

# Using a web analytics system as the basis for integration with CPA services

## Vyacheslav I. Zhukov

*Assistant Professor, Department of Innovation and Business in Information Technologies  
National Research University Higher School of Economics  
Address: 20, Myasnitskaya Street, Moscow, 101000, Russian Federation  
E-mail: vzhukov@hse.ru*

## Mikhail M. Komarov

*Associate Professor, Department of Innovation and Business in Information Technologies  
National Research University Higher School of Economics  
Address: 20, Myasnitskaya Street, Moscow, 101000, Russian Federation  
E-mail: mkomarov@hse.ru*

### Abstract

In modern e-business, there are many ways to attract potential customers to the site both with the help of offline and online methods. Companies usually use several channels to attract customers, differing in the placement of advertising, payment models and other parameters.

One of the most popular online methods is using CPA networks, which allow webmasters to place on their websites links to the advertised website and earn rewards for customers who purchased a service by clicking on the link. CPA networks work on the basis of payment for achieving targeted goals. A targeted goal can occur both online and offline. The most important task is to link the source of attraction (usually certain UTM tags) to the target goal of the client, since remuneration for the CPA network should only occur for orders from customers who are drawn to the CPA network. There are problems fixing operations, linking to source of attraction, storing and providing access to these data.

In this paper, we give a brief overview of various approaches to solving the problem: log analysis, use of marketing pixels and web analytics tools. We have analyzed the benefits and challenges of these methods, which were given to solve the task of fixing the target actions of clients and provided access to the data for the CPA network. Also in this article, we have described a practical case of integration with the CPA network based on the use of end-to-end web analytics. The advantages, disadvantages and limitations of the proposed method are set out in this paper.

**Key words:** web analytics, Internet marketing, CPA networks, end-to-end web analytics, Google Analytics, CRM.

**Citation:** Zhukov V.I., Komarov M.M. (2017) Using a web analytics system as the basis for integration with CPA services. *Business Informatics*, no. 4 (42), pp. 47–54.  
DOI: 10.17323/1998-0663.2017.4.47.54.

## Introduction

**R**unning an online business usually requires dealing with the problem of landing target customers using the global network of Internet. This problem can be solved using both online and offline methods. The online methods suggest advertising in one form or another being placed on external sources linking to the company site. Banners, contextual advertising, display advertising, links in the articles and other approaches are used for that purpose.

To determine the payment for placement of advertisements on an external source, a few models are used: payment for the placement for a defined period, payment for the number of views (cost per view, CPV), or a frequently used scheme of payment for a thousand of impressions (cost per millennium, CPM), payment for clicks (cost per click, CPC), payment for the target actions (cost per action, CPA) [1]. A payment model for the landed customers as a special case of the CPA model also exists. Each of the models has its pros and cons: as for the CPC or CPV the result cannot be ensured in contrast to CPA providing the advertiser pay only for the landed customers who make the payments. Thus, using CPA allows us to significantly reduce the risk of potential financial losses [2]. However, it should be noted that the CPA model demands that significant technical improvements be made by the advertiser that are not required when using CPV or CPM because the website of the advertisement placement usually provides convenient ways of publishing it.

CPA networks operate based on the CPA model. The CPA networks are binding and medium agents between the site owners (often with low traffic) and the business that places the ads on these websites. Webmasters publish the ads on their sites. Users click on these ads, and the webmasters get the payments for the landed customers in case the users make the target actions. The type of action and its

price are defined by the advertiser and they are named and offer. The CPA network is in fact an intermediate link between the huge amount of the webmasters, their sites and the companies which want to attract the audience from these sites to their own websites for the purpose of selling goods and services.

Usually the advertiser uses a couple of channels of customers' acquisition in the Internet. Special parameters (codes) should be added to the link to determine the acquisition source. Usually they are the UTM codes [3]. The special identifiers assigned to customers are often used for the referral programs [4].

Before doing the targeted action, the customer client can enter the site via several channels (for example, from contextual ads and then from the search results). To consider this fact during the process of analyzing acquisition source effectiveness, you should define the attribution as a rule of assigning the target action to the acquisition source or the rule of distributing value by the conversion value between the sources. The attribution of the last relevant click [5] is usually used for the CPA networks without taking free traffic sources into account (direct traffic, unpaid search traffic) i.e. the last payment click attribution.

Insofar as the payment is based on the number and types of the target actions, there appears to be a problem of recording these actions on the side of the advertiser and passing firm data about them to the CPA network in order to measure the payment for the webmasters.

### 1. Problem-solving approaches

A few methods of recording the target actions of the users on the website exist: who got in the site from a certain channel to provide access to them for the CPA network: the web server log analysis service, web beacons activation and using the web analysis record-keeping systems (trackers).

### 1.1. Web server log analysis service

In this case, all the “mechanics” of the attribution and attaining the target actions are developed on the server as the log analysis service [6] that records all the references to the web pages; submitting the transaction forms and other activities of the users. The referral and the UTM codes used for the user to get into the site are recorded in particular. It also records the customer’s IP-address, his identifier and other service information (browser and its version, operating system, time of the reference etc.) [7]. These logs are usually saved in one of the following formats [8]:

- ◆ NCSA Common Log;
- ◆ NCSA Combined Log;
- ◆ NCSA Separate Log;
- ◆ W3C Extended Log.

Further, the service for building the reports on this data including the attribution model logic is developed. If necessary, this data is combined with the other data sources of the company, for example, with the CRM-systems. The data is linked based on the client’s identifier that should be the same for the same customer in all the systems. To integrate with the CPA network, a special API is carried out to provide access to the information about the payments of the customers landed from the CPA network (defined based on the UTM codes).

The ability to get absolute accuracy of the calculations and carry out a larger amount of assorted target actions can be seen as the advantage of this approach. In addition, the possibility exists to record the actions done beyond the website, for example, to identify users who paid for the service or the product in cash, using payment terminals or by bank transfer due to the ability of integration with other company’s internal data sources.

We should consider the substantial expenses for the development of this service (or the

deployment of a ready-made solution) and its further support as the disadvantage. It is worth mentioning the issue of confidence, since the business can provide false information for the CPA network and it is quite difficult to check this.

### 1.2. Web beacons

This approach is based on the activation of external pixels when certain events take place on the website [9, 10]. Special pixels of the CPA network are placed on the company’s site. They are activated if certain target actions have been done. They usually are gif 1x1 px. [11], located on the CPA network’s sever. The pixel’s activation activates passing the parameters and the cookies of the user. These pixels are activated only for the users attributed based on the UTM codes as those landed from the CPA network. Thus, CPA networks get information about attaining the target actions. Usually this solution is used for relatively simple target actions that take place on the site (viewing articles, clicking certain buttons or submitting a request).

The advantages of this approach are transparency (as the logics of pixel activation can be checked by the CPA network) and unsophisticated deployment in the customer’s site.

The disadvantage is poor scalability of the approach: when increasing the number of recorded target actions each of them should be separately marked with a separate pixel with other parameters of activation. It is worth mentioning the problems related to the modernization of the website’s interface: it is necessary to keep the domain logic of activating these pixels. There is a problem of the complexity of recording events that can take place beyond the website, for example, when the client pays for the service. Another disadvantage is the partial accuracy of data due to losses related to JavaScript turned off at the user’s browser and ad blocking tools.

### 1.3. Using the goals in the web analytics systems

In this case, we use the goals that are already customized in the web analytics systems (for example, Google Analytics<sup>1</sup>, Yandex.Metrika<sup>2</sup>, Piwik<sup>3</sup>, etc.) and the abilities of these systems to set the conversions attributes by the sources. The CPA network specialists have access to the web analytics trackers with filtering on the traffic sources corresponding to the CPA network (usually it is a specific code in the UTM parameters).

For specialists, there is CPA network access to the trackers of web analytics. Based on this data, reports are generated and the amount of remuneration is determined. The reports are built and the payment value is determined based on this data.

Transparency is also an advantage of this method, since the CPA network always can check the consistency of collecting the goals and attributions. We should also consider the relative cheapness of integration: usually the counters are already installed on the client's site and the goals are configured and are used for analysis tasks within the company.

Loss of data is the disadvantage. Its reasons coincide with the "web beacons" approach.

## 2. Case description

### 2.1. Task assignment

Company "A" works in the field of domain sales and providing hosting services. It decided to use a CPA network to increase sales. To solve the assigned task, the "B" CPA network was chosen. Company "A" specified the "paid order" to be the target action. However, since the company sells a wide range of different services with different marginality, the payment ("offer") for the purchase by the customer led

by the CPA network, should also vary depending on the type of service. It is also important that a huge amount of payments be done not via the website but via the payment terminals, by bank transfers etc.

The CPA network demanded availability of the API for access to the data on service payments of the users led from this CPA network with passing information about the category of the purchased service; the attribution model – the last payment click. Company "A" did not have such a service and its development would require extensive resources. That brought to nought the beneficial effect of using the CPA network.

Previously, company "A" implemented the end-to-end web analytics approach [3] to carry out more accurate analysis of advertising campaign efficiency and deep analysis of the customer's behavior from the moment of client acquisition to the site until the moment of the real payment, based on the integration of Google Analytics with the internal self-developed billing solution via the Measurement Protocol (MP API) [12, 13]. That allowed them to record the target actions of the users (payments) done beyond the website.

To carry out technical integration into the CPA network with a view to economy of resources, the decision was made to use the existing system of end-to-end analytics but improving it.

### 2.2. Description of the integration of Google Analytics with the billing

When the customer enters into a purchase order, the Client ID (cid) is transmitted from the Google Analytics system (used by Google Analytics to identify the individual user of the website) as the parameter, and when the customer pays for the order, a special pur-

<sup>1</sup> <https://www.google.com/analytics/>

<sup>2</sup> <https://metrika.yandex.ru/>

<sup>3</sup> <https://piwik.org/>

pose request – the Enhanced Ecommerce<sup>4</sup> request – is sent from the billing to Google Analytics with binding to this customer. This data multiplexing allows tracking the customer’s full path from getting into the site to the real payment for a specific service and the profit of selling this service.

The payment hit includes the following information:

- ◆ the number of the customer’s order;
- ◆ SKU of the service;
- ◆ category of the service;
- ◆ name of the service;
- ◆ revenue;
- ◆ profit;
- ◆ clientID.

### 2.3. “Last payment click” emulation

To complete the tasks of the regular web analytics in the company, the Google Analytics system was customized according to the standard attribution model – the last non-direct click. However, it doesn’t match up for the integration with the CPA network since, besides the paid traffic, the unpaid traffic (mostly organic search and referral traffic) can dominate and “erase” the paid acquisition source.

This would take place if the user followed the paid source and then, for example, followed the results of the organic search and bought the goods. At that time, he searched for the name of the company.

To implement the “last payment click” model for the integration with the CPA network, there was an additional Javascript assembly unit developed on the front-end side of the website that emulated the “last payment click” logics via the specific user parameter of Google Analytics set on the user level (saved during the whole life time of the cli-

entID). The UTM codes were recorded to these parameters and sent to Google Analytics. The information was available for analysis and building the reports; at the same time the processes and systems already used in the company to analyze customer behavior on the site were not damaged.

### 2.4. Data access API

The mentioned organization allowed making Google Analytics a general store of the data on the full track of the client’s interaction with the site: from the acquisition source until the moment of the real payment. At the same time, since Google Analytics already has the instruments of getting the data via API, there was no necessity to develop the service for this data access.

The individual Google Analytics view was customized for the CPA network. It included the CPA networks of the traffic source that were transmitted via the UTM codes. The information was available through Google Analytics Core API [14].

### 2.5. Interaction patterns and stages

The integration diagram with the CPA network shown on *figure 1* provides for carrying out the following steps.

**Step 1.** The customers get into the client’s site from the webmasters’ resources via the URL with special UTM parameters that contain the information about the network and the webmaster identifier.

**Step 2.** When the site is being loaded, the information about the UTM codes in the specific parameters on the user level is also transmitted to Google Analytics besides the standard information about the page view (cid etc.).

**Step 3.** When ordering the service on the site,

<sup>4</sup> <https://developers.google.com/>

the cid (Client ID, client identifier in Google Analytics service) is sent to billing.

**Step 4.** After the service payment, billing transmits the hit with the contents of the paid services and the cid attached to the order to Google Analytics via MP. After receiving the payment data, Google Analytics relates it with the specified user (based on cid) whose visit information was recorded at step 2.

**Step 5.** Using Google Analytics Core API, the CPA network uploads the information about the payments of the services and their categories carried out by the customers who got into the site of company “A” from the resources of the webmasters who work with this CPA network. The subsequent cost clearings between the company and the CPA network are carried out based on this information.

### 3. Advantages and restrictions of the suggested method

The described method has a variety of advantages and restrictions that should be taken into account when solving the problem of integration with the CPA network.

One of the main advantages of the method is the fact that a small improvement of the infrastructure developed for the analytical tasks allows us to solve the integration problem. This method would allow the following things for companies that use end-to-end analytics in their work:

- ◆ significantly reduce the time and costs for solving the problem of integration with the CPA network;
- ◆ to test this channel of solving the market-

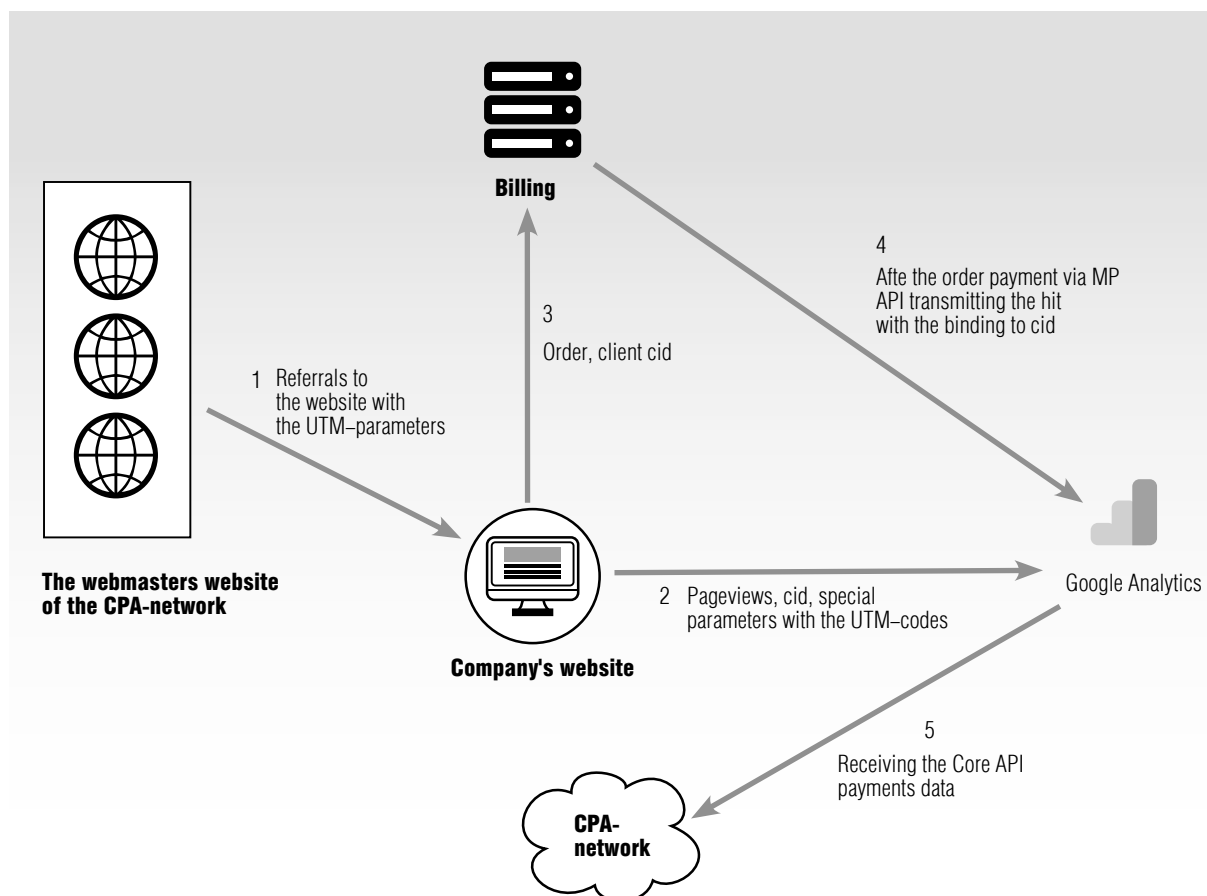


Fig. 1. Integration solution with the CPA network

ing problems of increasing the customer base, growth of profits or launching a new service as quickly as possible;

- ◆ quickly stop working with a promotion channel in case of its inefficiency;

- ◆ the ability to take service categories into account when calculating payment for the webmasters.

At the same time, the given method has a couple of technical restrictions that cause partial loss of data.

First of all, Google Analytics can get under the operation of the anonymizers and ad blockers for example, AdBlock<sup>5</sup>, that can cause the loss of payment data for the customers attracted from the CPA network who use these kinds of blockers. This problem can be completely solved only by using the previously described approach of log analysis, but due to the high cost of its development and support the common method is to simply increase the webmasters' reward by a certain percentage. This method is more profitable than log analysis system implementation and support for small volumes of the orders.

Secondly, some customers can get to know about the services (be attracted to the site) using one device (e.g., a cellphone) and buy the service using another device. This situation can also cause the loss of payment data in this circuit. To solve this problem, you can use one of the methods to identify the customer from various devices [13] by additionally passing the userID from billing or CRM to Google Analytics. This solution would be successful if the user logged in to the site on each of the devices.

## Conclusion

As far as the information about the real service payments of the site's users was bound with their information in the Google Analytics system and this information was available to be received by the CPA network, the integration task was solved. However, the transmitting of the goods categories and their names allowed estimation of the various payments for the webmasters depending on the service paid by the attracted customer.

Due to the built-in functionality of Google Analytics (which includes integration with other company sources of information and the possibility to access this information using API), it became possible to solve two business tasks at once: to bind the visitor's clickstream on the site to the payment data for deeper and more accurate analysis of the customer's behavior (end-to-end web analytics) and provide integration with the CPA network.

This method expands the approach to carry out the integration with the CPA network based on systems of web analytics, upgrading it by the method of end-to-end web analytics. The method allows one to record the actions of the customer beyond the website and at the same time to bind these actions with the visitor's clickstream and the acquisition source. This approach can also be used to record other types of target actions that take place offline, preceded by visiting the website. ■

## References

1. Rzemieniak M. (2015) Measuring the effectiveness of online advertising campaigns in the aspect of e-entrepreneurship. *Procedia Computer Science*, no. 65, pp. 980–987.
2. Hu Y., Shin J., Tang Z. (2012) *Performance-based pricing models in online advertising: Cost per click versus cost per action*. Atlanta: Georgia Institute, 2012.

<sup>5</sup> <https://adblockplus.org/ru/android>

3. Korobkov S.A. (2016) Osnovnye komponenty sistemy skvoznoy marketingovoy Internet-analitiki dlya malogo biznesa [Main components of a system of end-to-end marketing Internet-analytics for small business]. *International Scientific and Research Journal*, no. № 11–1 (53), pp. 48–50 (in Russian).
4. Brown B.C. (2009) *The complete guide to affiliate marketing on the web: How to use and profit from affiliate marketing programs*. Ocala, Florida: Atlantic Publishing Company.
5. Berman R. (2015) *Beyond the last touch: Attribution in online advertising*. Available at: <http://ron-berman.com/papers/attribution.pdf> (accessed 01 December 2017).
6. Grace L.K.J., Maheswari V., Nagamalai D. (2011) Analysis of web logs and web user in web mining. *International Journal of Network Security and its Applications*, vol. 3, no. 1, pp. 99–110.
7. Goel N., Jha C.K. (2013) Analyzing users behavior from web access logs using automated log analyzer tool. *International Journal of Computer Applications*, vol. 62, no. 2, pp. 29–33.
8. Booth D., Jansen B.J. (2009) A review of methodologies for analyzing websites. *Handbook of research on web log analysis* (eds. B.J. Jansen, A. Spink, I. Taksa). Hershey, PA: IGI Global, pp. 143–164.
9. D., Kaushik A. (2009) Web Analytics 2.0: Empowering customer centricity. *Search Engine Marketing Journal*, vol. 2, no. 1, pp. 5–11. Available at: <https://pdfs.semanticscholar.org/f804/b8c0f66b28220d-64060e27892dbe0a7a3baa.pdf> (accessed 01 December 2017).
10. Mayer J.R., Mitchell J.C. (2012) Third-party web tracking: Policy and technology. Proceedings of *2012 IEEE Symposium on Security and Privacy (SP 2012)*. San Francisco, California, 20–23 May 2012, pp. 413–427.
11. Dwyer C. (2009) Behavioral targeting: A case study of consumer tracking on Levis.com. Proceedings of the *Fifteenth Americas Conference on Information Systems (AMCIS 2009)*. San Francisco, California, 6–9 August 2009, pp. 1–10.
12. Waisberg D. (2015) *Google Analytics integrations*. John Wiley & Sons.
13. Weber J. (2015) *Practical Google Analytics and Google Tag Manager for developers*. Apress.
14. Clifton B. (2012) *Advanced web metrics with Google Analytics*. John Wiley & Sons.