Detecting Fake Check-Ins in Location-based Social Networks Through Honeypot Venues

Ke Zhang
University of Pittsburgh

Konstantinos Pelechrinis
University of Pittsburgh

Prashant Krishnamurthy
University of Pittsburgh

1. PROBLEM AND OUR CONTRIBUTION

The proliferation of location-based social networks (LBSNs) has offered many conveniences to its participants, such as place recommendation, tracking of friends, monetary rewards from venues visited and a cheap way of advertisement for local businesses. However, users can misuse the offered features and the major threat for the service providers is that of **fake check-ins**. Users can easily manipulate the localization module of the underlying application and declare their presence in a counterfeit location. The incentives for these behaviors can be both earning monetary rewards as well as virtual rewards. The latter is part of the "gamification" employed by LBSNs, which essentially transforms the LBSN into a mobile game for attracting new users and keeping the existing ones active. Given that people respond to incentives, users may be tempted to generate fake check-ins in order to "beat the game" by earning more virtual rewards or even to just create a virtual geo-social profile [1]. We refer to this type of cheating users as **game cheaters**. Even though game cheaters might not have any direct effect on the LBSN or the participating venues, eventually a large volume of noisy spatial data can significantly degrade services offered from the LBSN providers (such as recommendations) or third parties using these data (e.g., urban planners).

Our work proposes a novel, **honeypot** venue (HV)-based solution, to identify game cheaters. Essentially, this idea originates from the honeypot machines that are traditionally deployed to detect malicious attackers in a computer system/network. In LBSNs, game cheaters are in general attracted by venues that can facilitate their goal for as many as possible virtual rewards. In other words, they do not care for the specifics of the venues as long as the latter satisfy their goals. Hence, the LBSN service provider can create *fake* venues - the honeypots - that appear attractive to game cheaters. Given that under being use of the system no one should be present in that locale, users that check-in to HVs are automatically flagged as (potential) game cheaters. As compared to traditional location proof-based methods [2], our solution does not require the cooperation of third parties (e.g., certification providers, telecom providers etc.) and can be deployed and controlled purely by the LBSN provider. Furthermore, the method is neither hardware dependent, not does it require special hardware at the mobile devices of the end users. Hence, it is ideal for immediate deployment.

2. DESIGN AND MODEL

In general, the design of the HV should be such that maximizes the check-in probability P_{HV} of a cheater at venue HV. A simple **behavioral model** for this probability could be the following. Let us consider C to be the set of appealing features for a gamer cheater. For simplicity let us consider two representative features drawn from the Foursquare paradigm, the largest LBSN to date; the number of possible points n earned from the check-in and the probability m of becoming the "mayor" of the venue, i.e., $C = \{n, m\}$ [3]. Then the probability P_{HV} should be thought as a function of the elements of C: $P_{HV} = f(C) = f(n, m)$.

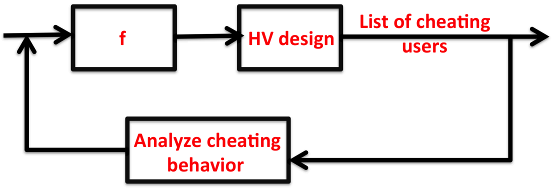

Figure 1: Reinforcement learning of model f.

In the case considered, the function f should clearly be a non-decreasing function with respect to both variables n and m. Of course, the exact shape of f is not known, but it can be reinforced by observing the behavior of identified cheaters. This constant feedback process, will facilitate the design of more effective HVs and is graphically depicted in Figure 1. As we can see the behavioral model f, and as a consequence the deployed HVs, is regularly refined through analyzing the behavioral data of game cheaters that are being detected.

Of course, even honest users can accidentally check-in to a HV. However, these instances are not expected to be excessive. Therefore, a **suspiciousness level** $l(u)$, can be defined for every user u. $l(u)$ can be a function h of a variety of factors such as the number of check-ins to HVs, $q(u)$, of user u and the number of distinct HVs, $r(u)$, that u has checked-in at, that is, $l(u) = h(q(u), r(u))$. By simply defining a threshold L against which we compare $l(u)$, we can make a decision if u is a game cheater or not.

3. CONCLUSION AND FUTURE WORK

In this preliminary study, we propose a novel scheme based on the primitives of honeypots for detecting fake check-ins in location-based services. We introduce the basic system design and model, while a more detailed description can be found in [4]. Our next steps are to build realistic models for the behavior of game cheaters (i.e., identifying function f) by analyzing check-in data and evaluate the detection performance of a prototype system in an actual LBSN.

4. REFERENCES

[1] Why i cheat at foursquare: http://ciaoenrico.com/2010/06/17/cheating-foursquare/.

[2] S. Saroiu and A. Wolman. Enabling new mobile applications with location proofs. *ACM HotMobile*, 2009.

[3] Foursquare's point system: http://aboutfoursquare.com/points-leaderboard/.

[4] Konstantinos Pelechrinis, Prashant Krishnamurthy, and Ke Zhang. Gaming the game: Honeypot venues against cheaters in location-based social networks. *CoRR*, arXiv:1210.4517, 2012.