Available online at www.sciencedirect.com

ScienceDirect

Procedia - Social and Behavioral Sciences 195 (2015) 1840 – 1845

Procedia Social and Behavioral Sciences

World Conference on Technology, Innovation and Entrepreneurship

Privacy Preserving Secure Online Advertising

Mustafa Oger[a]*, Isa Olmez[a], Erinc Inci[a], Serkan Kücükbay[a], Fatih Emekci[b]

[a]*Comodo Yazılım A.S., Çankaya, Ankara, Turkey*
[b]*Turgut Özal Üniversitesi, 06010, Ankara, Turkey*

Abstract

Online advertisement has generated 48.09 billion dollars in 2013 (IAB, 2014) in US sponsoring most of the free services and content on the Internet today. However, problems stemming from the rivalry to get a share of this high revenue can disrupt the system jeopardizing online privacy and security. There have been a lot of malicious cases where fraudsters have deployed methods such as malware or phishing to get a share from the market thus compromising the security of end users and thus harming the end users. In addition to this, tracking methods deployed to increase the targeting accuracy of the advertisements reveals a lot of private information about the users without any permission from them. The rules around collection and usage of personal data is ambiguous and can be harmful in various ways such when it is shared with third parties. In response to these threats, end users have been inclining to use ad-avoidance tools which disable user tracking or block ad requests completely. This situation harms the Internet ecosystem which heavily relies on online advertising. To remedy this problem, Comodo is developing secure and privacy-friendly advertising solutions around its "trusted ads" concept using methods such as anonymous user tracking, safe ad content delivery and continuous ad inspection (malware/virus scanning). These methods guarantee secure advertisements delivery to end users, while preserving effectiveness of ad delivery. The beta versions of the products have been deployed and tested in real life for 6 months. Based on our observations, users are concerned about their privacy; but they also want customized content according to their needs and make money from ads as long as it is not harmful to them. Based on this data, we are willing to expand our system and make it one of the major ad-delivery networks.

Peer-review under responsibility of Istanbul Univeristy.

Keywords: Secure Advertising; Privacy Protection; Anti-Malvertising; Anonymous Tracking; Ad Targeting; Ad Network

* Corresponding author. Tel.: +90-312-999-0044
E-mail address: mustafa.oger@comodo.com

Peer-review under responsibility of Istanbul Univeristy.
doi:10.1016/j.sbspro.2015.06.405

1. Introduction

Online advertising is a form of advertising which uses Internet to deliver promotional materials (ads) to users. It has different methods like email advertising, social media advertising, search engine advertising, mobile advertising and display advertising. Online advertising is one of the biggest businesses on the Internet. Considering the growth of the Internet and the advances achieved with mobile communication, it is expected to grow even more.

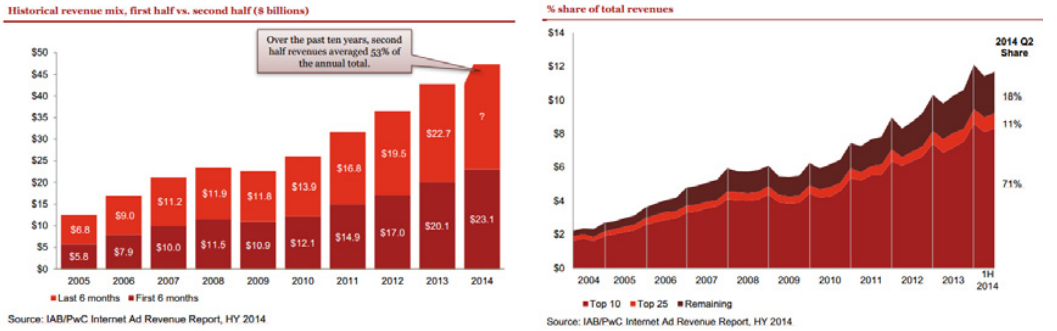

Fig. 1. (a) Historical ad revenues; (b) Market share of ad networks.

Fig. 1 (a) above shows the growth of online advertising market in US. Fig. 1 (b) shows the market share in US. Since the market volume and possible revenues are high; most of the powerful technology companies compete to get a bigger share. These figures solely indicate the magnitude of online advertising business.

The business model of online advertisement hides the details so that we think services are free of charge; however the maintenance of those online services and the continuity of the free content have a certain cost. The revenues generated from online advertisements cover the cost of these services. As a result, online advertising can be seen as the key that makes free Internet possible for end users.

Although online advertisement sponsors free Internet services, it also has some drawbacks for end users:

- It is possible to distribute malware with online ads and infect end users' computer. This attack is called malvertising and there are well known cases which many Internet users suffered.
- The ads can be guaranteed to be malware-free but the web page that users are redirected when they click on those ads can have potential threats. Users cannot know the outcome when they click on an ad in advance.
- The personal data is collected by nearly all online advertising companies without users' consent. Personal information can be used to abuse users or can be sold to third parties.
- Internet users don't have any control on what types of ads are displayed to them; so they can see an adult ad or a gambling ad on their favorite news webpage.
- Online advertising uses resources of end users, such as Internet bandwidth and computing power. This consumption causes decrease on quality of Internet experience for users such as increase in page load times.

Several researches have been conducted to better understand the attitude of end users towards online advertising. The following research (Zogby, 2013) as shown in Fig. 2 (a) shows that end users want to get free ad-supported content, while a small fraction is willing to pay for ad-free content. This infers the most organic, collaborative way of Internet. In addition, another poll shows that users also care about their privacy and security. It can also be seen from Fig. 2 (b) that though users want free content, they are still very concerned about their online privacy and security. This poll shows that tracking of personal activity and theft of personal data such as email address or credit card information are the most concerning topics for end users.

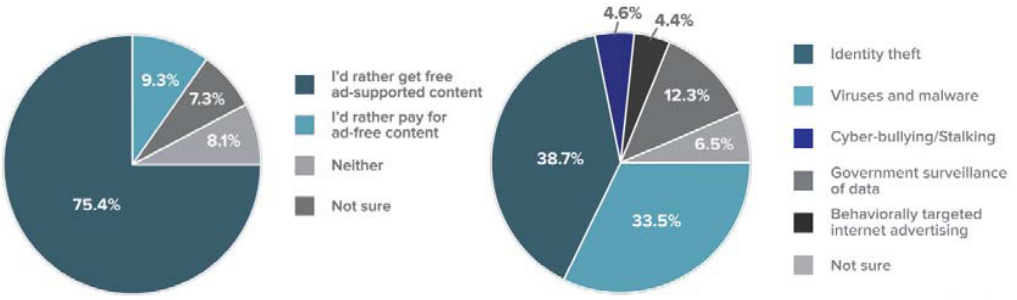

Fig. 2. (a) Poll for willingness to pay for content; (b) Poll for concerns about online advertising

Internet users who are victims of mentioned potential threats are looking for methods to protect themselves. As a result, ad blockers which are small software blocking all online advertising traffic, are being used by more users day by day. Even though these tools provide some degree of privacy, they are preventing revenue generation thus harming free Internet ecosystem. In other words, ad blocking with its drawbacks does not provide an effective solution.

This situation needs enhancements on multiple levels. Firstly, the system should guarantee secure advertisement delivery. Another enhancement can be outlined as integrating end user to ad delivery process as a key component. End users can decide whether an ad can be shown or not and can also decide how his/her personal information will be used once it is collected. This model assigns the end user as the decision maker and puts any usage of anonymous personal data under end user's control.

In accordance with aforementioned arguments, Comodo is taking steps to implement its "trusted ads" concept which ensures secure advertisement experience while preserving privacy of end users.

2. Methodology

Comodo aims to provide trusted ads that blocks malicious attacks and provide relevant ads to end user without intruding privacy of users. This will help to maintain free Internet services while providing additional gain of better targeted ads.

Secure advertisement is a crucial point in Comodo's "trusted ads" concept. To provide secure advertisements and avoid malvertising, we implemented two important features: Continuous ad inspection and continuous advertiser inspection. In continuous ad inspection, a well-defined approval process is applied for all ads in the system. When a new ad is injected into system or when any property changed of an existing ad, this approval process is triggered. During this approval process, ad content and properties are checked against malvertising and only safe ones are accepted. In continuous advertiser inspection, destination URLs of the ads are periodically checked and if any threat is detected respective ads are automatically blocked. With this method we ensure that users are protected when they clicked on an ad that is delivered by Comodo ad server.

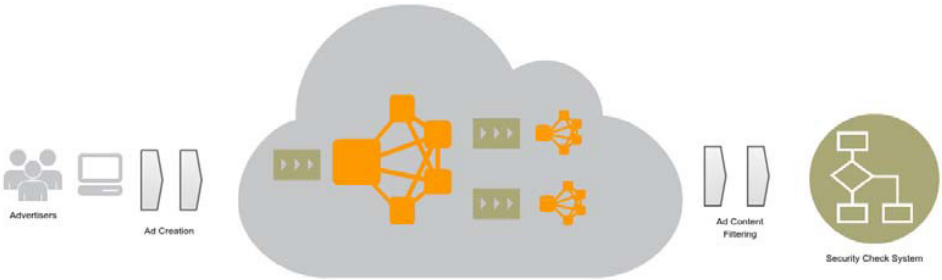

Fig. 3. Security check system

One of the least anticipated risks occur after an advertisement is created in the system. This case is generally overlooked and can cause severe damage on both ad network and publisher sides. As an example, in 2009 banner feed of The New York Times was hacked. According to spokeswoman, the culprit provided apparently legitimate ads for a week and then the ads were switched to the virus alert malvertisement (DailyFinance, 2009). To protect our system from similar threats, ad inspection processes run for each ad continuously as shown in Fig. 3.

User privacy is another crucial point in Comodo's "trusted ads" concept. Opt-in and opt-out choices are available to users and they are clearly stated to the users, along with the consequences of their choices. If the users opt in anonymous tracking, they can be tracked anonymously for better ad targeting. These users are identified by a unique identification number (UID) without storing any personal information like name or email address. In the case of opt-out, users can remove all their private data from the system. This method assists the system to protect user privacy, in addition to help keeping track of anonymous users.

3. Technology of Comodo Ad Server

Comodo Ad Server is a distributed system which will run on hundreds of servers. There are services provided on different servers communicating with each other directly or via data stores. To build such a system which is scalable, maintainable and high-performing, there are technical challenges that should be overcome. We briefly stated the challenges and our solutions below. Overview of the system architecture can be seen in Fig.4.

Fast Advertisement Delivery

Comodo ad server should deliver ads as fast as possible. This requirement exists to ensure that ad delivery does not hinder web experience for the end user and it should be fast enough leaving time for the user to engage with the ad content. Initially, MySQL database server was used to store advertisement inventory. A cluster of MySQL servers tried to handle massive incoming advertisement requests. However, this approach proved insufficient against high request rates. A new approach that performs and scales well is designed. ElasticSearch formed the base of the new approach, decreasing response times and increasing throughput multiple times.

Offline Data Analytics

Comodo ad server should have a capable data analytics system to gain insights about the business. Data analytics is the most valuable tool to analyze user base and its dynamics, it is used to analyze the performance of existing advertisement methods. Using these analyses, decision makers can develop new business strategies. Performance and efficiency and performance of these new business strategies are also measured by data analytics tools. This learn-build-measure cycle feeds itself and helps increase overall quality of Comodo ad server. To handle these requirements, Apache Hadoop and other big data technologies like Apache Pig, Hive, Oozie are used. Current system supports recurring data analytics pipelines, on-demand reports and ad-hoc queries.

Real-time Analytics and Fraud Detection

Comodo ad server should have real-time data analytics processes to handle the large amount of continuous data and make better use of it in many possible ways. In Comodo ad server, data generated by the system enters a real-time analytics pipeline. Every incoming data is categorized, aggregations are performed and results are persisted to data stores at scheduled times. Fraud detection is also handled by this system, where continuous click stream is investigated and fraudulent clicks are ruled out. Apache Kafka and Apache Storm are being used as main parts of this system. There are other complementary technologies like HDFS, Cassandra and ElasticSearch mainly used for storing outputs of analytics pipelines.

Monitoring

Comodo ad server should have a real-time, comprehensive monitoring infrastructure to evaluate the health of system, see the bottlenecks and help foresee possible problems. Since this will be a distributed system running on multiple servers, a problem in one part of the system may affect the whole system. To monitor health status of servers, Zabbix monitoring tool is used. Monitoring checks vary from hardware level checks to application level checks. To better monitor application level features, Kibana is being used along with ElasticSearch.

Data Stores

Comodo ad server has different data structures and processing needs across its components. To achieve this with one data store is just not feasible. Every component has different storage needs and has different access patterns to these storages. Redis, Cassandra and MySQL are examples to these data stores as seen on Fig.4.

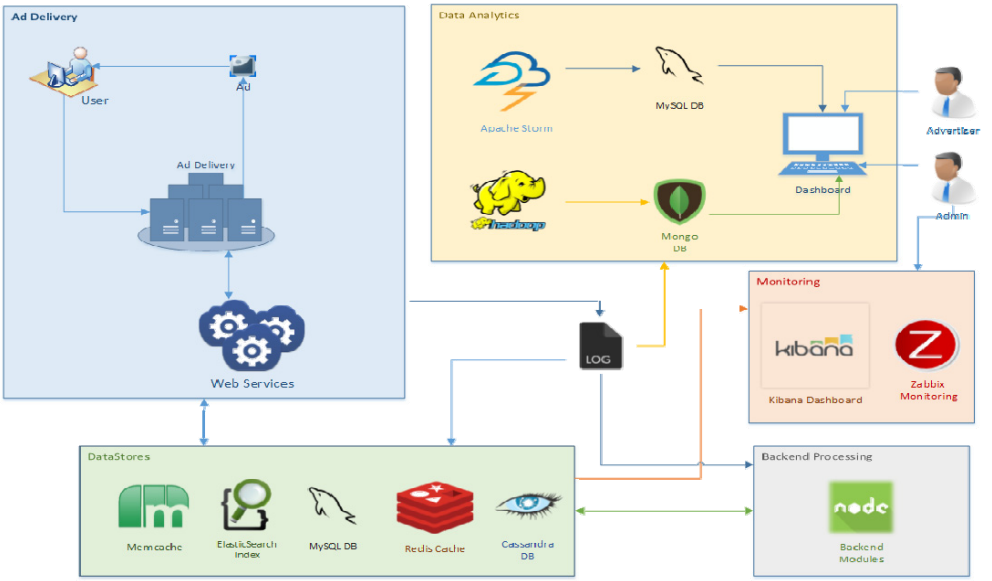

Fig.4. System architecture

4. Current Situation - Future Projections

Comodo ad server strives for delivering secure ads worldwide. Future projections about specific areas have been given to further analyze the future development.

Daily Impressions: In its current situation, Comodo ad server serves 20+ million impressions while handling 100+ million request. With its current scalability and capacity, it can handle up-to 150+ million impression per day. Mid-term target is to build up a system which can handle up-to 2 billion impression per day.

Delivery Speed: Comodo ad server can deliver a banner ad around 500ms while it can deliver a text ad around 400ms. With the continuation of architectural enhancements, in mid-term delivery speed will be under 300ms for both types of ads while keeping request handling capacity at the targeted level.

Retargeting Speed: Comodo ad server has strong retargeting capabilities, such as contextual retargeting, search retargeting, URL based retargeting. Currently when a retargeting trigger (search phrase, user visit to advertiser site etc.) was detected, it is processed and activated on the system under 10 seconds. Mid-term plan is to be able to detect and use the retargeting triggers under 5 seconds.

Data Processing Rate: Comodo Ad Server currently generates 25+ GB of structured raw data every day. This data is processed at real time with Apache Storm and periodically by scheduled Hadoop jobs feeding data analytics dashboards. Besides scheduled jobs, on-demand reports and ad-hoc queries also query data stores. Data processing capability of the system will continue to grow proportional to increase in ad delivery load.

5. Conclusion

Trust in online advertisement plays an important role in sustaining economy of Internet. Recent studies show that as end-users have privacy and security concerns about online advertisements, they begin to deploy counter measures to block advertisements. This situation threatens the Internet ecosystem. To remedy this problem, Comodo ad server is developing privacy friendly and secure advertising solutions. With our solution, we are expecting to see decrease in the usage of ad blocking tools.

References

DailyFinance. (2009, September). *www.dailyfinance.com*. Retrieved from www.dailyfinance.com: http://www.dailyfinance.com/2009/09/14/malvertising-hits-the-new-york-times/

IAB. (2014, October). *www.iab.net.* Retrieved from www.iab.net: http://www.iab.net/media/file/IAB_Internet_Advertising_Revenue_Report_HY_2014_PDF.pdf

Zogby. (2013, April). *www.aboutads.info.* Retrieved from www.aboutads.info: http://www.aboutads.info/resource/image/Poll/Zogby_DAA_Poll.pdf