Comparison of Search Engines Non-Neutral and Neutral Behaviors

Pierre Coucheney, Patrick Maillé, Bruno Tuffin

Comparison of Search Engines Non-Neutral and Neutral Behaviors

Pierre Coucheney
Inria Rennes Bretagne Atlantique
Campus Universitaire de Beaulieu
35042 Rennes Cedex, France
Pierre.Coucheney@inria.fr

Patrick Maillé
Télécom Bretagne
2, rue de la Châtaigneraie
5576 Cesson Sévigné Cedex, France
patrick.maille@telecom-bretagne.eu

Bruno Tuffin
Inria Rennes Bretagne Atlantique
Campus Universitaire de Beaulieu
35042 Rennes Cedex, France
Bruno.Tuffin@inria.fr

ABSTRACT

Network neutrality has recently attracted a lot of attention but search neutrality is also becoming a vivid subject of discussion because a non-neutral search may prevent some relevant content from being accessed by users. We propose in this paper to model two situations of a non-neutral search engine behavior, which can rank the link propositions according to the profit a search can generate for it instead of just relevance: the case when the search engine owns some content, and the case when it imposes a tax on organic links, a bit similarly to what it does for commercial links. We analyze the particular (and deterministic) situation of a single keyword, and describe the problem for the whole potential set of keywords.

Categories and Subject Descriptors

C.2.3 [**Computer Systems Organization**]: Computer-Communication Networks—*Network Operations*

General Terms

Economics, Neutrality, Search engines

1. INTRODUCTION

Network neutrality has become a very hot topic in the past few years [7, 12], at the same time from political, economic, daily-life points of view, because it may refashion the Internet and in general the telecommunications vision and future. Basically, the principle is that Internet service providers (ISPs) complain that major content providers (CPs) have their traffic flowing through their networks, are resource consuming, and do not pay any fee for that while CPs' revenue represents an important (and increasing) part of the total network-related money. Because of that, ISPs threat to ask side payments to those CPs, to cut their access, or to downgrade their quality of service. This would violate the neutrality principle, stating that all consumers are entitled to reach meaningful content, and that packets should not be differentiated. That stance has received a lot of protests from CPs and users associations claiming, among other reasons, that it would be an impeachment of freedom of speech. This debate induces a lot of challenges from a modeling and analysis point of view, see [1–3, 6, 11], without being exhaustive on the list of relevant references.

Though, while we agree about the importance of discussing the neutrality of ISPs, we believe that some actors, namely search engines, are strangely forgotten in the debate (see also [4] for some elements on this). Indeed, one has to notice first that search engines play a key role in the Internet, since in most cases end users use them by dialing a keyword in order to reach the most relevant related content. A biased search engine could somewhat "cut" CPs from end users if it deliberately omitted them in the list of displayed links. Second, search engine advertising has become an important business, the combined revenue of the two main actors in the area, Yahoo! and Google, being more than \$11 billion in 2005 for instance [13], and this business is expected to count for about 40% of total advertising revenue [8]. Those revenues are obtained thanks to clearly declared sponsored links that can be found usually at the top or at the right of the page detailing the results of the search [9]. But what made the success of search engines (for example Google's breakthrough) is the relevance of the ranking of so-called *organic* links, i.e., non-commercial links displayed by search engines and ranked according to their appropriateness related to the keyword(s). In this case, a neutral ranking, based only on relevance is expected. A non-neutral search engine could push, in organic rankings, some paying content ahead of more relevant links. Another possibility for a search engine would be to favor some content that it owns. This kind of behavior is typically what has been recently reproached to Google, accused to favor Youtube content[1]; Google's CEO was forced to testify in front of the US senate, and is facing to be dismantled because of that. Remark that favoring CPs paying to be ranked higher or favoring your own content means in both cases favoring content bringing more revenue to you.

We aim in this paper at designing a model to understand a non-neutral behavior and to compare it with a neutral one. This is to our knowledge the first paper in that direction (in terms of modeling and analysis). While being non-neutral clearly brings additional revenue to the search engine, we also claim and assume that it would or will limit the number of searches because end users will be less satisfied (for example, Google would have been less successful

[1]See for instance `http://www.guardian.co.uk/technology/blog/2011/sep/21/eric-schmidt-google-senate-hearing`

if non-neutral at its birth, because the search results would not have been perceived as good).

We therefore describe in Section 2 the basic assumptions of our model. Section 3 compares the outcomes of a neutral and non-neutral behavior (for two situations of non neutrality: when the search engines owns some content, and when it imposes a tax on the content providers) when considering a single keyword, everything being deterministic in that case. Section 4 poses the problem when requests are for whatever keyword: in that case, there is a distribution over relevances and gains of providers (corresponding to keywords), and the optimal policy problem is presented as a dynamic programming one, with the non-classical issue that the "rewards" or "costs" do not depend only on the current state and transition, but on the whole policy.

2. MODEL DESCRIPTION

2.1 Users

We consider a rate of requests $\lambda(r)$ per unit of time that depends on r, the (average) relevance of displayed links (to be detailed later). λ is assumed to be an increasing function.

2.2 Content providers

Consider a typical search, and a set of m CPs (the requests can be focusing on a single keyword) that are related to (or interested in) the keyword. The relevance of CP i is R_i, and its average gain per click made by CP i is G_i.

2.3 Search engine

The search engine is assumed to display all possible CPs (no limited amount of slots; they may be displayed on a sufficient number of pages as currently done), i.e., the m CPs.

But the impact of the ranking is through the click-through-rate (CTR), assumed decreasing in the ranking. Let θ_j be the CTR of rank j, with $\theta_1 > \theta_2 > \cdots > \theta_m$.

We consider a single keyword. $\lambda(r)$ is the arrival rate of requests for that specific keyword. There are m content providers interested in it with the characteristics described above. The relevance of the ranking can be defined as

$$r = \sum_{i=1}^{m} \theta_{\pi(i)} R_i$$

where $\pi(i)$ is the rank of CP i.

In the neutral case, the ranking is based on relevance only (the R_i), while in the non-neutral case, it is based on the revenue it brings to the search engine (SE).

2.4 Utility functions

Each CP i (in the neutral case, and each CP not owned by the SE in the non-neutral case) is naturally assumed to make an average revenue per unit of time from clicks (it can be through direct sales or thanks to advertisement; we do not care and take it general)

$$U_i^{\mathrm{CP}} = \lambda(r)\theta_{\pi(i)} G_i$$

with $\pi(i)$ the rank of CP i.

For the SE: there is the neutral revenue coming from sponsored links: $h(\lambda(r))$ where h can easily be assumed linear, or increasing and concave (the more visits, the more chances advertisement slots are clicked). This is the only component in the neutral case. In the non-neutral case, we can have two options:

1. a number of links $m' < m$ are from the SE itself (as Youtube for Google), assumed w.l.o.g. to be $\{1, \ldots, m'\}$, and each time the link is clicked, a gain G_i is obtained, leading to an additional revenue $\sum_{i=1}^{m'} \lambda(r)\theta_{\pi(i)} G_i$.
2. We could also assume a behavior even "worse" by the SE, who puts only "sponsored" links in the organic ranking and gets a proportion α of the CP revenue: $\alpha \sum_i \lambda(r)\theta_{\pi(i)} G_i$.

About users, we can remark similarly to [10] that the rate $\lambda(r)$ can be reformulated as $\lambda(r) = \Lambda\Phi(r)$ with $\Lambda = \lambda(\infty)$ the maximal arrival rate, corresponding to the request rate for a maximal relevance level, and $\Phi(r)$ the probability that an arriving request accepts a relevance r, i.e., requires r or less (with independent relevance levels; in steady-state only requests with an appropriate relevance output are assumed to be sent). Conversely, if we have a rate λ, it corresponds to requests asking for an average relevance level up to $\Phi^{-1}(\lambda/\Lambda)$. If ϕ is the density associated to distribution Φ, the total relevance got from users (corresponding to a total user satisfaction) is

$$\mathrm{UR} = \int_0^r y\phi(y)dy.$$

3. NUMERICAL COMPARISON

As an illustration, consider $m = 3$ relevant links, with $(G_1, R_1) = (3, 2)$; $(G_2, R_2) = (2, 3)$ and $(G_3, R_3) = (1, 1)$. Let also $\theta_i = 1/2^i$, and

$$\lambda(r) = \sqrt{\min(r, 4)}.$$

The corresponding density is

$$\phi(y) = 1/(4\sqrt{y})\mathbb{1}_{\{y\in(0,4]\}}$$

and for a $r \in [0, 4]$, $\mathrm{UR} = r^{3/2}/6$. We finally assume a linear revenue from sponsored links for the search engine, $h(\lambda) = \beta\lambda$.

Then, a neutral engine would rank first CP 2, then CP 1 and CP 3, leading to

$$r = \sum_{i=1}^{3} \theta_{\pi(i)} R_i = 17/8,$$

and the request rate is $\lambda = \sqrt{17/8} \approx 1.4577$ and $\mathrm{UR} \approx 0.5163$. The revenue of the engine from sponsored links is $\beta\sqrt{17/8}$. The CPs' revenues are $U_1^{\mathrm{CP}} \approx 1.0933$, $U_2^{\mathrm{CP}} \approx 1.4577$ and $U_3^{\mathrm{CP}} \approx 0.1822$.

Consider now the cases when the engine seeks to be non-neutral.

1. *The engine does not own any content related to the search, but gets a proportion α of the revenue of each CP as a tax.*

   Then its revenue is

$$\begin{aligned} U^{\mathrm{SE}} &= h(\lambda(r)) + \alpha\lambda(r)\sum_{i=1}^{3}\theta_{\pi(i)}G_i \\ &= \sqrt{\sum_{i=1}^{3}\theta_{\pi(i)}R_i}\left(\beta + \alpha\sum_{i=1}^{3}\theta_{\pi(i)}G_i\right). \end{aligned}$$

In full generality, we have six possible rankings to look at, but clearly ranking CP 3 as the last is the best option since it is the least relevant and making the smallest profit. This leaves us with two possibilities, CP 1 shown first, so that $U^{\mathrm{SE}} = U_1^{\mathrm{SE}} = \sqrt{15/8}(\beta + 17\alpha/8)$ or CP 2 shown first (the neutral ranking too), so that $U^{\mathrm{SE}} = U_2^{\mathrm{SE}} = \sqrt{17/8}(\beta + 15\alpha/8)$. Therefore a non-neutral ranking may happen if $U_1^{\mathrm{SE}} > U_2^{\mathrm{SE}}$, that is if and only if $\beta < \alpha(17\sqrt{15/17}/8 - 15/8)/(1 - \sqrt{15/17}) \approx 1.996\alpha$. In other words, and logically, if the revenue made from a unit of request is not large enough with respect to the ratio of revenue got from CPs, the engine has an interest to be non-neutral. We are able to characterize here the threshold. Without loss of generality, let us fix $\beta = 1$. We can draw respectively in Figure 1 and Figure 2 the CPs' and engine's revenues and UR in terms of α, compared with the fully neutral case ($\alpha = 0$).

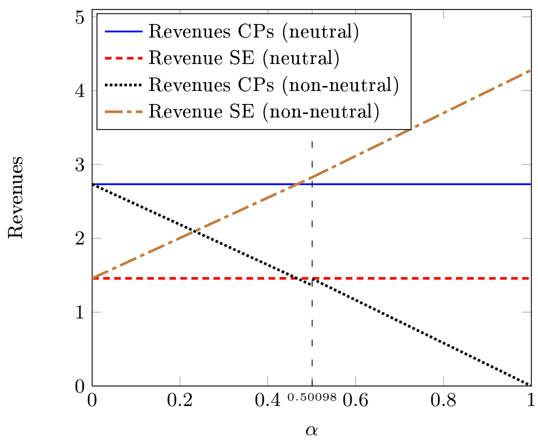

Figure 1: Revenues when the SE imposes a tax

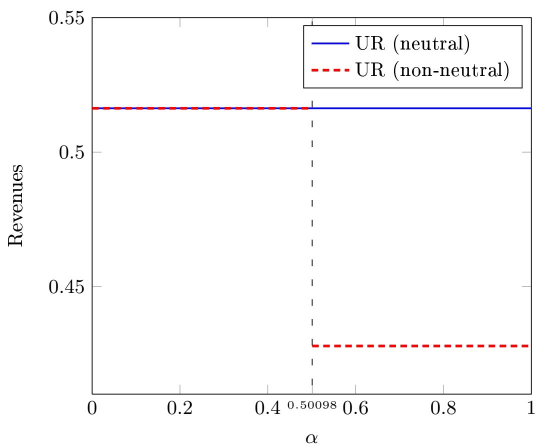

Figure 2: UR when the SE imposes a tax

One can note on Figure 1 that the engine always gains by being non-neutral; while the CPs lose, up to the point where all the money goes to the engine when it gets the whole revenue ($\alpha = 1$). There is interestingly a discontinuity in some of CPs' revenues at the threshold when the ranking changes. From Figure 2, the aggregated relevance is first the same in the neutral and non-neutral cases (for $\alpha < 0.50098$) because the ranking is the same and all the impact is on CPs, but is reduced as soon as the tax on CPs is above the threshold. In that case, both users and CPs lose from non-neutrality.

2. *CP 1 is owned by the engine.*

The engine earns money from sponsored links, but also from the content, so that its total revenue is

$$\begin{aligned} U^{\mathrm{SE}} &= h(\lambda(r)) + \lambda(r)\theta_{\pi(1)}G_1 \\ &= (\beta + 3\theta_{\pi(1)})\sqrt{\sum_{i=1}^{3}\theta_{\pi(i)}R_i}. \end{aligned}$$

The neutral case leads to $U^{\mathrm{SE}} = \sqrt{17/8}(\beta + 3/4)$. But displaying first CP 1 would produce as revenue $U^{\mathrm{SE}} = \sqrt{15/8}(\beta + 3/2)$. Hence, we are here too able to derive the max threshold on β such that a non-neutral ranking would be preferred: if $\beta < 3/2(\sqrt{15/17} - 1/2)/(1 - \sqrt{15/17}) \approx 10.8633$. Above this value, the revenue from sponsored links is too large to favor the CP. We can draw, in terms of β, the CPs' and engine's revenues (Figure 3), as well as UR (Figure 4). Above $\beta = 10.8633$, the system is neutral.

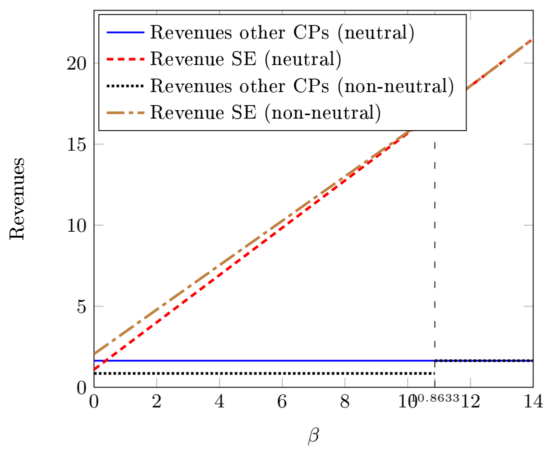

Figure 3: Revenues when the SE owns CP 1

4. THE CASE OF A GENERAL SET OF KEYWORDS

It is of more general interest to consider $\lambda(r)$ as the total rate of requests, not just for a single keyword, but for all ones. We propose here a model for that case. For that whole set of possible requests, we assume that the search engine does just know when one request arrives, that the set of CPs it will face is such that: (i) each time only m CPs are candidate to be displayed. This could be generalized to a random m, but we keep it like this for sake of simplicity (ii) the relevance of any seen CP comes from a distribution F (independent relevance between CPs) (iii) the gain of any

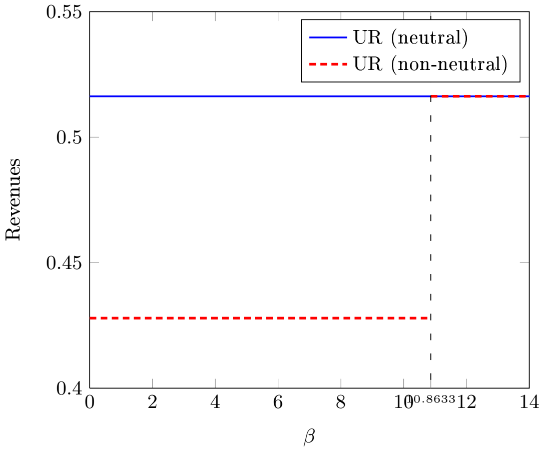

Figure 4: UR when the SE owns CP 1

seen CP comes from a distribution G (independent between CPs).

The SE therefore just knows the general distribution of CPs in terms of relevance and gain, and the independence between CPs. The average relevance r used in $\lambda(r)$ is the average relevance (to be detailed later) seen over the whole set of requests, since in steady-state the rate of requests (user satisfaction) depends on average relevance experienced.

In the neutral case, the average relevance r can be derived from order statistics [5] since the ranking is only based on relevance:

$$r = \sum_{j=1}^{m} \theta_j \int y \frac{m!}{(m-j-1)!\,j!} F(y)^{m-j-1} (1-F(y))^j f(y) dy. \tag{1}$$

The average revenue of a CP is $\mathbb{E}[U^{\mathrm{CP}}] = \lambda(r)\mathbb{E}[\theta_J G]$, where J is the random rank of the CP (uniform on $\{1, \dots, m\}$, which depends only on the relevance). In the case when relevance and gain are independent, we also get in the neutral case $\mathbb{E}[U^{\mathrm{CP}}] = \lambda(r)\mathbb{E}[\theta_J]\mathbb{E}[G]$.

In the non-neutral case, the ranking depends in general on *both* the relevance (through the rate $\lambda(r)$) and the gains. Looking at the case when non-neutrality comes from the SE getting a proportion α of CPs' revenues, the goal of the SE is to determine the optimal family π of permutations, defined for each possible configuration $((R_1, G_1), \dots, (R_m, G_m))$ (*but avoiding the dependence on the values to simplify the notations*), as the one such that

$$\pi = \mathrm{argmax}_{\mathrm{permut.}\ \delta}\ h(\lambda(r)) + \alpha\lambda(r)\mathbb{E}\left[\sum_{j=1}^{m} \theta_{\delta(j)} G_j\right]$$

with r the average over such optimally chosen permutations of relevance. This can be solved using dynamic programming, with the difficulty that the "reward" associated to a transition depends on the optimal policy through $\lambda(r)$, a non-classical assumption.

In the case when non-neutrality comes from the SE owning a CP, the idea is to get the optimal family π of permutations

$$\pi = \mathrm{argmax}_{\mathrm{permut.}\ \delta}\ h(\lambda(r)) + \lambda(r)\mathbb{E}\left[\sum_{j=1}^{m'} \theta_{\delta(j)} G_j\right].$$

Acknowledgement

The authors would like to acknowledge the support of the Euro-NF Network of Excellence through the INNIS specific joint research project, and Inria's collaborative research action MENEUR.

5. REFERENCES

[1] E. Altman, A. Legout, and Y. Xu. Network non-neutrality debate: An economic analysis. In *Proc. of Networking*, pages 68–81, Barcelona, Spain, May 2011. LNCS 6641, Springer Verlag.

[2] F. Boussion, P. Maillé, and B. Tuffin. Net neutrality debate: Impact of competition among ISPs. In *Proceedings of the Fourth International Conference on COMmunication Systems and NETworkS (COMSNETS)*, Bangalore, India, 2012.

[3] P. Coucheney, P. Maillé, and B. Tuffin. Network Neutrality Debate and ISP Inter-Relations: Traffic Exchange, Revenue Sharing, and Disconnection Threat. Rapport de recherche, Inria, Jan. 2012.

[4] J. Crowcroft. Net neutrality: The technical side of the debate: A white paper. *ACM SIGCOMM Computer Communication Review*, 7(1), January 2007.

[5] H. A. David. *Order Statistics*. Wiley, second edition, 1981.

[6] R. Frieden. Internet Packet Sniffing and its Impact on the Network Neutrality Debate and the Balance of Power between Intellectual Property Creators and Consumers. Technical report, SSRN, 2007. `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=995273`.

[7] T. Lenard and R. E. May. *Net Neutrality or Net Neutering: Should Broadband Internet Services be Regulated*. Springer, 2006.

[8] D. Liu, J. Chen, and A. Whinston. Competing keyword auctions. In *Proceedings of 4th Workshop on Ad Auctions*, Chicago, IL, USA, July 2008.

[9] P. Maillé, E. Markakis, M. Naldi, G. Stamoulis, and B. Tuffin. An overview of research on sponsored search auctions. *Electronic Commerce Research Journal*, 2012. To appear.

[10] H. Mendelson and S. Whang. Optimal incentive-compatible priority pricing for the M/M/1 queue. *Operations Research*, 38(5):870–883, 1990.

[11] P. Njoroge, A. Ozdaglar, N. Stier-Moses, and G. Weintraub. Investment in two sided markets and the net neutrality debate. Technical Report DRO-2010-05, Columbia University, Decision, Risk and Operations Working Papers Series, 2010.

[12] A. Odlyzko. Network neutrality, search neutrality, and the never-ending conflict between efficiency and fairness in markets. *Review of Network Economics*, 8(1):40–60, 2009.

[13] H. Varian. Position auctions. *International Journal of Industrial Organization*, 25:1163–1178, 2005.