Real Time Click Fraud Prevention using multi-level Data Fusion

Article · October 2010

Source: DOAJ

CITATIONS
8

READS
184

3 authors, including:

Chamila Walgampaya
University of Peradeniya
31 PUBLICATIONS **86** CITATIONS
SEE PROFILE

Roman Yampolskiy
University of Louisville
231 PUBLICATIONS **3,317** CITATIONS
SEE PROFILE

Real Time Click Fraud Prevention using multi-level Data Fusion

Chamila Walgampaya, Mehmed Kantardzic, Roman Yampolskiy

***Abstract*—from the viewpoint of Dempster-Shafer evidence theory, information obtained from different sources can be considered as pieces of evidence, and as such, multi-sensor based CCFDP (Collaborative Click Fraud Detection and Prevention) system can be viewed as a problem of evidence fusion. In this paper we detail the multi level data fusion mechanism used in CCFDP for real time click fraud detection and prevention. Prevention mechanisms are based on blocking suspicious traffic by IP, referrer, city, country, ISP, etc. Our system maintains an online database of these suspicious parameters. We have tested the system with real world data from an actual ad campaign where the results show that use of multi-level data fusion improves the quality of click fraud analysis.**

***Index Terms*—Click Fraud Detection and Prevention, Information integration on the Web, Sensor Fusion.**

I. Introduction

Web search is a fundamental technology for navigating the Internet and it provides access to information for millions of users per day. Internet search engine companies, such as Google, Yahoo, and MSN have revolutionized not only the use of the Internet by individuals but also the way businesses advertise to consumers [11, 16]. Typical search engine queries are short and reveal a great deal of information about user preferences. This gives search engine companies a unique opportunity to display highly targeted ads to the user. These search services are expensive to maintain and depend upon advertisement revenue to remain free [16] for the end user. Many search service companies such as Google, Yahoo and MSN generate advertisement revenue by selling clicks. This business model is known as Pay-Per-Click (PPC) model.

In the PPC model, internet content providers are paid for each time an advertisement link on their website is clicked leading to the sponsoring company's content. There is an incentive for dishonest service providers to inflate the number of clicks their sites generate. In addition, dishonest advertisers tend to simulate clicks on the advertisements of their competitors to deplete their advertising budgets [17]. Generation of such invalid clicks either by humans or software with the intension to make money or deplete competitor's budget is known as click fraud (CF).

The diversity of CF attack types makes it hard for a single counter measure to attain desired results. Therefore, it becomes one of the new hot spots in research how to combine multiple data sources with multiple measures to provide the PPC system with more effective protection from CF. A real time click fraud detection and prevention system based on multi-model and multi-level data fusion is proposed in this paper. Each independent component can be considered as an invisible data mining module, in which "smart" software incorporates data mining into its functional components, often unbeknownst to the user [10]. Evidence for CF from multiple models are "fused" in this system, using Dempster-Shafer evidence theory [21], so that it achieves improved accuracy for detecting fraudulent traffic. Conversely it increases the quality of clicks reaching advertisers' websites.

This paper is organized as follows. In section II an introduction to the multi-sensor data fusion and Dempster-Shafer evidence theory is given. Related work, where Dempster-Shafer theory was used as the fusion mechanism, is presented in Section III. In section IV the fusion architecture of the CCFDP system is described, while a case study is explained in section V. Experimental results and discussion are given in section VI. Conclusions are given in section VII.

Manuscript received July 7, 2010. Chamila Walgampaya, Mehmed Kantardzic, and Roman Yampolskiy are with the Computer Engineering and Computer Science department at the University of Louisville, Louisville, KY 40292, USA. (phone: 502-852-3626; fax: 502-852-4713; e-mail: ckwalg01, mmkant01, roman.yampolskiy@ louisville.edu).

II. Multi-Sensor Data Fusion with Dempster-Shafer Evidence Theory

Data fusion is "a process dealing with the association, correlation, and combination of data and information from single and multiple sources to achieve refined position and identity estimates, and complete and timely assessments of situations and threats, and their significance" [15]. The resulting information is more satisfactory to the user when fusion is performed than simply delivering the raw data [24].

Different fusion methods are discussed in literature, such as statistical estimation [6, 9], Kalman filter [28], fuzzy integration [22], neutral networks [5], D-S evidence theory [27] and so on. Of these fusion methods, D-S evidence theory is widely known for better handling uncertainties. Moreover, it provides flexible information processing and can deal with asynchronous information [19].

In the following section, terminology of theory of evidence and the notation used in this paper are defined.

I. *Frame of discernment*: If Θ denotes the set of $\theta_N(\theta_N \in \Theta)$ corresponding to N identifiable objects, let $\Theta = \theta_1, \theta_2, \dots, \theta_N$ be a frame of discernment. The power set of Θ is the set containing all 2^N possible subsets of Θ, represented by $P(\Theta) = \{\phi, \{\theta_1\}, \{\theta_2\}, \dots, \{\theta_N\}, \{\theta_1, \theta_2\}, \{\theta_1, \theta_3\}, \dots, \Theta\}$ where ϕ denotes the null set.

ISBN: 978-988-17012-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

ii. *Basic Probability Assignment function* (BPA): The BPA is a primitive of evidence theory. The BPA, represented by m, defines a mapping of the power set to the interval between 0 and 1, where the BPA of the null set is 0 and the summation of the BPA's of all the subsets of the power set is 1. The value of the BPA for a given set A, represented as $m(A)$, expresses the proportion of all relevant and available evidence that supports the claim that a particular element of Θ belongs to the set A but to no particular subset of A. The elements of $P(\Theta)$ that have none-zero mass are called focal elements. Formally, this description of m can be represented with the following three equations:

$$m: P(\Theta) \Longrightarrow [0,1] \quad (1)$$
$$\sum_{A \in P(\Theta)} m(A) = 1 \quad (2)$$
$$m(\emptyset) = 0 \quad (3)$$

iii. *Belief function* Bel(A) : Given a BPA m, a belief function Bel is defined as:

$$Bel(A) = \sum_{B \subseteq A} m(B) \quad (4)$$

The belief function $Bel(A)$ measures the total amount of probability that must be distributed among the elements of A.

iv. *Combination of rule of evidence* m(C): Supposed $m_1$and m_2 are two mass functions formed based on information obtained from two different information sources in the same frame of discernment; according to Dempster's orthogonal rule we define $m(C) = (m_1 \oplus m_2)(C)$

if $C = \emptyset$

$$(m_1 \oplus m_2)(C) = 0$$

else

$$(m_1 \oplus m_2)(C) = \frac{\sum_{A \cap B = C} m_1(A) m_2(B)}{1 - K} \quad (5)$$

Where K represents basic probability mass associated with conflict defined as:

$$K = \sum_{A \cap B \neq \emptyset} m_1(A) m_2(B) < 1 \quad (6)$$

In our system, evidence supports a click to either be valid or invalid. Therefore it becomes a two class problem. Accordingly we have modified the calculation of $m(C)$ for the CCFDP system [18]. For a two class problem, we can simplify the equation for combination of evidence to:

$$S = \frac{\prod_{i=1,n} r_i}{\prod_{i=1,n} r_i + \prod_{i=1,n} (1 - r_i)} \quad (7)$$

Where r_i *is* the output from each is model and n is the number of models.

III. RELATED WORK

The Dempster-Shafer theory of evidence reasoning (D-S theory) has been widely discussed and used recently, because it is a reasonable, convenient, and promising method to combine uncertain information from disparate sources with different levels of abstraction. Carvalho [3] et al. proposed a general Data Fusion Architecture (DFA) based on Unified Modeling language (UML) and using a taxonomy based on the definitions of raw data and variables or tasks. Their DFA can be reconfigured according to the measured environment and availability of the sensing units or data sources, providing a graceful degradation in the view of the environment as resources change.

Clerentin [4] et al. has applied the D-S theory to study the cooperation between two omni-directional perception systems for mobile robot localization. In this paper, an absolute localization paradigm based on the cooperation of an omni-directional version system composed of a conical mirror and a CCD camera and a low cost panoramic range finder system is reported. Authors presented the absolute localization method that uses three matching criteria fused by the combination rules of the D-S theory.

Distributed databases allow us to integrate data from different sources which have not previously been combined. The Dempster-Shafer theory of evidence and evidential reasoning are particularly well suited to the integration of distributed databases. Cai et al. have studied the suitability of evidential functions to represent evidence from different sources. They have carried out evidential reasoning by the well-known orthogonal sum method [2].

In their article, Janez and Goretta [12] et al. present a strategy to report in automatic way significant changes on a map by fusion of recent images in various spectral bands. They have shown that D-S theory as a more suitable formalism for configurations of partial overlapping between map and images, which may be difficult or even impossible to formalize the approach suggested within a probability framework.

Tian et al. have described the use of D-S evidence theory and its data fusion technology in their intrusion detection model. This model merges alerts from different intrusion detection systems, makes intelligent inference by applying D-S evidence theory, and estimates the current security situation according to the fusion result [23].

The military typically operates in demanding, dynamic, semi-structured and large-scale environments. This reality makes it difficult to detect, track, recognize/ classify, and response to all entities within the volume of interest, thus increasing the risk of late response to the ones that pose actual threat. Benaskeur et al. proposed an Adaptive Data Fusion and Sensor Management information gathering and fusion process by automatically allocating, controlling, and coordinating the sensing and the processing resources to meet mission requirements [1].

IV. FUSION OF EVIDENCES OF CLICK FRAUD IN THE CCFDP SYSTEM

The collaborative click fraud detection and prevention (CCFDP) system was developed to collect data about each click, involving the data fusion between client side log and server side log [8]. In CCFDP there are three modules that contribute to the process of finding fraudulent clicks. They are rule based module, click map module, and outlier detection module. In each of these modules, output is a probabilistic measure of evidence for the click being fraudulent. Authors have discussed the functionality of each of these modules in detail before [13, 14]. In addition, CCFDP maintains an online fraudulent database of suspicious sources of clicks in terms of IP, referrer, country

ISBN: 978-988-17012-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

etc. When the score of an IP or a country etc. reaches a predefine threshold value the CCFDP system moves it to the online fraudulent database and inform the service providers with the instructions to block future traffic originating from these sources. Scores for each parameter are updated after a click found suspicious based on the combined evidences of the modules that we mentioned above.

The model-driven fusion process of CCFDP is depicted in Fig. 1. Real-time data feeds from three sources (k=3): server side, client side, and extended context of the click (S1, S2, S3). This is represented by sensors in Fig. 1. In the data preprocessing stage we standardize (align) the input data [25]. The concept of alignment is an integral part of the fusion process, and assumes "common language" between the inputs and includes the standardization of measurement units. The scores from the rules based module (DM model 1), outlier detection module (DM model 2), and click map module (DM model 3) are then combined using D-S evidence theory at the decision level (m=3). The combination of scores will be used to dynamically adjust advertising profiles in such a way that low quality sources of traffic will no longer be shown advertisements.

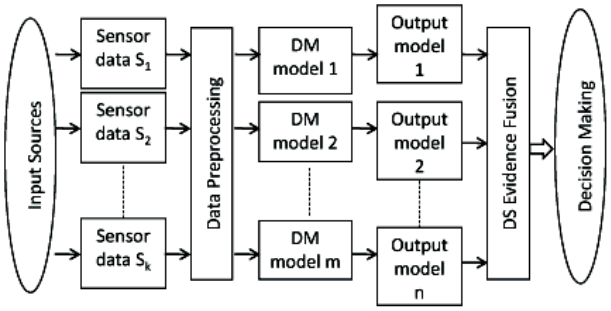

Fig.1: Model-driven fusion process of CCFDP

V. A CASE STUDY

In this section, we demonstrate the application of D-S evidence theory to combine evidences of sources.

Evidence 1: Repeated clicks from IP during past minute detected by the rule based module. (m_1)
Evidence 2: Java Script is allowed in the browser detected by the rule based module. (m_2)
Evidence 3: Country Morocco is detected suspicious by outlier module. (m_3)
In the two classes Fraud is represented by $\{F\}$ and non-Fraud is represented by $\{N\}$. Let $\Theta = \{F, N\}$, We define the power set and Basic Probability assignments as follows:

$$P(\Theta) = \{\phi, \{F\}, \{N\}\}$$
$$m_1(\phi) = 0, m_1(\{F\}) = 0.6, m_1(\{N\}) = 0.4$$
$$m_2(\phi) = 0, m_2(\{F\}) = 0.5, m_2(\{N\}) = 0.5$$
$$m_3(\phi) = 0, m_3(\{F\}) = 0.7, m_3(\{N\}) = 0.3$$

Calculation of $M_1 \oplus M_2$
For the convenience we use the fusion tables, introduced by Shafer in [21], to show the calculations. Fusion tables are given in Table 1 and Table 2.

Using equation 6: $K = 0.5 * 0.6 + 0.4 * 0.5 = 0.5$

Table 1: Fusion of Evidence 1 and Evidence 2

$(M_1 \oplus M_2)$	{F}0.5	{N}0.5
{F}0.6	{F}0.3	Φ0.3
{N}0.4	Φ0.2	{N}0.2

New Belief function

$$Bel_{m_1 \oplus m_2}(\{F\}) = \sum_{B \subseteq \{F\}} m(B) = 0 + 0.6 = 0.6$$

Calculation of $M_1 \oplus M_2 \oplus M_3$

Table 2: Fusion of Evidences 1,2 and Evidence 3

$(M_1 \oplus M_2 \oplus M_3)$	{F}0.7	{N}0.3
{F}0.3	{F}0.21	Φ0.09
Φ0.3	Φ0.21	Φ0.09
Φ0.2	Φ0.14	Φ0.06
{N}0.2	Φ0.14	{N}0.06

Using equation 6:
$K = 0.09 + 0.21 + 0.09 + 0.14 + 0.06 + 0.14 = 0.78$
New Belief function

$$Bel_{m_1 \oplus m_2 \oplus m_3}(\{F\}) = \sum_{B \subseteq \{F\}} m(B) = 0 + 0.78 = 0.78$$

In this example we considered the local suspicious scores of 0.6, 0.5, and 0.7. D-S evidence theory is used to find the final evidence. The belief that the click is fraudulent is 0.78.

VI. EXPERIMENTAL RESULTS AND DISCUSSION

The real time version of CCFDP is now available online at http://www.netmosaics.com. All of our experiments use click data from Hosting.com and thebestmusicsites.org websites. The process was started on January 7th, 2007 and is still in collecting data. As of June 30th, 2008 we have collected around 1,400,000 natural and 25,000 paid click data.

Initial version of CCFDP was designed using only a rule based system. The new CCFDP has outlier module and the click map module in addition to an improved rule based system with additional click context information. Experiments are performed on both old and new versions of CCFDP. In this research, initial experiments are conducted to observe and compare the changes in the scores of parameters such as IP, country, and referrer in both systems.

After all paid click data has been processed we have selected the top 10 IPs, countries, and referrers with the highest fraudulent scores to see if the fusion process has any effect on updating individual scores of these parameters. Tables 3 and 4 list the IPs, countries, and referrers that have the highest fraudulent scores respectively. The results are slightly modified to protect privacy of some publisher websites. For example the actual domain names and referrer names are replaced with dummy identifiers.

In Fig. 2 (top) the variation of scores for IPs are depicted. Except for one IP address (136.165.67.74) all others have higher fraudulent scores after combining the evidences from all the modules. In the rule based system, evidence is collected by considering only the changes detected in a limited neighborhood. For example with only the rule based system, it will be difficult to detect a Bot associated to a particular IP which sends http requests in the time intervals

ISBN: 978-988-17012-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

greater than 15 minutes. But with the outlier detection module that covers larger neighborhood1 of the clicks, the pattern becomes observable. Once a suspicious activity is detected this evidence will contribute to increase of corresponding partial scores in the CCFDP system. IP address with higher scores has increased probability of being blacklisted sooner.

Table 3: Top IP and Country Counts

Top IP Count		Top Country Count	
IP	Count	Country	Count
71.235.26.170	122	US	19784
68.88.239.191	112	IN	1278
136.165.67.74	94	CA	856
199.231.146.254	86	GB	666
89.139.234.179	82	NULL	574
203.162.3.146	80	MX	544
170.20.96.116	80	AU	534
71.193.114.12	72	TR	518
74.133.47.66	68	BR	506
74.192.144.103	68	PH	456

Table 4: Top Referrer Counts

Referrer	Count
No referrer (NULL)	8800
http://www.r1.com/	4568
http://www.r2.com/	2192
http://www.r3.com/	604
http://www.r4.com/	546
http://www.r5.com/	538
http://www.r6.com/	510
http://www.r7.com/	450
http://www.r8.com/	420
http://www.r9.com/	414

Once the IP addresses are on the blacklist the search provider will be notified to eliminate future traffic from the corresponding sources. This will improve the quality of the traffic redirected to the advertiser's website.

One of the biggest advantages of using a multi-model system in CCFDP is its ability to cover wider area in the time domain. While the rule based module deals with events within couple of minutes of each other the outlier detection module handles events in a 24 hour window. Fig. 2 (center) shows the final scores of top 10 countries from which we have received most of the traffic. With the rule based module alone we were unable to detect patterns and variations in the time axis. Therefore almost all countries have a score less than 0.1, which implies clicks from these countries are not suspicious at all. But with the outlier module, which keeps track of traffic for extended period of time, we were able to detect abnormal traffic from most of the countries. For example some of these countries send traffic only during certain hours of the day.

A similar behavior is observed with the top referrers of traffic to hosting.com site. Fig. 2 (bottom) shows the variation of scores of top 10 referrers. All these referrers appear normal when they are evaluated only with the rule based system. But when they are evaluated together with click map module and the outlier detection module referrer scores were drastically increased. Some of these referrers are from outside the US. When the countries suspicion score increases so does the scores of associated referrers. For example we mentioned in the above example that certain countries send traffic only in certain hours of the day. When we include the click context it is observed that most of these referrers are associated with those countries. This behavior will be very hard to detect if we are using only the rule based score.

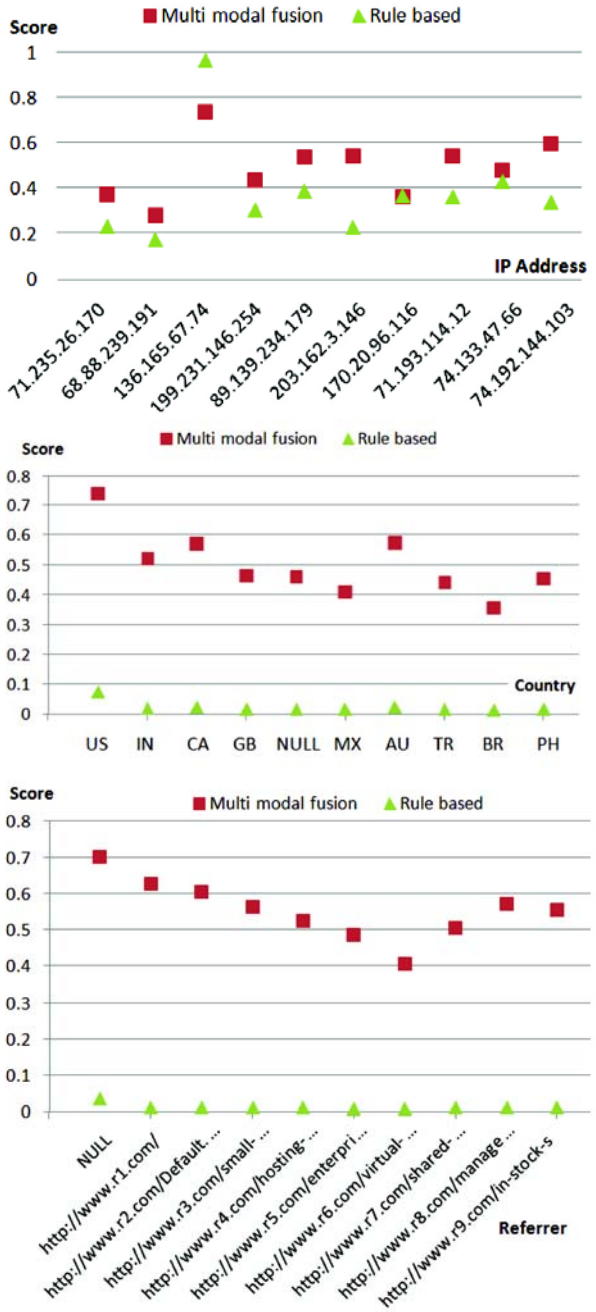

Fig.2: Variation of IP Score (top), Country Score (center), and Referrer Score (bottom)

In traditional system (rule based) country and referrer did not influence on the score almost at all. Inclusion of additional modules make country score and referrer score become much more sensitive. For example the new system

ISBN: 978-988-17012-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

includes country parameter in 73% of clicks from US in the final score.

Table 5: Distribution of clicks in each region in Fig.3

	I	II	III	IV
Rule based system	12198	1	3	520
Multi modal system	4197	4650	3817	643

Fig. 3 shows the distribution of final scores for all the clicks with two versions of CCFDP. The lighter graph (L) corresponds to the first version of CCFDP where only rule based module was used. The darker one (D) is the new version with multiple modules. Area I represents most of the valid clicks. This corresponds to the records with attributes which do not have presence in the fraudulent database and all key attributes satisfies the requirements defined in the algorithm to be a legitimate click. The percentage of traffic present in Area I with system L is much higher than that of system D. With the inclusion of multiple models the suspiciousness of clicks has increased and the graph is shifted to the Area II with the system D, which is still in the safer region. Area III shows the suspected clicks. These are records with the attributes present in the fraudulent database or attributes that exceed certain threshold values. It can be clearly seen in the graph how the scores have increased after fusing multiple pieces of evidence from different modules. Area IV includes invalid clicks. Blocked traffic is identified as clicks with highly suspicious scores usually greater than 0.9. As shown in Table 5 with the traditional system (rule based system) we were able to block only 520 fraudulent clicks but with the multi model system it was 643, which is about 24% additional clicks. We believe that advertisers should not be billed for any of these clicks.

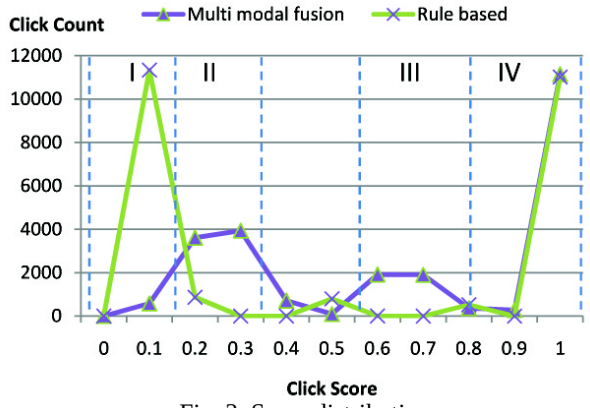

Fig. 3: Score distribution

Fig. 4 shows the percentage participation of each module in the final score calculation. Remember that the click map module is already used as a screening module to filter invalid clicks, where mouse clicks are recorded off-positioned to the advertisement. Light area of Fig. 4 represents the rule based module participation and dark area represents the outlier detection module participation.

We looked at the changes in quality of traffic after implementing the multi-model based CCFDP system. A summarized version is depicted in Fig. 5. The dataset was used in the rule based module alone and found that the average 53% of traffic is suspicious [13]. When running the outlier detection module alone on the dataset we discovered that about 34.6% of all clicks had one or more attributes that were found to have an outlying attribute-value pair count [14].

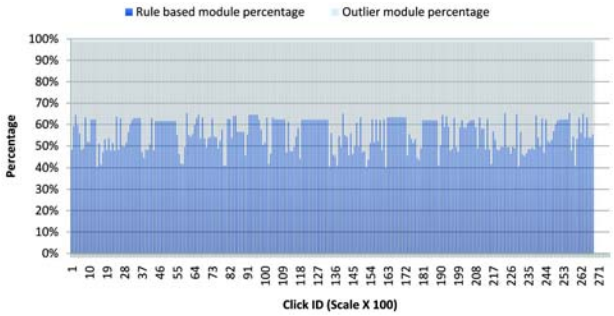

Fig. 4: Percentage Participation

Clicks found to have an outlier will contribute evidence effecting partial scores. The CCFDP system will compute the final score measuring suspicion for each click. And with the multi-model system classified about 64% of paid traffic as fraudulent. In addition we have observed the changes in the online fraudulent database. In the traditional system, only with rule base, the fraudulent database has recorded 71 IPs as fraudulent. The multi model system recorded 283 IPs as fraudulent with the same data set, which is nearly 4 times more than the traditional system. This is a greater improvement in terms of prevention of fraudulent traffic. As we discussed in Fig. 2, the traditional system has very little effect on country score and referrer score when calculating the total score. But with the multi model system scores for countries such as India, Morocco, and Mexico have shown enough suspicion. Clicks came from these countries received a higher fraudulent score but the system did not have enough suspicious clicks to block any of the countries completely. Similar results are observed for referrers.

We defined the quality of traffic as (1-score). Using only the rule based module and the outlier module we have about 47% and 65% quality scores respectively. With the combined model we were able to get much better traffic with about 36% of quality. With these results we can see that the multi model based CCFDP system is capable of improving the detection of fraudulent traffic at least by 10% compared to same models working alone.

Fig. 5: Improvement of quality of traffic

In addition we analyzed the volume of total clicks from Google and its partner network during these two periods of time.

ISBN: 978-988-17012-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

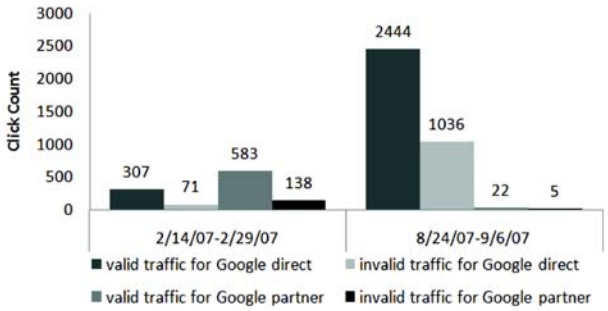

Fig. 6: Traffic analysis for Google

Fig. 6 shows the total and invalid traffics from Google and Google partner networks for the second month and the eighth month. First thing to observe is there is much higher volume of traffic in the eighth month compared to the second month. Second thing to observe is that traffic from Google partner networks in the eighth month is almost negligible. In the second month out of 307 direct Google referrals 71 are observed invalid, while 138 of 583 Google partner network referrals are detected invalid. In the eighth month total Google only traffic is 2444 and nearly 50% (1036) of that traffic is found to be invalid.

VII. CONCLUSIONS

In this paper we proposed a multi-model real time detection and prevention system for click fraud. The CCFDP system uses multi-level data fusion to enhance the description of each click, and to obtain better estimation of a click traffic quality. The CCFDP system analyzes the detailed user activities on both, server side and client side collaboratively to better evaluate the quality of clicks. Extended click record includes also context data available in fraudulent and blocking databases. Our system analyzes the extended click record using three independent data mining modules: rule based, outlier and click map. A score is assigned to each click based on the individual scores estimated by independent modules. Scores are combined using the Dempster-Shafer evidence theory. We have tested the system with a data from actual ad campaign in 2007 and 2008. Results show that higher percentage of click fraud is present even with most popular search engines such as Google. The multi-model based CCFDP estimated the average score as 64% where the 53% is the highest average score recorded by any individual module that ran the data alone. By these additional refinements we were also able to increase the quality of the click traffic by 10%.

REFERENCES

[1] A. R. Benaskeur and F. Rheaume. “Adaptive data fusion and sensor management for military applications,” *Aerospace Science and Technology*, 11(4):327 – 338, 2007.
[2] D. Cai, M. McTear, and S. McClean. “Knowledge discovery in distributed databases using evidence theory,” *International Journal of intelligent systems*, 15(8):745–761, 2000.
[3] H. Carvalho, W. Heinzelman, A. Murphy, and C. Coelho. “A general data fusion architecture,” *In International Conference on Information Fusion*, pages 1465–1472, 2003.
[4] A. Clerentin, L. Delahoche, and E. Brassart. “Cooperation between two omnidirectional perception systems for mobile robot localization,” *In Proceedings of the 2000 IEEE/RSJ Intemational Conference on Intelligent Robots and Systems*, pages 1499–1504, 2000.
[5] X. Dai and S. Khorram. “Data fusion using artificial neural networks: a case study on multitemporal change analysis,” *Computers, Environment and Urban Systems*, 23(1):19–31, 1999.
[6] H. Durrant-Whyte. “Integration, coordination and control of multi-sensor robot systems,” *Dissertation Abstracts International*, 47(10), 1987.
[7] Z. Fusheng and D. Feng. “Application of DS evidence theory in flow regime identification of two-phase horizontal pipe flow,” *In 27th Chinese Control Conference*, pages 758–762, 2008.
[8] L. Ge and M. Kantardzic. “Real-time click fraud detecting and blocking system,” *US Patent App. 11/413,983*, May 1 2006.
[9] G. Hager, S. Engelson, and S. Atiya. “On comparing statistical and set-based methods in sensor data fusion,” *In IEEE International Conference on Robot Automation*, 1993.
[10] J. Han and M. Kamber. “Data mining: concepts and techniques,” *Morgan Kaufmann*, 2006.
[11] N. Immorlica, K. Jain, M. Mahdian, and K. Talwar. “Click fraud resistant methods for learning click-through rates,” *Lecture Notes In Computer Science*, 3828:34–45, 2005.
[12] F. Janez, O. Goretta, and A. Michel. “Automatic map updating by fusion of multispectral images in the Dempster-Shafer framework,” *In Proceedings of SPIE*, volume 4115, page 245, 2000.
[13] M. Kantardzic, C. Walgampaya, B. Wenerstrom, O. Lozitskiy, S. Higgins, and D. King. “Improving Click Fraud Detection by Real Time Data Fusion,” *In IEEE International Symposium on Signal Processing and Information Technology*, 2008. ISSPIT 2008, pages 69–74, 2008.
[14] M. Kantardzic, B. Wenerstrom, C. Walgampaya, O. Lozitskiy, S. Higgins, and D. King. “Time and Space Contextual Information Improves Click Quality Estimation,” *e-Commerce 2009*, page 123, 2009.
[15] D. Lambert. “A blueprint for higher-level fusion systems,” *Information Fusion*, 10(1):6–24, 2009.
[16] M. Mahdian. “Theoretical challenges in the design of advertisement auctions,” *In The Capital Area Theory Symposia*. University of Maryland, Spring 2006.
[17] A. Metwally, D. Agrawal, and A. El Abbadi. “Detectives: detecting coalition hit inflation attacks in advertising networks streams,” *In Proceedings of the 16th international conference on World Wide Web*, ACM, 2007.
[18] NetMosaics. *NetMosaics Inc. Internal Documentation*, 2009.
[19] N. Ouyang, Z. Liu, and H. Kang. “A method of Distributed Decision Fusion based on SVM and DS evidence theory,” *In 5th International Conference on Visual Information Engineering*, pages 261–264, 2008.
[20] U. S. DOD, Data fusion lexicon, Data Fusion Subpanel of the Joint Directors of Laboratories, Technical Panel for C3, Environmental Research Inst. Of Michigan Arlington VA, 1991.
[21] G. Shafer. “A mathematical theory of evidence,” *Princeton university press Princeton*, NJ, 1976.
[22] B. Solaiman, L. Pierce, and F. Ulaby. “Multisensor data fusion using fuzzy concepts: application to land-cover classification using ERS-1/JERS-1 SAR composites,” *IEEE Transactions on Geoscience and Remote Sensing*, 37(3):1316–1326, 1999.
[23] J. Tian, W. Zhao, R. Du, and Z. Zhang. “DS evidence theory and its data fusion application in intrusion detection,” *Lecture notes in computer science*, 3802, 2005.
[24] L. Wald. “The present achievements of the EARSeL-SIG’Data Fusion, In A Decade of Trans-European Remote Sensing Cooperation,” *Proceedings of the 20th Earsel Symposium*, Dresden, Germany, 14-16 June 2000, page 263. Taylor &Francis, 2001.
[25] E. Waltz. “Information understanding: integrating data fusion and data mining processes,” *In IEEE International Symposium on Circuits and Systems*, pages 553–556, 1998.
[26] E. Waltz and J. Llinas. “Multisensor data fusion,” *Artech House*, Boston, London, 1990.
[27] H. Wu, M. Siegel, R. Stiefelhagen, and J. Yang. “Sensor fusion using Dempster-Shafer theory,” *In IEEE Instrumentation and Measurement Technology Conference Proceedings*, volume 1, pages 7–12, 2002.
[28] C. Yukun, S. Xicai, and L. Zhigang. “Research on Kalman-filter based multisensor data fusion,” *Journal of Systems Engineering and Electronics*, 18(3):497–502, 2007.
[29] C. Yukun, S. Xicai, and L. Zhigang. “A Data Fusion Based Intrusion Detection Model,” *First International Workshop on Education Technology and Computer Science*, 1:1017–1021, 2009.
[30] D. Zeng and J. Xu. “Data Fusion for Traffic Incident Detection Using D-S Evidence Theory with Probabilistic SVMs,” *Journal of Computers*, 3(10):36–43, 2008.

ISBN: 978-988-17012-0-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)