# Security Games in Online Advertising: Can Ads Help Secure the Web?

Nevena Vratonjic, Jean-Pierre Hubaux, Maxim Raya
*School of Computer and Communication Sciences*
*EPFL, Switzerland*
*firstname.lastname@epfl.ch*

David C. Parkes
*School of Engineering and Applied Science*
*Harvard University, USA*
*parkes@eecs.harvard.edu*

## Abstract

*Some ISPs are trying to become part of the online advertising market. Such ISPs either: (i) cooperate with online advertising entities (e.g., ad networks) by providing users' private information to achieve better ad targeting in exchange for a share of the revenue, or (ii) modify the ad traffic on-the-fly such that they divert part of the online advertising revenue for themselves. This is a very important issue because online advertising is at the core of today's business model and it fuels many "free" applications and services. We study this behavior using game theory to model the interactions between ISPs and ad networks, and we analyze the effects on the Web caused by ISPs taking part in online advertising. Our results show that if the users' private information can improve ad targeting significantly and if ad networks do not have to pay a high share of revenue to the ISPs, ad networks and ISPs will cooperate to jointly provide targeted online ads. Otherwise, ISPs will divert part of the online ad revenue for themselves. In that case, if the diverted revenue is small, ad networks will not react. However, if their revenue loss is significant, the ad networks will invest into improving the security of the Web and protecting their ad revenue.*

## 1 Introduction

The traditional role of ISPs is to provide Internet access to end users. ISPs are supposed to provide this service by only faithfully forwarding end users' communication, in compliance with the Network Neutrality Policy [1]. Recently, several cases of ISPs meddling with users' traffic and violating the Network Neutrality policy have been reported [2] [3] [4]. Reis et al. [5] show that more than $1\%$ of Internet traffic is modified on-the-fly between web servers and end users. The majority of the modifications are performed on the ad traffic (e.g., ad injection, ad blocking) by ISPs.

Due to their topological position between end users and the Internet, ISPs can observe all the traffic of their end users. Based on the observed traffic, ISPs can extract users' private information, their preferences and interests, and can profile their online behavior. In the EU, to comply with data retention legislations [6] [7], ISPs have to obtain and keep records of their users' activities for a period between six months and two years, and upon request provide them to law enforcement agencies. This directive has imposed a significant burden on ISPs as it increases their storage costs and it requires investing into new technologies for packet inspection (e.g., Deep Packet Inspection [8]). There is no clear answer on how ISPs will obtain a return on that investment.

One possibility for ISPs to generate additional revenue is to take part in the online advertising business. Online advertising is the main business model on the Web today and it generates huge revenues (e.g., $\$23.4$ billion in the US in $2008$ [9]). However, ISPs are not part of the traditional online advertising systems. The online advertising revenue model includes ad networks (e.g., the Google ad network), advertisers and web publishers. In this revenue model, ISPs are bypassed because the only service they provide is to forward the traffic to and from end users. Hence, ISPs might be tempted by the high online advertising revenues and might try to become participants in

the online advertising business, especially because the user information in their possession could have high commercial value (e.g., due to its unavailability to other online entities). According to observed cases in practice, the behavior of ISPs can be either *cooperative* or *non-cooperative*.

A *cooperative* ISP collects and provides information about users' online behavior with the goal of improving ad targeting. This rich data about users can help better matching ads to users' interests, resulting in higher click-through rates on ads and consequently increasing the ad revenue [10]. Cooperative ISPs generate revenue by charging ad networks for user profiles. There are several examples in practice of ISPs that shared their users' data with ad companies (e.g., Phorm [11]), despite many concerns about the users' privacy [12].

A *non-cooperative* ISP diverts part of online advertising revenues for its own benefit by performing some of the attacks described in [4] [13]. For example, it injects ads into the content of web pages on-the-fly [5] or replaces legitimate ads with its own [4].

To the best of our knowledge, this is the first in-depth quantitative analysis of ISPs becoming strategic in the online advertising business. We study the effect of strategic ISPs on the Web using game theory as a tool to analyze mutually dependent actions of ISPs and the current participating entities in online advertising systems (e.g., ad networks). Our analysis shows that the outcome of the game between ISPs and ad networks mostly depends on: (i) the value of the users' private information and (ii) the share of the revenue that ad networks offer to ISPs. If the collected users' private information improves ad targeting significantly and ad networks do not have to pay a high price for it to the ISPs, the latter tend to be *cooperative* and they improve the quality of ad targeting jointly with ad networks. Otherwise, ISPs tend to be *non-cooperative*. Non-cooperative ISPs can divert a very small fraction of clicks from all the websites without causing any reaction from ad networks. However, if ISPs become greedy and divert a high fraction of clicks, ad networks will secure the high value websites first (by paying for SSL certificates and thus enabling the use of HTTPS), i.e., the websites that generate high volumes of clicks on their web pages. This means that the significance of the threat creates incentives for ad networks to protect their ad revenues, which could result in improved web security. Improved web security would not only benefit ad networks, but websites and users as well, because the security of all the online content, not only ads, would be improved. The results also show that ISPs will probably never try to divert a very high fraction of clicks from very popular websites, as that would cause a higher loss for ad networks, which would then secure the websites and prevent ISPs from obtaining any revenue from those websites.

The rest of the paper is organized as follows. After a brief presentation of the state of the art in Section 2, we present the system model in Section 3 and the various threats and countermeasures in Section 4. We present a game-theoretic model with two players, the ISP and the ad network and identify equilibrium outcomes of that game in Section 5. In Section 6, we provide further analytical refinements of our model and a numerical example to study the practical impact of the obtained results in Section 7. We conclude the paper in Section 8.

## 2 Related Work

There are two main categories of literature that are relevant to our work: research on fraud in online advertising and analyses of security investments on the Internet.

Research on online advertising fraud is mostly focused on click fraud [14] [15] [16]. Many problems that stem from online advertising and security gaps, especially the consequences for the end users, are addressed in [17]. The economics of click fraud are briefly addressed in [16]. An economic analysis [18], based on a game-theoretic model of the online advertising market, shows that ad networks that deploy effective algorithms for click fraud detection gain a significant competitive advantage. If it is the case that some ad networks do not fight click fraud, mechanisms are proposed in [19] to protect online advertisers from paying for fraudulent clicks. In comparison, our model does not address click fraud and introduces a new strategic player - the ISP - in addition to the traditional entities in online advertising (i.e., ad networks, advertisers and publishers). Our results show that this player can yield significant implications for the security of the Internet.

Related to the second issue - finding the right incentives to increase the security of the Internet - there are several contributions in the literature. Part of the research focuses on how risk management and cyberinsurance could be used as a tool for security management [20] [21] [22]. The game-theoretic approach of [23] on strategic security investment models how users choose between investments in security (e.g., firewalls) or insurance (e.g., backup) mechanisms. The positive effect of cyberinsurance on the investment of agents in self-protection is analyzed using a game-theoretic model in [24]. The main conclusion of this work is that cyberinsurance is not a good incentive for self-protection without regulation. Another line of work proposes a centralized certification mechanism to encourage ISPs to secure their traffic and analyzes the resulting scheme using game theory [25]. In contrast to these works, our analysis shows that Internet security can be increased, under given conditions, without any central oversight and thanks to self-interested decisions by only a few key players (namely, the ad networks).

## 3 System Model

We consider a system consisting of the online advertising system and an ISP, as depicted in Figure 1.

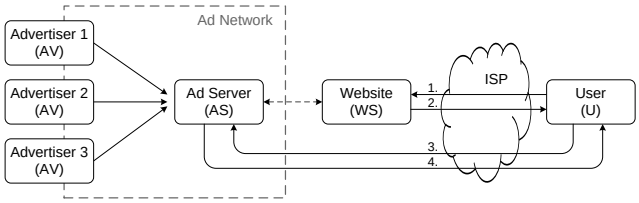



Figure 1. The system consists of an online advertising system and an ISP.

### 3.1 Online Advertising Systems

To have their ads appear with the appropriate web content, Advertisers (AV) subscribe with an ad network whose role is to automatically embed ads into web pages. Ad networks have contracts with Websites (WS) that want to host advertisements. When a User (U) visits such a website (Figure 1, step 1), while downloading the content of the web page (step 2), the user's browser will be directed to communicate with one of the Ad Servers (AS) belonging to the ad network (step 3). The ad server chooses and serves (step 4) the most appropriate ads to the user, such that users' interests are matched and the potential revenue is maximized. Throughout the rest of the paper, we use the terms "ad network" and "ad server" (that belongs to the ad network) interchangeably. We also use the terms "user" and "user's browser" interchangeably.

A user-generated click on an advertisement directs the user's browser to the advertised website and is called *click-through*. The event of a click-through being followed by a predefined users' action on the advertiser's website (e.g., online purchase or registration for a newsletter) is called *click conversion*. According to the terms of a contract, an advertiser pays a certain amount of money to the AS whenever a click-through or a click conversion on an advertisement occurs. The AS gives a fraction of the ad revenue to the WS that hosted the ad on which a click-through or a click conversion occurred. Throughout the rest of the paper, we use the term "clicks" to refer to the user-generated clicks on ads that create ad revenue for the AS and the WS.

The AS values an associated website based on the volume of clicks and ad traffic generated by the website's visitors clicking on hosted ads. Popular websites that attract a great number of visitors generate more clicks on ads, thus also create a high ad revenue for the associated AS and themselves.

### 3.2 Internet Service Providers

Traditionally, an ISP provides Internet access to end users and is topologically placed between users and the Internet. We say that the system operates in the *nominal mode* when the ISP only

faithfully forwards users' traffic. However, to capture the recent behavior of ISPs, in our system model the ISP can also either take advantage of the users' private information and operate alone as an ad network offering higher quality clicks to the set of its advertisers (*non-cooperative behavior*) or cooperate with ad servers by sharing users' private information to jointly improve ad targeting (*cooperative behavior*).

## 4 Threats and Countermeasures

Given that ISPs are in the position to observe all the traffic of its subscribers and that recently they had to invest in technologies that enable profiling of their subscribers' online behavior, ISPs can collect a high volume of users' private data. Such a rich data would be of immense value for ad networks as it can be used to improve the quality of matching ads to users' interests [10]. Consequently, ad networks could generate even higher ad revenues. Ad networks are already deploying mechanisms (e.g., third-party cookies) themselves to track users' interests. However, the collected information cannot be as rich as the ISPs are able to obtain, because ISPs have access to all the users' traffic (unless it is encrypted). Thus, ad networks might be willing to subsidize ISPs to profile users' online behavior in exchange for a share of ad revenue. When the ISP and the AS are cooperative the system operates in the *cooperative mode*.

Some ISPs might gain more revenue when being non-cooperative. A non-cooperative ISP plays a role similar to the role of ad networks: it uses the obtained information about users' interests and performs advertising services for a set of its own advertisers. As the ISP is the last hop in forwarding the traffic towards its subscribers, it can free-ride on the existing traffic to deliver ads of its choice to the end users. The ISP can simply perform modifications of the content of webpages on-the-fly between servers and users with the goal of modifying the original ads or injecting new ads. Another technique is for the ISP to replace entire web pages by modifying users' DNS traffic on-the-fly and redirecting users to servers of the ISP's choice.[1] Thus, the affected users would see altered ads, which are different from the original ads embedded into the webpages by a legitimate AS associated to the browsed website. When users click on the altered ads, the clicks generate revenue for the ISP instead of the AS and we say that the ISP has diverted the clicks from the AS. Consequently, the non-cooperative ISP diverts a part of the ad revenue from the AS. When the ISP is non-cooperative and diverts clicks (i.e., ad revenue) from the AS the system operates in the *non-cooperative* mode.

Depending on the AS's loss of ad revenue caused by the ISP diverting clicks, the AS might decide to deploy a countermeasure and prevent exploits by the non-cooperative ISP. A straightforward solution to prevent on-the-fly modifications is to deploy HTTPS instead of HTTP to deliver web content and ads. HTTPS provides data integrity and in case encryption is used would also reduce the amount of information ISPs can collect about users. Given the system architecture (Figure 1), data integrity is necessary in both communication channels[2]: (i) between users and websites and (ii) between users and ad servers. So far, HTTPS with valid authentication certificates is used mostly by a small fraction of websites (e.g., e-commerce). The major part of costs of implementing HTTPS at a web server is the cost of obtaining a valid $X.509$ authentication certificate[3]. Usually, websites' owners are not willing to bear this cost. Therefore, if the AS wants associated websites to deploy HTTPS, it has to cover the costs itself. Deploying HTTPS at ad servers is easy as they typically belong to major companies that already have valid authentication certificates.

Usually, users ignore security warnings related to certificate verification failures because a high number of websites use self-signed certificates. However, if websites use valid certificates, browsers can differentiate between: (i) the case of a website having a self-signed certificate and (ii) the case when an adversary tampers with a valid certificate or the content of a website. Consequently, browsers can deploy more sophisticated policies in handling associated security risks in these two

1. However, in this case the websites might detect the decrease in the number of visits and become suspicious.

2. Only data integrity property of HTTPS is necessary, encryption is optional.

3. Data integrity can be provided with Message Authentication Codes which are cheap in terms of computation and communication overhead. Thus, the per transaction cost of serving content over HTTPS instead of HTTP is negligible compared to the ad revenue.

cases and can display specifically targeted security warnings that alert users to not accept the content that has been altered by the adversary.

Each website maximizes its revenue by choosing the ad network whose ads it will host. A WS can be associated with the AS or with the ISP. This association is known, as the WS has a contract with the associated ad network. If the WS has willingly decided to associate with the ISP then the WS's ad revenue is not affected by the deviating behavior of the ISP. The concerned websites are the ones that have chosen to host the ads of the AS, but due to the actions of the non-cooperative ISP, the WS's web pages are displayed with ads of the ISP. Consequently, the WS loses the ad revenue.

When the WS that is originally associated with the AS is affected by the non-cooperative behavior of the ISP, it can only decide whether to accept to deploy HTTPS or not. As explained, the major cost of deploying HTTPS instead of HTTP at the WS is the cost of a certificate. If this cost is paid by the AS, then the remaining costs (e.g., per transaction computational and communication overhead of HTTPS compared to HTTP) are negligible compared to the ad revenue. Thus, in the presence of the non-cooperative ISP, if the AS is willing to bear the costs, the WS's revenue is maximized when it accepts to deploy HTTPS together with the AS. Since the AS bears the costs, we say that the AS *secures* the WS.

## 5 Game-Theoretic Model

We propose a game-theoretic model of the relationship between an ISP and an ad network. The strategic decision facing an ISP is to be cooperative or not with the ad network. In the case of a cooperative ISP, an ad network can offer a share of its revenue in exchange for the users' private information based on which it improves ad targeting. In the case of a non-cooperative ISP, an ad network can deploy security mechanisms to prevent the ISP from diverting the revenue. We study within this model the possible outcomes of this tension between the ISPs and ad networks.

### 5.1 Actions

We denote the two entities, an ISP and an Ad Server, as players $ISP$ and $AS$, respectively. We model the possible behavior of $ISP$ with the following three actions:

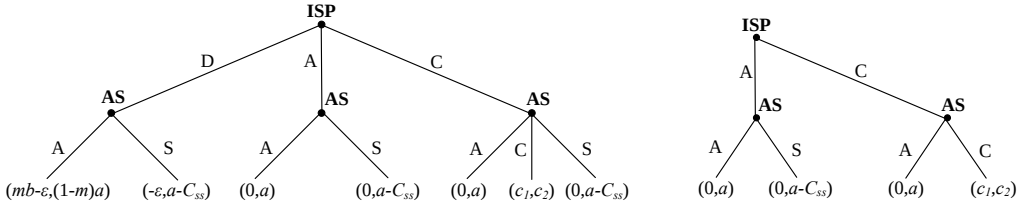
- ***Divert (D)***: $ISP$ diverts from $AS$ a fraction $m$ of the clicks generated at a website WS associated with the AS. In practice, this means that $ISP$ modifies the traffic on-the-fly. $ISP$ diverts the revenue from $AS$ because it charges for the clicks that were supposed to be associated to the original $AS$ with which the WS had an agreement. This action models the *non-cooperative* behavior of $ISP$.
- ***Cooperate (C)***: $ISP$ shares with $AS$ the collected private information about users in order to help $AS$ improve the quality of ad targeting. In return, it receives from $AS$ a share of the generated revenue. This action models the *cooperative* behavior of $ISP$.
- ***Abstain (A)***: $ISP$ takes no action. This models the initial behavior of $ISP$ when it operates in the nominal mode.

The player $AS$ can choose between the following three actions:

- ***Abstain (A)***: $AS$ does not react to the changed behavior of $ISP$. This models the initial behavior of $AS$ operating as in the nominal mode.
- ***Cooperate (C)***: $AS$ cooperates with $ISP$ by providing a share of its revenue in exchange for the users' private information.
- ***Secure (S)***: $AS$ secures a given website to prevent the $ISP$ from diverting clicks. The one-time cost ($C_{ss}$) of securing the website depends on the secure solution that is implemented. Our model applies, in general, to all solutions in which the AS pays a per website one-time cost ($C_{SS}$) to secure ad serving. In the case of HTTPS, $AS$ can buy a digital certificate from a Certification Authority (e.g., VeriSign) thus enabling the WS to communicate with users over the HTTPS protocol. HTTPS provides integrity of the content, hence preventing $ISP$ from meddling with users' traffic.

## 5.2 The Game

We model the problem as a dynamic, finite multi-stage game with perfect and complete information between $AS$ and $ISP$. We assume that $AS$ can detect on-the-fly modifications of the ad traffic using mechanisms such as web tripwires [5] and $ISP$ can observe if HTTPS has been deployed at a given WS or not, hence it is a game with perfect information. The game consists of $n$ stage games, where each stage game is an extensive-form game in which $ISP$ plays first and $AS$ plays second. This models the behavior observed in practice, where ISPs act first by taking part in the online advertising business and then the AS can react. We model the game as a finite game because business relationships usually have a finite duration. The length of the business relationship, known to the players, determines the value of $n$. If the website is not secured, in each stage game $ISP$ chooses among the actions $\{D,C,A\}$ and then $AS$ chooses among the actions $\{A,C,S\}$, as illustrated in Figure 2(a). If $AS$ secures the website at some stage of the multi-stage game, $ISP$ cannot divert clicks until the end of the game and $AS$ cannot secure the website again. Thus, in all of the following stages, if the website is secured the single stage game is as illustrated in Figure 2(b).




(a) Single stage game if a website is not secured

(b) Single stage game if a website is secured

Figure 2. Extensive-form single stage games. $ISP$ always plays first.

Note that in the model, we consider clicks on ads generated by $ISP$'s subscribers at a single website. The results are extended to the case of multiple websites in Section 7. The WS is not modeled as a player in the game because its revenue is maximized when the AS's revenue is maximized. As explained in Section 4, the WS always complies with a decision (to deploy HTTPS or not) that is made by the associated AS. The symbols used in the model are given in Table 1.

Table 1
TABLE OF SYMBOLS FOR THE GAME-THEORETIC MODEL.

| Symbol | Definition |
|---|---|
| $m$ | Fraction of clicks $ISP$ diverts |
| $\varepsilon$ | Cost of diverting clicks |
| $u_{AS}$ | Ad Server's total payoff |
| $u_{ISP}$ | $ISP$'s total payoff |
| $a$ | Ad Server's total payoff in the nominal model |
| $b$ | $ISP$'s per fraction revenue when diverting clicks |
| $c_1$ | $ISP$'s total payoff in the cooperation model |
| $c_2$ | Ad Server's total payoff in the cooperation model |
| $C_{ss}$ | One-time cost of securing a website |
| $n$ | Number of stages of the multi-stage game |
| $k_1$ | Stage at which $AS$ secures the website |
| $k_2$ | Stage at which $ISP$ starts diverting clicks |

## 5.3 Analytical Analysis and Results

In this section, we first explain the single stage games presented in Figure 2 and then we present the outcome of the multi-stage game.

In a stage game, when $ISP$ plays $A$ and it is not part of the online advertising system, $AS$ earns the nominal revenue $a$ and $ISP$ earns nothing. This corresponds to the case when both

players play $A$ and it represents the system operating in the nominal mode (i.e., when the $ISP$ only faithfully forwards the traffic). Thus, the payoffs of the $ISP$ ($u_{ISP}$) and the $AS$ ($u_{AS}$) are $(u_{ISP}, u_{AS})$=$(0, a)$.

Cooperation only emerges if both players are willing to cooperate, i.e., if they both play $C$. Therefore, $AS$ can choose action $C$ only if $ISP$ has played $C$. Let the corresponding payoffs in case of cooperation be $(u_{ISP}, u_{AS})$=$(c_1, c_2)$.

If $ISP$ plays $D$ followed by $AS$ playing $A$, a fraction $m$ of the clicks is successfully diverted, which brings revenue $mb$ to $ISP$. $ISP$ has to pay a small cost ($\varepsilon$) in every stage to divert clicks due to resources invested in mounting and performing attacks (e.g., parsing the code of a web page, identifying ads and replacing or adding ads).[4] Therefore, $ISP$'s payoff is $mb - \varepsilon$. When the diversion of a fraction $m$ of clicks is successful, $AS$ loses a part of its revenue proportional to the fraction of clicks being diverted, $ma$. Thus, the payoffs when $ISP$ successfully diverts clicks from $AS$ are $(u_{ISP}, u_{AS})$=$(mb - \varepsilon, (1 - m)a)$.

$AS$ can decide to play $S$ to prevent the loss of its revenue. $AS$ has to pay a one-time cost $C_{ss}$ which makes its payoff $a - C_{ss}$ in the stage when it secures the WS. After securing the website, $AS$ does not have to pay any other costs and it secures its nominal revenue $a$ in all future stages. Depending on whether $ISP$ has tried to divert clicks or not in the stage game when $AS$ implements security, it either has a cost $\varepsilon$ or not, which corresponds to payoffs $(u_{ISP}, u_{AS})$=$(-\varepsilon, a - C_{ss})$ and $(u_{ISP}, u_{AS})$=$(0, a - C_{ss})$, respectively.

To solve the finite multi-stage game with perfect information, we apply *backward induction* to determine the *subgame perfect Nash equilibrium* (SPNE) of the game [26]. The resulting outcome depends on the values of several parameters of the model. We perform an exhaustive analysis for all the possible values of the model parameters. There are five cases:

$$\text{Case 1} : ma \geq C_{ss} \text{ and } c_2 > a \tag{1}$$
$$\text{Case 2} : ma \geq C_{ss} \text{ and } c_2 \leq a \tag{2}$$
$$\text{Case 3} : ma < C_{ss} \text{ and } c_2 \leq a \tag{3}$$
$$\text{Case 4} : ma < C_{ss} \text{ and } c_2 > a \text{ and } c_1 \geq mb - \varepsilon \tag{4}$$
$$\text{Case 5} : ma < C_{ss} \text{ and } c_2 > a \text{ and } c_1 < mb - \varepsilon \tag{5}$$

In practice, the values of the parameters can be estimated by each of the players and they determine to which of the five cases of the model the system corresponds to.

Next, we present the results for each of the cases. We focus on the outcomes of the SPNE and we present the full SPNE strategy sets in Appendix A. For proofs, also see Appendix A.

***Result 1****: In Case* 1*, there is a unique SPNE where the outcome is (Cooperate,Cooperate) in every stage game and the corresponding total payoffs, summed over* $n$ *stages, are:*

$$\begin{aligned} u_{ISP} &= nc_1 \\ u_{AS} &= nc_2 \end{aligned} \tag{6}$$

In Case 1, if $ISP$ diverts a large fraction ($m \geq \frac{C_{ss}}{a}$) of clicks, the best response of $AS$ is to implement security because the cost of deploying a secure protocol ($C_{ss}$) is smaller than the loss of revenue due to the diversion of clicks ($ma \geq C_{ss}$). If $AS$ implements security, $ISP$ does not earn any revenue and it only pays the cost of mounting the attack, $u_{ISP} = -\varepsilon$. Therefore, it is better for $ISP$ either to abstain, in which case its payoff would be $u_{ISP} = 0$, or to offer cooperation, in which case its payoff would be $u_{ISP} = c_1$ if $AS$ accepts the cooperation. Thus, in Case 1, cooperation is the best action for $ISP$. Whether $ISP$ and $AS$ cooperate now depends on the action of $AS$. In Case 1, cooperation is also more profitable for $AS$ ($c_2 > a$), hence $AS$ accepts cooperation.

4. In practice, the cost $\varepsilon$ might not be exactly the same in each stage of the game. However, the variations are insignificant and since $\varepsilon$ is negligible compared to the ad revenue, assuming a constant cost per stage does not influence the results.

***Result 2****: In Case* 2*, there is a unique SPNE where the outcome is (Cooperate,Abstain) in every stage game and the corresponding total payoffs, summed over* $n$ *stages, are:*

$$\begin{aligned} u_{ISP} &= 0 \\ u_{AS} &= na \end{aligned} \tag{7}$$

As $m \geq \frac{C_{ss}}{a}$ holds in Case 2 as in Case 1, the best action for $ISP$ is to offer cooperation, as explained for Case 1. However, in Case 2 $AS$ obtains a higher revenue when operating alone than when cooperating with $ISP$ ($a \geq c_2$), thus $AS$ does not accept cooperation and the system operates in the nominal mode in every stage game.

***Result 3.1****: In Case* 3*, if* $m < \frac{C_{ss}}{na}$*, there is a unique SPNE where the outcome is (Divert,Abstain) in every stage game and the corresponding total payoffs, summed over* $n$ *stages, are:*

$$\begin{aligned} u_{ISP} &= n(mb - \varepsilon) \\ u_{AS} &= n(1-m)a \end{aligned} \tag{8}$$

If $ISP$ diverts such a small fraction $m < \frac{C_{ss}}{na}$ of clicks as in Result 3.1, the loss of revenue it imposes to $AS$ is not significant enough to cause $AS$ to secure the website, i.e., the cost of a secure solution exceeds the loss of revenue. Therefore, $ISP$ diverts a fraction $m$ of clicks in all stages and $AS$ does not react.

***Result 3.2****: In Case* 3*, if* $\frac{C_{ss}}{na} \leq m < \frac{C_{ss}}{a}$*, there are two SPNE that result in two different outcomes. The first outcome is: (Divert,Abstain) in the first* $k_1$ *stage games, where* $k_1 = \lfloor \frac{\varepsilon}{mb-\varepsilon} \rfloor$ *and* $0 < k_1 < n$*, (Divert,Secure) in the stage game* $k_1 + 1$ *and (Abstain,Abstain) till the end. The corresponding total payoffs, summed over* $n$ *stages, are:*

$$\begin{aligned} u_{ISP} &= k_1(mb - \varepsilon) - \varepsilon \\ u_{AS} &= k_1(1-m)a + a - C_{ss} + (n - k_1 - 1)a \end{aligned} \tag{9}$$

*The second outcome is (Abstain,Abstain) in the first* $k_2$ *stage games, where* $k_2 = \lceil \frac{nma - C_{ss}}{ma} \rceil$ *and* $0 < k_2 < n$*, and (Divert,Abstain) in the last* $n - k_2$ *stage games. The corresponding total payoffs, summed over* $n$ *stages, are:*

$$\begin{aligned} u_{ISP} &= (n - k_2)(mb - \varepsilon) \\ u_{AS} &= k_2 a + (n - k_2)(1-m)a \end{aligned} \tag{10}$$

Result 3.2 means that if $ISP$ wants to divert a high fraction of clicks, i.e., $\frac{C_{ss}}{na} \leq m < \frac{C_{ss}}{a}$, it cannot do so in all stages but only in a limited number of stages of the game. The two outcomes show that $ISP$ has two options to divert clicks. In the first outcome, $ISP$ diverts clicks in the first $k_1$ stage games, which causes $AS$ to secure the website in the stage game $k_1 + 1$ because the loss of revenue for $AS$ is higher than the cost of deploying the secure protocol. In the remaining stages, $ISP$ cannot divert clicks and there is no cooperation, as $AS$ earns more when operating alone ($a \geq c_2$), hence the system operates in the nominal mode. The second outcome means that $ISP$ has another possibility to divert clicks and avoid $AS$ securing the website. If $ISP$ abstains in the first $k_2$ stage games, it can then divert clicks in the remaining $n - k_2$ stage games till the end, with a fraction $m < \frac{C_{ss}}{(n-k_2)a}$. Intuitively, $ISP$ can divert clicks in a larger number of stage games but with a smaller fraction, or for a smaller number of stage games but with a larger fraction.

***Result 4****: In Case* 4*, there is a unique SPNE where the outcome is (Cooperate,Cooperate) in every stage game and the corresponding total payoffs are given by (6).*

In Case 4, as both $AS$ and $ISP$ earn more when cooperating than in any other mode ($c_2 > a$ and $c_1 \geq mb - \varepsilon$), their best actions are to always cooperate.

***Result 5.1**: In Case 5, if $m < \frac{(n-1)(a-c_2)+C_{ss}}{na}$, there is a unique SPNE where the outcome is (Divert,Abstain) in every stage game and the corresponding total payoffs are given by (8).*

The result shows that when $ISP$ diverts a small fraction of clicks the loss of revenue for $AS$ is not significant enough to invest in securing the WS.

***Result 5.2**: In Case 5, if $\frac{(n-1)(a-c_2)+C_{ss}}{na} \leq m < \frac{C_{ss}}{a}$, there are two SPNE that result in two different outcomes. The first outcome is: (Divert,Abstain) in the first $k_1$ stage games, where $k_1 = \lfloor \frac{\varepsilon+c_1}{mb-\varepsilon-c_1} \rfloor$ and $0 < k_1 < n$, (Divert,Secure) in the stage game $k_1+1$ and (Cooperate,Cooperate) till the end. The corresponding total payoffs, summed over $n$ stages, are:*

$$\begin{aligned} u_{ISP} &= k_1(mb - \varepsilon) - \varepsilon + (n - k_1 - 1)c_1 \\ u_{AS} &= k_1(1 - m)a + a - C_{ss} + (n - k_1 - 1)c_2 \end{aligned} \tag{11}$$

*The second outcome is (Cooperate,Cooperate) in the first $k_2$ stage games, where $k_2 = \lceil n - \frac{C_{ss}-a+c_2}{ma-a+c_2} \rceil$ and $0 < k_2 < n$, and (Divert,Abstain) in the last $n - k_2$ stage games. The corresponding total payoffs, summed over $n$ stages, are:*

$$\begin{aligned} u_{ISP} &= k_2c_1 + (n - k_2)(mb - \varepsilon) \\ u_{AS} &= k_2c_2 + (n - k_2)(1 - m)a \end{aligned} \tag{12}$$

Result 5.2 shows that, as in Case 3, if $ISP$ wants to divert a higher fraction of clicks it has two possibilities: (i) divert in the first $k_1$ stage games (the first outcome), or (ii) divert in the last $n - k_2$ stage games (the second outcome). The difference between the outcomes in Cases 3 and 5 is that when in Case 3 the system operates in the nominal mode, in Case 5 $AS$ and $ISP$ cooperate. For $ISP$, cooperation is always better than operating in the nominal mode when it earns nothing. However, $AS$ benefits more when operating alone than when cooperating ($a \geq c_2$) in Case 3, so it does not agree to cooperate. In Case 5 cooperation is more profitable ($c_2 > a$), hence $AS$ agrees to cooperate.

The obtained outcomes of the multi-stage game for all the possible cases of parameters are presented in Table 2. Each column corresponds to a SPNE of the multi-stage game and each row corresponds to the achieved outcomes in each stage of the multi-stage game. Note that stages $k_1$ and $k_2$ are different in Case 3.2 and Case 5.2 and can be calculated with the expressions presented in Result 3.2 and Result 5.2. For the simplicity of presentation we abstract this in Table 2 and use the same symbols $k_1$ and $k_2$ for the both cases.

Table 2
OUTCOMES OF THE MULTI-STAGE GAME.

| **Stage of the multi-stage game** | **Case 1** | **Case 2** | **Case 3.1** | **Case 3.2** | | **Case 4** | **Case 5.1** | **Case 5.2** | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | (C,C) | (C,A) | (D,A) | (D,A) | (A,A) | (C,C) | (D,A) | (D,A) | (C,C) |
| 2 | (C,C) | (C,A) | (D,A) | (D,A) | (A,A) | (C,C) | (D,A) | (D,A) | (C,C) |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| $k_1$ | (C,C) | (C,A) | (D,A) | (D,A) | (A,A) | (C,C) | (D,A) | (D,A) | (C,C) |
| $k_1 + 1$ | (C,C) | (C,A) | (D,A) | ***(D,S)*** | (A,A) | (C,C) | (D,A) | ***(D,S)*** | (C,C) |
| $k_1 + 2$ | (C,C) | (C,A) | (D,A) | ***(A,A)*** | (A,A) | (C,C) | (D,A) | ***(C,C)*** | (C,C) |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| $k_2$ | (C,C) | (C,A) | (D,A) | (A,A) | (A,A) | (C,C) | (D,A) | (C,C) | (C,C) |
| $k_2 + 1$ | (C,C) | (C,A) | (D,A) | (A,A) | ***(D,A)*** | (C,C) | (D,A) | (C,C) | ***(D,A)*** |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| $n$ | (C,C) | (C,A) | (D,A) | (A,A) | (D,A) | (C,C) | (D,A) | (C,C) | (D,A) |

# 6 Refinement of the Game-theoretic Model

In order to understand the implications of this game-theoretic model in reality, we will apply the analysis of Section 5.3 to the real data set. Thus, we must first refine the game-theoretic model by estimating the values of the parameters using the data that characterize an online advertising system in practice. We consider three different modes of operation: (i) *Nominal* (Figure 3), (ii) *Non-cooperative* (Figure 4) and (iii) *Cooperative* (Figure 5), that capture possible interactions between entities of the system. The symbols used below are given in Table 3.

Table 3
TABLE OF SYMBOLS FOR THE NUMERICAL ANALYSIS.

| Symbol | Definition |
|---|---|
| $\mathcal{K}$ | Set of Advertisers |
| $\mathcal{K}_1$ | Set of Advertisers associated only with the Ad Server |
| $\mathcal{K}_2$ | Set of Advertisers associated both with the Ad Server and the ISP |
| $h$ | Fraction of revenue paid by the Ad Server to websites |
| $l$ | Fraction of revenue paid by the Ad Server to ISP when cooperating |
| $s$ | Fraction of revenue paid by the ISP to a third party for providing targeted advertising |
| $\beta_j$ | Fraction of clicks that become conversions |
| $Q$ | Volume of clicks |
| $v_{k,j}$ | Advertiser $k$ valuation of $j$'s clicks |

## 6.1 Nominal Mode

The system operating in the nominal mode is depicted in Figure 3. It corresponds to the case when $ISP$ is faithfully forwarding the traffic and does not try to take part in the online advertising system. A number of clicks, $Q$, is generated by users at the website WS. The clicks are registered by $AS$ that distributes them among associated AVs. We assume that $AS$ distributes clicks uniformly at random among the AVs. In practice, the volume of clicks given to each advertiser is typically determined in an auction based on advertisers' bids on given keywords. Modeling the auction process would add complexity to the problem and is out of the scope of this paper, therefore we assume that all advertisers receive the same amount of clicks. We also assume that there is no click fraud, i.e., all the clicks from one ad network have the same conversion probability. Let the conversion probability of a click from $AS$ be $\beta_1$.

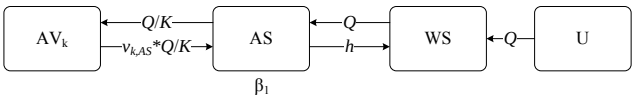



Figure 3. Nominal Mode.

Advertiser $AV_k \in \mathcal{K}$, where $\mathcal{K}$ is the set of all AVs associated to $AS$ and $K = |\mathcal{K}|$, selects its valuations $v_{k,AS}$ on clicks such that its revenue from $AS$ is maximized. The valuations are directly proportional to the conversion probability of the clicks (i.e., the quality of the clicks) received from $AS$ [18].

For the clicks that turn into conversions, AVs pay $AS$, and $AS$ pays a fraction $h$ of that amount to the $WS$ where the clicks were generated. We assume that AVs pay $AS$ an amount of money equal to their valuations of clicks (i.e., bids). Therefore, the nominal payoff of $AS$, $a$, is:

$$u_{AS} = \frac{Q}{K}(1-h)\sum_{k\in\mathcal{K}} v_{k,AS} = a \tag{13}$$

## 6.2 Non-cooperative Mode

If $ISP$ chooses to become part of the online advertising system and to divert clicks from $AS$, the system can be modeled as in Figure 4. $ISP$ diverts a fraction $m$ of $Q$ clicks generated at the website $WS$ and distributes it uniformly at random among the set of its own associated advertisers.

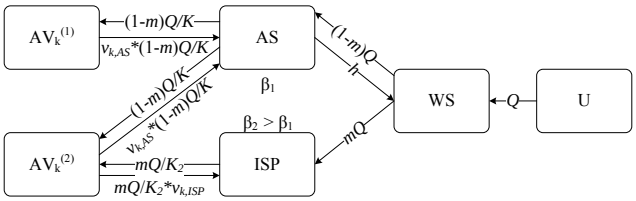



Figure 4. Non-cooperative Mode.

In the non-cooperative model, we assume two types of AVs:

- Advertisers of type 1, $AV^{(1)}$, are associated only with $AS$ because they care about their reputation and they do not associate with $ISP$ even if it would increase their revenues. The set of $AV^{(1)}$ is represented by $\mathcal{K}_1$, where $K_1 = |\mathcal{K}_1|$.
- Advertisers of type 2, $AV^{(2)}$, are associated with both $ISP$ and $AS$. $AV^{(2)}$ are willing to associate with $ISP$, because working with both $AS$ and $ISP$ generates more revenue. The set of $AV^{(2)}$ is represented by $\mathcal{K}_2$, where $K_2 = |\mathcal{K}_2|$.

There are no advertisers associated only with $ISP$, because advertisers that do not care about their reputation have higher revenue when associated with both $ISP$ and $AS$ than in the case when they are associated only with $ISP$. Therefore, we have $\mathcal{K} = \mathcal{K}_1 \cup \mathcal{K}_2$ and $K_1 + K_2 = K$.

An advertiser $AV_k^{(2)}$, associated with both $AS$ and $ISP$, selects its valuations $v_{k,AS}$ and $v_{k,ISP}$ on clicks such that its revenues from $AS$ and $ISP$ are maximized.

The conversion probability of clicks coming from $ISP$ ($\beta_2$) is higher than the conversion probability of clicks coming from $AS$ ($\beta_1$), i.e., $\beta_2 > \beta_1$, due to better ad targeting based on users' private information. Therefore, an advertiser places higher valuations on clicks from $ISP$ than on clicks from $AS$, i.e., $v_{k,ISP} > v_{k,AS}$. The difference in valuations on clicks from two different ad networks can be expressed as [18]:

$$v_{k,ISP} = \frac{\beta_2}{\beta_1} v_{k,AS} \tag{14}$$

Given that ISPs do not necessarily have the resources to perform ad targeting themselves, we assume that they rely on a third party entity, as observed in practice [11]. The partnering entity provides ad targeting technology and in return, $ISP$ gives the partner a fraction $s$ of its revenue. The payoffs of $AS$ and $ISP$ in the non-cooperative model are:

$$u_{AS} = \frac{(1-m)Q}{K}(1-h)\sum_{k\in\mathcal{K}} v_{k,AS} = (1-m)a \tag{15}$$

$$u_{ISP} = \frac{mQ}{K_2}(1-s)(\sum_{k\in\mathcal{K}_2} v_{k,ISP}) - \varepsilon = mb - \varepsilon \tag{16}$$

where

$$b = \frac{Q}{K_2}(1-s)\sum_{k\in\mathcal{K}_2} v_{k,ISP} \tag{17}$$

### 6.3 Cooperative Mode

When cooperating with $AS$ (Figure 5), $ISP$ provides users' private information $P$ that $AS$ uses to improve ad targeting, i.e., to improve the conversion probability of a click. The benefit for AVs is that they receive clicks that have higher probability of conversion ($\beta_2$) which is why they offer higher valuations ($v_{k,ISP}$) for those clicks. Thus $AS$ earns more money for $Q$ clicks when cooperating than when operating alone. In return for users' private information, $AS$ gives a fraction $l$ of the revenue to $ISP$.

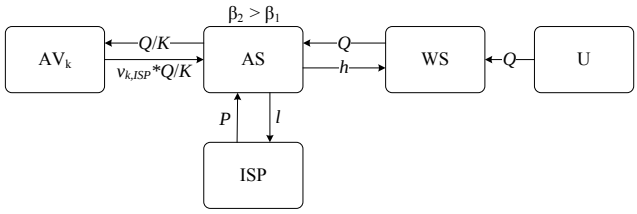



Figure 5. Cooperative Mode.

In the cooperative model, based on (13) and (14), the payoffs of $AS$ and $ISP$ are:

$$u_{AS} = \frac{Q}{K}(1-h-l)\sum_{k\in\mathcal{K}} v_{k,ISP} = \frac{\beta_2}{\beta_1}\frac{1-h-l}{1-h}a = c_2 \tag{18}$$

$$u_{ISP} = \frac{Q}{K}l\sum_{k\in\mathcal{K}} v_{k,ISP} = \frac{\beta_2}{\beta_1}\frac{l}{1-h}a = c_1 \tag{19}$$

Cooperation is good for $AS$ when $l \leq (1-h)(1-\frac{\beta_1}{\beta_2})$, i.e., when the cooperation revenue ($c_2$) is greater than the nominal revenue ($a$), based on (18).

## 7 Numerical Analysis

In this section, we evaluate the impact of the results in Section 5.3 on the Web using the above equations and a real data set. We extend the analysis to multiple websites. Note that the outcome of the game can be different for different websites, e.g., $AS$ can decide to secure only some of the websites while cooperating with $ISP$ for the others. We are interested in the outcomes of the game for the most popular $1000$ websites.

### 7.1 Evaluations on a Real Data Set

The exact values of parameters that characterize the system in practice are difficult for us to obtain. Many of them are kept confidential (e.g., $Q$ and $h$) and some are difficult to quantify (e.g., the value of users' private information). However, this information is available to the participating entities of the game, namely ad networks and ISPs, thus our model is applicable in practice.

We use the following representative values of system parameters in our evaluations: (i) $AS$ pays $h = 10\%$ of the revenue to its referrers per click conversion [27]; (ii) $ISP$ gives $s = 30\%$ of the revenue to a third party ad targeting company (varying the values of $s$ has no significant effect on the results); (iii) the cost of each certificate is \$399 [28]; (iv) the cost of mounting an attack is $\varepsilon \leq \$100$ (writing and deploying scripts to perform on-the-fly modifications of the ad traffic have a negligible cost, especially compared to the ad revenue and hence, the value of $\varepsilon$ has no effect on the results in practice) and (v) advertisers pay $\$0.5$ per click conversion [29].

We infer the generated volume of clicks on ads on the $1000$ most popular websites on the Web, based on the data of page views on each website in June 2009 (Figure 6), obtained from *Compete.com*. Based on the measurements reported in [30], $58\%$ of the top websites host ads and there are $8$ ads per page on average. The probability that a click occurs on an ad is around

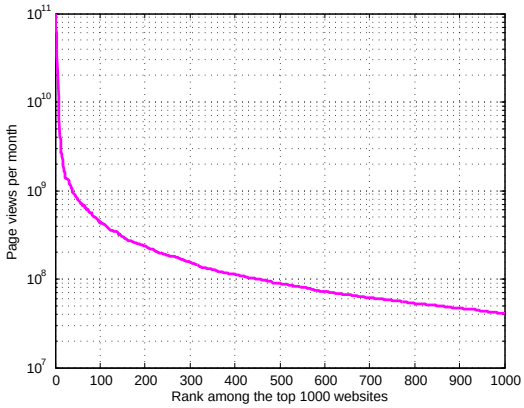



Figure 6. Popularity of the top 1000 websites based on page views per month.

$0.1\%$ [31]. Consequently, to convert the number of page views into the number of clicks on ads on each website, we use the following formula: $Q_i$=(Page views on the website $i$)$*0.58*8*0.001$.

There are two system parameters that influence the outcomes of the game: the fraction of shared revenue when cooperating ($l$) and the improvement of ad targeting ($\frac{\beta_2}{\beta_1}$). Thus, we take into account different values of the two parameters and analyze their effects. The fraction $m$ of clicks diverted by non-cooperative $ISP$ is also kept as a parameter of the analysis. We vary this parameter, and then consider the equilibrium outcome for each of the $1000$ most popular websites, as predicted by the analysis in Section 5.3. Our numerical results show that the outcomes are mostly determined by the values of the three parameters: $l$, $\frac{\beta_2}{\beta_1}$ and $m$. By varying values of other system parameters we conclude that they only insignificantly change the absolute values of the results but not the main observations.

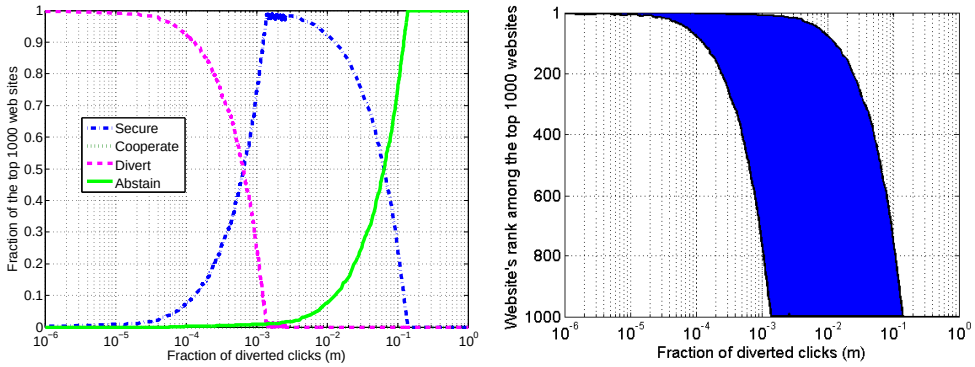
## 7.2 Numerical Results

In the case of a non-cooperative $ISP$, the outcomes of the multi-stage game for the $1000$ most popular websites are depicted in Figure 7(a). To obtain the non-cooperative scenario, we consider that the fraction of shared revenue when cooperating is high ($l = 0.4$) and ad targeting is not significantly improved ($\frac{\beta_2}{\beta_1} = 1.75$). The $AS$ is not willing to cooperate and pay such a high price for not so valuable user profiles, thus we observe the non-cooperative behavior. Outcomes are represented with the four curves in Figure 7(a). Each curve represents a fraction of websites for which the outcome of the game is the same.[5]

All the values of the *Cooperate* curve are equal to zero, which shows that, in this scenario, cooperation will not be established in any stage of the multi-stage game, for any of the websites.[6] The *Divert* curve represents the fraction of websites from which $ISP$ successfully diverts a fraction $m$ of clicks during all stages of the multi-stage game.[7] The *Secure* curve represents the fraction of websites that $AS$ will secure at some stage of the multi-stage game, due to $ISP$ diverting clicks.[8] The fraction of websites for which $ISP$ will abstain during all stages of the multi-stage game is represented with the *Abstain* curve.[9]

Results show that $ISP$ can divert a small fraction ($m < 0.001\%$) of clicks from all of the $1000$ websites (*Divert* curve equal to one) without causing $AS$ to react (*Secure* curve equal to zero). This amount of click diversion could be done in practice either by a very small ISP modifying all the traffic of its subscribers or by a large ISP selectively modifying only a tiny portion of the traffic it forwards.

5. For a given $m$, the sum of the values of the four curves is always equal to one.
6. The SPNE that correspond to Result 1, Result 4 and Result 5.2 are not achieved in the non-cooperative scenario.
7. The SPNE that correspond to Result 3.1 and Result 5.1.
8. The SPNE that corresponds to Result 3.2.
9. The SPNE that corresponds to Result 2.




(a) Outcomes in the non-cooperative scenario (*Cooperate* curve is equal to zero)

(b) Rank of the websites that should be secured

Figure 7. Outcomes of the game in the non-cooperative scenario applied to real data.

If $ISP$ starts diverting a higher fraction of clicks, it causes $AS$ to deploy security and protect the concerned websites. Thus, we observe that the fraction of websites that will be secured among the top 1000 websites (*Secure* curve) is increasing for higher values of $m$. Consequently, the fraction of websites from which $ISP$ successfully diverts clicks (*Divert* curve) is decreasing. When $ISP$ diverts $m = 0.14\%$ of clicks, almost all ($98.7\%$) of the 1000 websites should be secured.

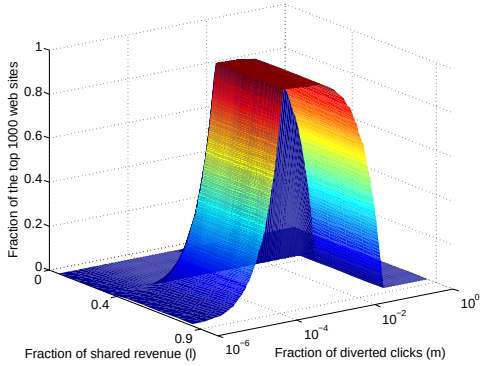
If $ISP$ is to divert a higher fraction ($m > 0.14\%$) of clicks, it would try do so only for the websites for which the condition $ma < C_{ss}$ holds, i.e., for which the revenue that $ISP$ would divert from $AS$ is smaller than $AS$'s cost of deploying the security mechanism. Otherwise, $AS$ would secure the websites in the first stage of the game, which would cause $ISP$ to only pay the cost of mounting the attack without any gain. Thus, if the condition $ma < C_{ss}$ does not hold for a given website, $ISP$ abstains during all stages of the multi-stage game. The fraction of such websites for which $ISP$ abstains during all stages of the multi-stage game is higher for higher values of $m$, as represented with the increase of the *Abstain* curve following the increase of $m$. This implies that the fraction of websites from which $ISP$ will try to divert clicks becomes smaller, thus resulting in fewer websites that need to be secured by $AS$ (corresponding decreasing values of the *Secure* curve). Further, results show that $ISP$ will not try to divert a high fraction ($m > 14\%$) of clicks from any of the websites, but rather choose to abstain (*Abstain* curve equal to one).

The *Secure* curve in Figure 7(a) only shows the fraction of websites that will be secured, but we are also interested in which websites are those. The colored area in Figure 7(b) corresponds to the popularity ranks of the websites that should be secured for a given value of $m$. Intuitively, since the $ISP$ diverts the same fraction $m$ of clicks from all the websites and more popular websites generate higher ad revenue, the loss of ad revenue for $AS$ is higher for more popular websites. As the cost of securing a website is the same for all the websites, it is better for $AS$ to first secure the most popular websites (i.e., those that generate highest ad revenue) among the ones $ISP$ tries to divert clicks from. In this way, $AS$ protects more ad revenue at the same cost.

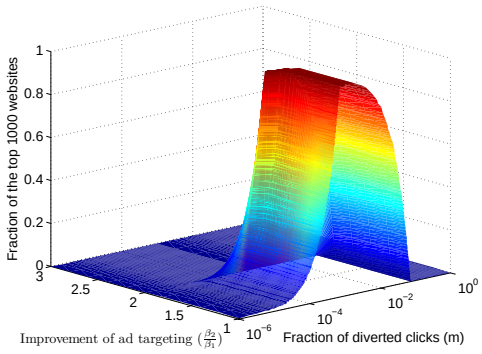
Based on the results in Figure 7(a), for small values of $m$ ($m \leq 0.14\%$) $ISP$ tries to divert clicks from all of the websites and the fraction of websites to be secured increases with the increase of $m$. The colored area in Figure 7(b) shows that for $m \leq 0.14\%$ $AS$ secures the fraction of websites starting from the most popular ones, i.e., the highest ranked websites according to their popularity.

However, as $m$ increases ($m > 0.14\%$) $ISP$ will stop trying to divert clicks from the most popular websites. We concluded earlier that $ISP$ will not try to divert a given fraction $m$ of clicks from websites for which the condition $ma < C_{ss}$ does not hold, as it would obtain a negative payoff. For a given $m$, this becomes true first for the most popular websites that generate high ad revenue $a$. Therefore, $ISP$ would only try to divert clicks from the less popular websites. Consequently, the threat exists only for the less popular websites and the most popular among those

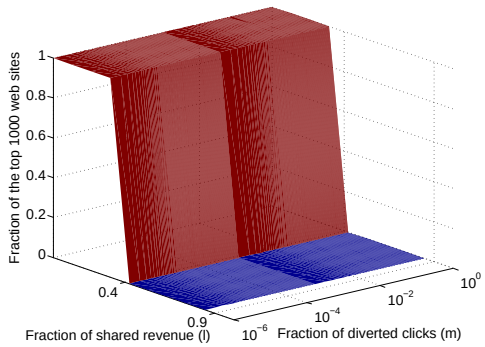
are the ones that will be secured by $AS$. For example, for $m = 5\%$, $60\%$ of the websites will be secured by $AS$ (*Secure* curve equal to 0.6 in Figure 7(a)) that correspond to websites ranked from 400 to 1000 (Figure 7(b)). For the highest ranked $40\%$ there is no need to implement security as $ISP$ would abstain from diverting clicks from those, knowing that $AS$ would immediately implement security as $ma > C_{ss}$.



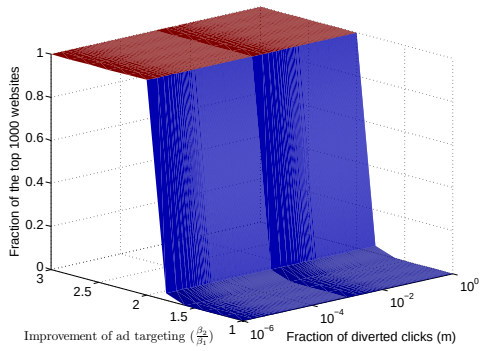

(a) Fraction of the top 1000 websites that should be secured (depending on $l$)




(b) Fraction of the top 1000 websites that should be secured (depending on $\frac{\beta_2}{\beta_1}$)




(c) Fraction of top the 1000 websites for which cooperation is achieved (depending on $l$)




(d) Fraction of the top 1000 websites for which cooperation is achieved (depending on $\frac{\beta_2}{\beta_1}$)

Figure 8. Effects of the parameters $l$ and $\frac{\beta_2}{\beta_1}$ on the game outcomes.

Next, we analyze the effect of the parameters $l$ and $\frac{\beta_2}{\beta_1}$ on the results. Figures 8(a) and 8(b) represent the *Secure* curve for different values of parameters $l$ and $\frac{\beta_2}{\beta_1}$, respectively. The graphs show that non-cooperative behavior occurs when $ISP$ demands a high share ($0.4 \leq l \leq (1-h)$) for the users' profiles and when ad targeting cannot be significantly improved ($\frac{\beta_2}{\beta_1} < 2$). Observe that the fraction of the websites to be secured follows the same pattern as in Figure 7(a). Thus, following our analysis of the non-cooperative behavior, the threat of $ISP$ diverting ad revenue can lead to improved Web security. In our example, if $ISP$ modifies around $0.14\%$ of the clicks, almost all of the websites should be secured.

The graphs in Figures 8(c) and 8(d) represent the fraction of the 1000 most popular websites for which $ISP$ and $AS$ cooperate during all stages of the multi-stage game.[10] The results show that if $AS$ does not have to give a high share of its revenue to $ISP$ ($0 < l < (1-h)(1-\frac{\beta_1}{\beta_2})$) or if the users' private information can significantly improve ad targeting ($\frac{\beta_2}{\beta_1} \geq 2$), $ISP$ and $AS$ cooperate for all of the websites.

We do not show the equilibrium outcomes *Abstain* and *Divert*, as they also follow the patterns in Figure 7(a).

10. The SPNE that correspond to Result 1 and Result 4.

## 8 Conclusion

In this paper, we have investigated the recent problem of ISPs becoming strategic participants in the online advertising business. We have proposed a game-theoretic model of this problem to study the behavior and interactions of the ISPs and ad networks. We have applied our model to the real data of the 1000 most popular websites to understand the meaning of the results in practice. Our analysis shows that whether an ISP will be *non-cooperative* or *cooperative* mostly depends on the value of the users' private information obtained by ISPs and on their share of the advertising revenue. The effect on the Web is positive in both cases: When ISPs are cooperative, users receive better targeted ads and both ISPs and ad networks earn higher revenues; when ISPs are non-cooperative, Web security can be improved as a side effect of protecting the ad revenue.

## References


[1] J. Crowcroft, "Net Neutrality: The technical side of the debate: A white paper," *SIGCOMM Computer Communication Review*, vol. 37, no. 1, pp. 49–56, 2007.

[2] (2008) Growing number of ISPs injecting own content into websites. [Online]. Available: http://www.techdirt.com/articles/20080417/041032874.shtml

[3] (2008) ISPs meddled with their customers' Web traffic, study finds. [Online]. Available: http://www.pcworld.com/businesscenter/article/144682/isps_meddled_with_their_customers_web_traffic_study_finds.html

[4] B. April, F. Hacquebord, and R. Link, "A Cybercrime Hub," *A Trend Micro White Paper*, August, 2009.

[5] C. Reis, S. D. Gribble, T. Kohno, and N. C. Weaver, "Detecting In-Flight Page Changes with Web Tripwires," in *NSDI'08: Proceedings of the 5th USENIX Symposium on Networked Systems Design and Implementation*.

[6] "Directive 2006/24/EC of the European parliament and of the council," *Official Journal of the European Union*, 2006.

[7] "Retention of Communications Data under Part 11: Anti-Terrorism, Crime and Security Act 2001, Home Office."

[8] K. Mochalski and H. Schulze, "Deep Packet Inspection - Technology, Applications and Net Neutrality," *ipoque White Paper*, 2009.

[9] "IAB Internet Advertising Revenue Report, 2008 full-year results," *Interactive Advertising Bureau*, 2009.

[10] (2009) Making ads more interesting. [Online]. Available: http://googleblog.blogspot.com/2009/03/making-ads-more-interesting.html

[11] (2008) Phorm. [Online]. Available: http://www.phorm.com/

[12] (2009) Telecoms: Commission launches case against UK over privacy and personal data protection. [Online]. Available: http://europa.eu/rapid/pressReleasesAction.do?reference=IP/09/570

[13] N. Vratonjic, J. Freudiger, M. Felegyhazi, and J.-P. Hubaux, "Securing Online Advertising," EPFL, Switzerland, Technical Report 2008-017, 2008.

[14] N. Daswani and M. Stoppelman, "The Anatomy of Clickbot.A," in *HotBots'07: Proceedings of the first conference on First Workshop on Hot Topics in Understanding Botnets*. USENIX Association, 2007.

[15] M. Gandhi, M. Jakobsson, and J. Ratkiewicz, "Badvertisements: Stealthy Click-Fraud with Unwitting Accessories," *Digital Forensic Practice*, vol. 1, no. 2, 2006.

[16] M. Jakobsson and Z. Ramzan, *Crimeware*. Reading, MA: Addison-Wesley, 2008.

[17] B. G. Edelman, "Securing Online Advertising: Rustlers and Sheriffs in the New Wild West," *SSRN eLibrary*, 2008.

[18] B. Mungamuru, S. Weis, and H. Garcia-Molina, "Should Ad Networks Bother Fighting Click Fraud? (Yes, They Should.)," Stanford InfoLab, Technical Report, July 2008.

[19] B. G. Edelman, "Deterring Online Advertising Fraud Through Optimal Payment in Arrears," *SSRN eLibrary*, 2009.

[20] R. Böhme, "Cyber-Insurance Revisited," in *WEIS'05: Proceeding of Workshop on the Economics of Information Security*.

[21] L. A. Gordon, M. P. Loeb, and T. Sohail, "A framework for using insurance for cyber-risk management," *Communications of the ACM*, vol. 46, no. 3, pp. 81–85, 2003.

[22] N. Shetty, G. Schwartz, M. Felegyhazi, and J. Walrand, "Competitive Cyber-Insurance and Internet Security," in *WEIS'09: Proceeding of Workshop on the Economics of Information Security*.

[23] J. Grossklags, N. Christin, and J. Chuang, "Secure or insure?: a game-theoretic analysis of information security games," in *WWW'08: Proceeding of the 17th international conference on World Wide Web*.

[24] M. Lelarge and J. Bolot, "Economic incentives to increase security in the internet: The case for insurance," in *INFOCOM'09: Proceedings of the 29th IEEE Conference on Computer Communications*.

[25] X. Zhao, F. Fang, and A. B. Whinston, "An economic mechanism for better Internet security," *Decision Support Systems*, vol. 45, no. 4, pp. 811–821, 2008.

[26] D. Fudenberg and J. Tirole, *Game Theory*. MIT Press, 1991.

[27] AdBrite Referral Program. [Online]. Available: http://www.adbrite.com/mb/affiliate_info.php

[28] VeriSign Inc. [Online]. Available: http://www.verisign.com/ssl/buy-ssl-certificates/secure-site-services/index.html

[29] U.S. average CPC by category, May and June 2009. [Online]. Available: http://www.clickz.com/3634374

[30] Krishnamurthy, Balachander, and C. E. Wills, "Cat and mouse: content delivery tradeoffs in Web access," in *WWW '06: Proceedings of the 15th international conference on World Wide Web*, 2006.

[31] "2008 Year-in-Review Benchmarks," *DoubleClick Research Report*, June, 2009.

# Appendix A.
# Proofs

We use induction to prove that the payoff expressions in Section 5.3 hold for any $n \geq 1$. Next, we apply backward induction to these payoffs to solve the multi-stage game of $n$ stages. The backward induction algorithm constructs a SPNE in finite games of perfect information [26]. We only present proofs of the results for Case 3 as they are more complex. Results for Case 5 can be proven in the same way as for the Case 3. Proofs for Cases 1, 2 and 4 are trivial.

Applying backward induction to the single stage game (Figure 2(a)) in Case 3 results in a unique SPNE with the strategy (D,AAA). The corresponding total payoffs in the game outcome, (Divert,Abstain), are: $((mb-\varepsilon),(1-m)a)$.

To prove the payoff expressions for the $n$ stage game, we prove that they hold for 1 stage, we assume they are true for $j$ stages and prove that they hold for $j+1$ stages. We assume the relevant subgames (denoted by $SG$) and the respective payoffs in the multi-stage game with $j$ stages:

- $SG_1 : (j(mb-\varepsilon), j(1-m)a)$, which corresponds to $ISP$ successfully diverting clicks in all stage games;
- $SG_2 : (k_1(mb-\varepsilon)-\varepsilon, k_1(1-m)a+a-C_{ss}+(j-k_1-1)a) = (k_1(mb-\varepsilon)-\varepsilon, (j-k_1m)a-C_{ss})$, $0 < k_1 < j$, which corresponds to $ISP$ successfully diverting clicks in the first $k_1$ stage games, resulting in $AS$ implementing security in the stage game $k_1+1$, followed by the system operating as in the nominal mode till the end;

- $SG_3 : ((j-k_2)(mb-\varepsilon), k_2a+(j-k_2)(1-m)a) = ((j-k_2)(mb-\varepsilon), (k_2m+j(1-m))a)$, $0 < k_2 < j$, which corresponds to the system operating as in the nominal mode in the first $k_2$ stage games, followed by $ISP$ diverting clicks till the end.

If we set $j = 1$ in $SG_1$ ($SG_2$ and $SG_3$ do not exist in this case), we obtain the outcome of a single stage game. Now let us extend the $j$ stage game with an additional stage game and solve the multi-stage game of $j + 1$ stage games. For all subgames where the security was not implemented, in the unique SPNE in the $j + 1$st stage game the outcome is (Divert,Abstain). Therefore, we add the payoffs $(\Delta u_{ISP}, \Delta u_{AS}) = (mb-\varepsilon, (1-m)a)$ to the payoffs of $ISP$ and $AS$ obtained after $j$ stage games. In the subgames where security has been implemented, in the unique SPNE in the $j + 1$st stage game the outcome is (Abstain,Abstain). We add the payoffs $(\Delta u_{ISP}, \Delta u_{AS}) = (0, a)$ to the payoffs of $ISP$ and $AS$ obtained after $j$ stage games.

The obtained payoffs after $j + 1$ stage games are:

- $SG_1 : (j(mb-\varepsilon)+(mb-\varepsilon), j(1-m)a+(1-m)a) = ((j+1)(mb-\varepsilon), (j+1)(1-m)a)$;
- $SG_2 : (k_1(mb-\varepsilon)-\varepsilon+0, k_1(1-m)a+a-C_{ss}+(j-k_1-1)a+a) = (k_1(mb-\varepsilon)-\varepsilon, (j+1-k_1m)a-C_{ss})$, $0 < k_1 < j+1$;
- $SG_3 : ((j-k_2)(mb-\varepsilon)+(mb-\varepsilon), k_2a+(j-k_2)(1-m)a+(1-m)a) = ((j+1-k_2)(mb-\varepsilon), (k_2m+(j+1)(1-m))a)$, $0 < k_2 < j+1$.

Observe that the payoffs of the game with $j + 1$ stages can be obtained by replacing $j$ with $j + 1$ in the payoffs of the game with $j$ stages. As this also holds for $j = 1$, these payoffs hold for any $j$ by induction.

Now we can solve the game with $n$ stages. Applying backward induction to this game, we obtain three SPNE:

- For $m < \frac{C_{ss}}{(n-k_1)a}$, there is a unique SPNE that corresponds to the outcome of the $SG_1$, where $ISP$ always diverts clicks. The SPNE strategy set is (D,AAA) in every stage game.
- For $\frac{C_{ss}}{(n-k_1)a} < m < \frac{C_{ss}}{a}$, there are two SPNE. In the first SPNE, that corresponds to $SG_2$, $ISP$ diverts clicks for $k_1$ stage games, where $k_1 > (n-k_2) + \frac{\varepsilon}{mb-\varepsilon}$ and $0 < k_1 < n$, $AS$ secures the website in $k_1 + 1$st stage game and the system operates as in the nominal mode till the end. The SPNE strategy set in the first $k_1$ stage games is (D,AAA), in the stage game $k_1 + 1$ the strategy set is (D,SAA), and in every stage game till the end the strategy set is (A,AAA). In the second SPNE, that corresponds to the outcome of the $SG_3$, the system operates as in the nominal mode for the first $k_2$ stage games, where $k_2 = \lceil \frac{nma-C_{ss}}{ma} \rceil$ and $0 < k_2 < n$, followed by $ISP$ diverting clicks till the end. The SPNE strategy set in the first $k_2$ stage games is (A,SAA) and (D,AAA) in the last $n - k_2$ stage games.

To obtain the results in Section 5.3, we chose the values of $k_1$ and $k_2$ as follows. In $SG_1$ and $SG_3$, $AS$ does not implement security, therefore $k_1 = 0$. In $SG_2$, as $ISP$ does not divert clicks after $AS$ implements the secure solution, we need to set $k_2 = n$. In $SG_3$, the choice of $k_2$ is determined by the threat of $AS$ implementing security. In a given stage game, $AS$ compares the cost of securing a website:

$$(n-k_1)a - C_{ss} \tag{20}$$

to the loss in revenue from diverted clicks:

$$k_2a + (n-k_2)(1-m)a \tag{21}$$

For $m < \frac{C_{ss}}{(n-k_2)a}$ the loss of revenue is smaller than the cost of securing a website and $AS$ lets $ISP$ divert a fraction of clicks in every stage game from $k_2$ till $n$. ■