Fortifying TripAdvisor against Reputation-System Attacks

Conference Paper · December 2014

DOI: 10.13140/2.1.1919.0728

4 authors:

Francesco Buccafurri
Mediterranean University of Reggio Calabria
177 PUBLICATIONS **1,490** CITATIONS
SEE PROFILE

Gianluca Lax
Mediterranean University of Reggio Calabria
107 PUBLICATIONS **600** CITATIONS
SEE PROFILE

Serena Nicolazzo
Mediterranean University of Reggio Calabria
27 PUBLICATIONS **109** CITATIONS
SEE PROFILE

Antonino Nocera
Mediterranean University of Reggio Calabria
63 PUBLICATIONS **484** CITATIONS
SEE PROFILE

Fortifying TripAdvisor against Reputation-System Attacks

Extended Abstract (Work in Progress)

Francesco Buccafurri[§], *Member, IEEE*, Gianluca Lax, *Member, IEEE*, Serena Nicolazzo, Antonino Nocera
DIIES, Università Mediterranea di Reggio Calabria,
Via Graziella, Località Feo di Vito, 89122 Reggio Calabria, Italy
{bucca, lax, s.nicolazzo, a.nocera}@unirc.it
[§]Corresponding Author

***Abstract*—The reputation model of TripAdvisor suffers from several drawbacks making the system vulnerable to users' misbehavior. This results in a significative loss of credibility of the system and possible legal disputes. In this paper, we propose an approach to address the above problem which does not imply an increase in invasiveness or a reduction of usability and guarantees backward compatibility. The approach normalizes scores given by reviewers on the basis of their estimated trustworthiness.**

I. Introduction

TripAdvisor is a travel website collecting reviews of travel-related contents. On the basis of these reviews, an aggregative score of each content is also shown. The reputation model underlying TripAdvisor is robust provided that all actors behave honestly. In this case, the difference between the judge of two different users about an operator (i.e., the subject of the review) derives only from the subjective evaluation pertaining to or characteristic of an individual [1]. However, the assumption that actors behave honestly is not realistic: indeed, one of the main problems of TripAdvisor is the presence of users who pollute the reputation system through malicious behavior either simply for vandalism or, more likely, to take some advantage. The most relevant attacks on the TripAdvisor reputation system can be classified into several categories. *Self-Promoting* attacks concerns the possibility that an operator increases its reputation by fake positive feedbacks. Typically, such attacks are contrasted by excluding positive feedbacks but this solution is not applicable in the case of TripAdvisor because positive feedbacks are already used and provide meaningful information whose elimination would compromise the effectiveness of the reputation system. On the contrary, *Slandering* attacks are carried out to decrease the reputation of other operators (typically competitors) by providing fake negative feedbacks about their services or products. The two attacks above usually exploit accounts of real users who are disposed to write reviews (positive or negative) about some operator, for friendship, money or other advantage. In some cases, such users do not even have a real transaction with the operator. Consider also that a sort of market of feedbacks exists, in which a company collects TripAdvisor users and offers their (fake) reviews to any bidder. Besides this real-account market, there is also the possibility to create fake accounts: indeed, the account registration procedure of TripAdvisor does not implement any functionality to validate the digital identity. In this case, a *Sybil* attack may be carried out to make a fake account seemly licit. To this aim, a sybil attack generates interaction between fake accounts to mutually increase their reputation in the system. Finally, *Whitewashing* attacks exploit the difficulty to establish users identities: in this case, a user, who wasted his reputation, creates a new account to change its reputation to the default initial trustworthiness value [2]. In this paper, we propose to contrast such attacks by normalizing the score given from a user to an operator on the basis of the degree of trustworthiness of the review.

II. Our proposal

Taking into account the considerations given in Section I about the limits of the reputation system of TripAdvisor, we measure the trustworthiness of a review according to two aspects: the trustworthiness of the digital identity of the user and the trustworthiness of the transaction. It is worth noting that our proposal does not imply an increase in invasiveness or a reduction of usability of TripAdvisor because the normalized scores can be applied optionally and selectively determining only different levels of review trustworthiness, but not modifying the standard user's activities.

As for the digital identity, it is evident that many attacks rely on the absence of any degree of trustworthiness of the digital identity of the accounts registered in TripAdvisor. Indeed, to create an account it is required only to provide a mail address, thus, with no limitation in the number of accounts the same user can create. In our proposal, we think of adding two additional options to register a *certified* account. The first option implies to sent a *token* to the user who registers an account via sms or instant messaging systems uniquely associated to a phone number (such as, for example, Whatsapp or Viber), thus introducing the necessity of providing a valid phone number to complete the registration. In the second option, to obtain a *certified* account, we think of using information from social networks, thus associating a TripAdvisor account to an existing profile on selected social networks, in which user identification is guaranteed (e.g., Facebook, Twitter, Google+). This strategy allows us to benefit from the authentication techniques already implemented from the social networks themselves, in which effective checks are carried out to prevent a user can create

fake or multiple accounts [3]Importantly, such social networks are able to detect suspicious activity, such as accessing by a device from an area very far from that in which it was last logged, thus providing a further support against identity theft attacks. Observe that, TripAdvisor allows internetworking solutions, through which the digital identity of a user may coincide with an identity related to a certain profile of a social network. However, currently no difference in terms of trustworthiness is done when a review comes from a 'simple' user or a user logged via social network.

Concerning the second aspect we discuss in our proposal, i.e., the trustworthiness of the transaction, we observe that many of the attacks described in Section I are based on the possibility to make a review about an operator with no guarantee that the reviewer has really interacted with the operator. Think of a TripAdvisor member labelled as an expert by the community, who provides fake reviews to lower the reputation of an activity, possibly very far from places he has visited in the past. This possibility, although difficult to counteract, can be limited by giving the user the option to provide a proof that he visited the activity to review. Moreover, it would be easy for an operator hit by a negative review claiming to be victim of a slandering attack carried out by a competitor operator. This would open a dispute difficult to solve because TripAdvisor does not provides any mechanism to support this process: a user can not provide a proof of his review and an operator can not be sure of not being a victim of an attack by a competitor. The use of evidence in support of their review, although not compulsory to use, allows us to quantify the degree of trustworthiness of a transaction. There are several factors that can be used to make a transaction *certified*. We identified the following ones:

1) The possession of images of the product or service offered. In addition to the words, a reviewer should include a photo of a expensive but inadequate dish or a dirty bath in luxury hotel. Note that TripAdvisor already allows user to insert images, but these images are not used to evaluate the quality of the review. This makes the implementation of our first factor quite immediate.
2) Presence of witnesses. A user could indicate other people who can confirm his presence in an activity. The witness does not concern the trustworthiness of the review content, but only the presence of the reviewer in that activity.
3) A confirmation by the operator. In this case, we expect that the operator issues a *ticket* to the reviewer, and the ticket has to be inserted into the system to provide a review. To contrast the possibility that an operator carries out self-promoting attacks (by generating tickets do not associated with any transaction), the idea is that the reviewer has a sequence of tickets, each generated by an operator who he has visited, but for reviewing the i-th operator, he has to use the ticket supplied from the $(i-1)$-th operator. The first ticket is generated from the system when creating the reviewer account.

Once that we have described how to certify the digital identity and a transaction, we need to quantify the degree of trustworthiness of a review done from an account after a transaction. We explicitly note that the optimal tuning of this degree is outside the scope of this preliminary work and will be the object of a future work which surely will require an experimental validation. Just to give a possible tuning, we think that an initial possibility could be the following. A standard review (i.e., all reviews done in the current reputation system of TripAdvisor) are assigned with trustworthiness equal to 0.50. A review done from a certified account or by a certified transaction has trustworthiness equal to 0.75. Finally, the trustworthiness of a review is 1 when it is done from a certified account by a certified transaction.

The overall score assigned to an operator is computed as the mean of the scores received at each transaction weighted by the trustworthiness of each review. This is equivalent to normalizing the scores given by reviewers on the basis of their trustworthiness.

III. Conclusion and Future Work

This paper is a first step towards a solution of a very topical problem concerning the vulnerability of the TripAdvisor reputation system. The idea underlying our approach is that a number of pieces of information coming from different sources can be used to estimate the trustworthiness of a review also basing on the proof that a transaction actually occurred. The work is in progress. We have to refine the model and, importantly, to validate it on real-life TripAdvisor data. As a first step, we plan to take a snapshot of the reviews concerning an operator at a given year (e.g., 2010) and to infer certified identities as those accounts linked to social networks and certified transactions as that in which the reviewer has provided pictures. The ground truth of the operator reputation is obtained looking at its score on TripAdvisor at a very recent time (possibly, the last year) because we expect that with the increasing of the number of the reviews, the pollution given by fake reviews is negligible. Then, we compare the performance of the reputation model of TripAdvisor with and without normalized scores.

Acknowledgment

This work has been partially supported by the Program "Programma Operativo Nazionale Ricerca e Competitività" 2007-2013, Distretto Tecnologico CyberSecurity funded by the Italian Ministry of Education, University and Research.

References

[1] J. K. Ayeh, N. Au, and R. Law, "Do We Believe in TripAdvisor? Examining Credibility Perceptions and Online Travelers Attitude toward Using User-Generated Content," *Journal of Travel Research*, vol. 52, no. 4, pp. 437–452, 2013.

[2] A. A. Irissappane, S. Jiang, and J. Zhang, "Towards a comprehensive testbed to evaluate the robustness of reputation systems against unfair rating attack." in *UMAP Workshops*, vol. 12, 2012.

[3] Y. Boshmaf, I. Muslukhov, K. Beznosov, and M. Ripeanu, "Design and analysis of a social botnet," *Computer Networks*, vol. 57, no. 2, pp. 556–578, 2013.