needed]

- Average amount of time that visitors spend on the site each time they visit. It is calculated as the sum total of the duration of all the sessions divided by the total number of sessions. This metric can be complicated by the fact that analytics programs can not measure the length of the final page view. Single page visit/singleton - A visit in which only a single page is viewed (this is not a 'bounce').

- A visit in which only a single page is viewed (this is not a 'bounce'). Site overlay is a report technique in which statistics (clicks) or hot spots are superimposed, by physical location, on a visual snapshot of the web page.

is a report technique in which statistics (clicks) or hot spots are superimposed, by physical location, on a visual snapshot of the web page. Click-through rate is a ratio of users who click on a specific link to the number of total users who view a page, email, or advertisement. It is commonly used to measure the success of an online advertising campaign for a particular website as well as the effectiveness of email campaigns.

Off-site web analytics [edit]

Off-site web analytics is based on open data analysis, social media exploration, share of voice on web properties. It is usually used to understand how to market a site by identifying the keywords tagged to this site, either from social media or from other websites.

Common sources of confusion in web analytics [edit]

The hotel problem [edit]

The hotel problem is generally the first problem encountered by a user of web analytics. The problem is that the unique visitors for each day in a month do not add up to the same total as the unique visitors for that month. This appears to an inexperienced user to be a problem in whatever analytics software they are using. In fact it is a simple property of the metric definitions.

The way to picture the situation is by imagining a hotel. The hotel has two rooms (Room A and Room B).

Day 01 Day 02 Day 03 Total Room A John John Mark 2 Unique Users Room B Mark Anne Anne 2 Unique Users Total 2 2 2 ?

As the table shows, the hotel has two unique users each day over three days. The sum of the totals with respect to the days is therefore six.

During the period each room has had two unique users. The sum of the totals with respect to the rooms is therefore four.

Actually only three visitors have been in the hotel over this period. The problem is that a person who stays in a room for two nights will get counted twice if they are counted once on each day, but are only counted once if the total for the period is looked at. Any software for web analytics will sum these correctly for the chosen time period, thus leading to the problem when a user tries to compare the totals.

[Analytics Poisoning](#) [edit]

As the internet has matured, the proliferation of automated bot traffic has become an increasing problem for the reliability of web analytics.[citation needed] As bots traverse the internet, they render web documents in ways similar to organic users, and as a result may incidentally trigger the same code that web analytics use to count traffic. Jointly, this incidental triggering of web analytics events impacts interpretability of data and inferences made upon that data. IPM provided a proof of concept of how Google Analytics as well as their competitors are easily triggered by common bot deployment strategies.[11]

[Web analytics methods](#) [edit]

[Problems with cookies](#) [edit]

Historically, vendors of page-tagging analytics solutions have used third-party cookies sent from the vendor's domain instead of the domain of the website being browsed. Third-party cookies can handle visitors who cross multiple unrelated domains within the company's site, since the cookie is always handled by the vendor's servers.

However, third-party cookies in principle allow tracking an individual user across the sites of different companies, allowing the analytics vendor to collate the user's activity on sites where he provided personal information with his activity on other sites where he thought he was anonymous. Although web analytics companies deny doing this, other

companies such as companies supplying banner ads have done so. Privacy concerns about cookies have therefore led a noticeable minority of users to block or delete third-party cookies. In 2005, some reports showed that about 28% of Internet users blocked third-party cookies and 22% deleted them at least once a month.[12] Most vendors of page tagging solutions have now moved to provide at least the option of using first-party cookies (cookies assigned from the client subdomain).

Another problem is cookie deletion. When web analytics depend on cookies to identify unique visitors, the statistics are dependent on a persistent cookie to hold a unique visitor ID. When users delete cookies, they usually delete both first- and third-party cookies. If this is done between interactions with the site, the user will appear as a first-time visitor at their next interaction point. Without a persistent and unique visitor id, conversions, click-stream analysis, and other metrics dependent on the activities of a unique visitor over time, cannot be accurate.

Cookies are used because IP addresses are not always unique to users and may be shared by large groups or proxies. In some cases, the IP address is combined with the user agent in order to more accurately identify a visitor if cookies are not available. However, this only partially solves the problem because often users behind a proxy server have the same user agent. Other methods of uniquely identifying a user are technically challenging and would limit the trackable audience or would be considered suspicious. Cookies reach the lowest common denominator without using technologies regarded as spyware.[13]

Secure analytics (metering) methods [edit]

It may be good to be aware that third-party information gathering is subject to any network limitations and security applied. Countries, Service Providers and Private Networks can prevent site visit data from going to third parties. All the methods described above (and some other methods not mentioned here, like sampling) have the central problem of being vulnerable to manipulation (both inflation and deflation). This means these methods are imprecise and insecure (in any reasonable model of security). This issue has been addressed in several papers,[14][15][16][17] but to date the solutions suggested in these papers remain theoretical.

See also [edit]

References [edit]

Reference

[The Pelvic Girdle: An integration of clinical expertise and research](#)

[Essentials of Descriptive-Interpretive Qualitative Research: A Generic Approach
\(Essentials of Qualitative Methods\)](#)