# LocBorg: Hiding Social Media User Location while Maintaining Online Persona (Vision Paper)

Victor Zakhary, Cetin Sahin, Theodore Georgiou, Amr El Abbadi
University of California, Santa Barbara, CA 93106
[victorzakhary,cetin,teogeorgiou,amr]@cs.ucsb.edu

## ABSTRACT

Social media streams analysis can reveal the characteristics of people who engage with or write about different topics. Recent works show that it is possible to reveal sensitive attributes (e.g., location, gender, ethnicity, political views, etc.) of individuals by analyzing their social media streams. Although, the prediction of a user's sensitive attributes can be used to enhance the user experience in social media, revealing some attributes like the location could represent a threat on individuals. Users can obfuscate their location by posting about random topics linked to different locations. However, posting about random and sometimes contradictory topics that are not aligned with a user's online persona and posts could negatively affect the followers interested in her profile. This paper represents our vision about the future of user privacy on social media. Users can locally deploy a cyborg, an artificial intelligent system that helps people to defend their privacy on social media. We propose LocBorg, a **loc**ation privacy preserving cy**borg** that protects users by obfuscating their location while maintaining their online persona. LocBorg analyzes the social media streams and recommends topics to write about that are similar to a user's topics of interest and aligned with the user's online persona but linked to other locations.

## CCS CONCEPTS

•**Security and privacy** → *Social aspects of security and privacy;*

## KEYWORDS

Location privacy, Online Persona Utility, Content-based Analysis

## 1 INTRODUCTION

Social Media, such as Twitter, Facebook, and Instagram, are prevalent and pervasively reflect everyday activity, communication, interaction, and socializing. Social media users develop on-line persona that reflect their overall interests, activism, and diverse orientations. Many such users have numerous followers that are interested in their postings which reflect this specific persona. However, due to the rise of machine learning and deep learning techniques, it is possible to accurately and automatically predict an individual user's sensitive attributes (e.g., location, sexual orientation, political views, ethnicity, age, etc) based on their posts, likes, and interactions on social network platforms [18, 23, 30]. Facebook likes analysis was able to distinguish between Democrats and Republicans with 85% accuracy [18]. Also, it is possible to predict an individual user's location solely using content-based analysis of the user's posts [6, 7]. These sensitive attributes determine a user's *online persona*[1] and can be used to better serve personalized trending topics, suggest pages to like, accounts to follow, notify users about hyper-local events, and significantly enhance interest-based marketing and directed advertising campaigns.

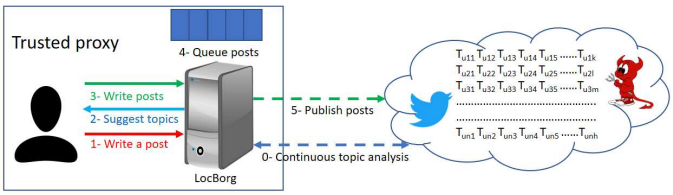



**Figure 1: LocBorg continuously analyzes twitter stream and recommend topics to tweet about in order to obfuscate user location.**

Although predicting a user's sensitive attributes can be used to enhance the user's experience in social networks, revealing some attributes like the location could represent a threat to individuals. An oppressive government can use this location information to arrest or minimally harass political activists. As TechCrunch previously reported [1], Geofeedia is one of a bevy of technologies used, secretly, by police to monitor the locations of activists and the contents of their online discussions. A *#BlackLivesMatter* activist might want to hide her location from the police while continuing to post about topics specifically related to her political movement. A user can try to hide their location by disabling the geo-tagging feature of their posts and hiding the IP address using an IP obfuscation browser like Tor [2]. However, Content-based location prediction models can successfully and accurately predict a user location only based on the content of their posts. A *#BlackLivesMatter* activist who frequently posts about topics that discuss BLM events in New York city is most probably a resident of New York city. Also, it has been shown that user location can be inferred from the locations of followers and friends [13]. Although, it is important to protect user location against both attacks, in this paper, the focus is on protecting the user location only against content-based attacks.

In this paper we present our vision for the future of user privacy on social media. To defend themselves against the large social media providers, with their large computational resources and analytical powers, users too need strong software systems that help them protect their own concerns. We therefore envision that

[1] persona for short.


*SIGSPATIAL'17, Los Angeles Area, CA, USA*

DOI: 10.1145/3139958.3140057

users can locally deploy a cyborg, which is an artificial intelligent software system that helps people defend their privacy on social media. Here we propose our vision for *LocBorg*, a location hiding cyborg that helps social network users achieve location privacy while preserving their persona. As shown in Figure 1, LocBorg is deployed locally at a user's trusted machine and does not require any central service deployment or coordination between different cyborg instances. LocBorg continuously analyzes social network streams and suggests topics of similar interest to the user but linked to different locations. In addition to her own specific postings, a user writes *obfuscating* posts about these different topics to hide her specific location among other users of similar personas but in different locations. This is represented by the dark gray arrows in Figure 2. A *#NoBanNoWall* activist can hide her location by writing posts about the *#NoBanNoWall* protests linked to different locations while maintaining her persona.

Hiding the location of individuals who use Location-Based Services (LBS) has been addressed by many research efforts. Gong et. al [14] use collaborative *pseudonyms* where different users collaborate and synchronously change their identities to hide an individual's exact locations. Brown et. al [5] depend on group collaboration to also hide the location of an individual. However, these techniques are prone to content-based attacks and collaboration between users might not be applicable in the social network context.

Mokbel et. al [22] use location generalization and k-anonymity [24, 26] to obfuscate the exact location of an LBS query. Location generalization typically takes a location query and generalizes it to a range query, which contains a superset of the results of the obfuscated query. A query that asks *"what is the nearest gas station to me?"* can be answered by sending a generalized query *"List all the gas stations in area A that includes my location"*. This approach is oriented towards location based services and can easily be implemented using spatial indexes (e.g. R-Trees). The black and the gray rectangles in Figure 2 are examples of this approach. The larger the generalization box, the more location privacy is achieved, and the more communication and processing overhead added.

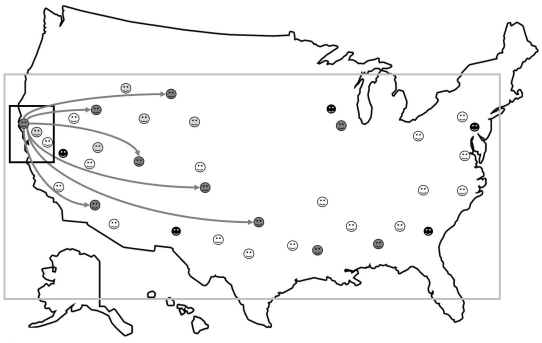


**Figure 2: The black and the gray boxes uniformly generalize the exact location within small and large surrounding areas respectively. The dark gray arrows generalize the location among other locations that have users of similar persona. People with similar persona are represented by the same color.**

Location generalization is quite appropriate for LBS that use spatial indexes. However, applying this same approach to obfuscate social media user locations might be persona unfriendly, especially if applied uniformaly. A user who posts about topics that are linked to her generalized location but not aligned with her persona might not be able to maintain her online persona, and thus lose interest from social media followers. Location generalization introduces a trade-off between location privacy and persona utility and might suggest that a *#NoBanNoWall* activist post about *#BuildTheWall* if *#BuildTheWall* is one of the trending topics within the uniformly generalized location of this activist. LocBorg is radically different from prior LBS privacy preserving approaches, and is specifically designed for hiding location in social media, while preserving a user's online persona. LocBorg generalizes a user location between other locations that have people of similar persona to the user. As shown in Figure 2, people of similar persona are represented using the same color and LocBorg's generalization is shown using the dark gray arrows (the location of *#NoBanNoWall* activists in California represented by the gray color is generalized among other locations that have *#NoBanNoWall* activists). We envision LocBorg to achieve the following goals:

- LocBorg is user centric. It is designed to preserve the user's location while maintaining their persona.
- LocBorg hides a user location against powerful attackers (e.g. the social network service provider).
- The user should have control over the location privacy granularity that needs to be achieved.

## 2 RELATED WORK

Location information not only represents an individual's physical location but also discloses their habits, lifestyle, and maybe personal secrets. Therefore, location privacy has been the focus of a large body of research including publishing a dataset with location information and providing privacy-preserving location based services.

A well-known privacy model, *k-anonymity* [24, 26, 27], and its successors *l-diversity* [21] and *t-closeness* [19] aim to hide location information among a set of indistinguishable locations with different adversarial models. Although these techniques have been applied in many cases, if these locations are not equally likely to be the real location, an attacker might be able to infer the real location, which will obviously violate location privacy. In the context of LBS, Kido et al. [17] and Shankar et al. [25] use similar approaches to hide the exact location information of the user by sending $k-1$ other fake location queries/points. Another approach built on top of *k-anonymity* is the *cloaking* approach which expands the region centered at the user's exact location until it covers $k-1$ other locations/users [4, 8, 22, 28]. Such techniques might lead an adversary to gain more information about the exact location of an individual with high confidence by observing a continuous stream of post publications in social media. Differential privacy [9] is another privacy notion that has been used in the context of location privacy [3, 16, 20]. It ensures that the removal or addition of a single record does not significantly affect the outcome of any analysis. It adds controlled noise to the location information. In this way, the attacker cannot infer the exact location of an individual but an obfuscated location. Similar to the anonymity based solutions, the obfuscated location might still not be hidden in a stream. Recall

that such services still want to achieve some level of location utility since LBS relies on this information and highly obfuscated versions of the exact location would make service totally useless. This limits earlier efforts to achieve the privacy level that we want to achieve in the context of social media.

Social networks have unique data model characteristics, where the data is a stream of posts. Handling such information poses its own challenges since the attacker can have access to many posts, which collectively might reveal sensitive information about the users. Dwork et al. [10] propose privacy preserving algorithms, called *pan-private*, for streaming data. However, the main focus of these algorithms is to deal with attacks where the attacker might be in control of the machine where the algorithm is running but does not have access to the stream, while in our case the attacker has access to every public social post.

In the context of social networks, the earlier works focus on information inference due to the structure of social networks. In [30], Zhelava and Getoor attempt to infer private sensitive attributes using public and private user profiles. However, the authors do not provide any solution to prevent such inference attacks. Later, Yakout et al. [29] proposed a system called Privometer, which measures how much privacy leaks from certain user actions (or from their friends' actions) and creates a set of suggestions that could reduce the risk of a sensitive attribute being successfully inferred. Similar to the Privometer, [15] proposes sanitation techniques to the structure of the social graph by introducing noise, and obfuscating edges in the social graphs to prevent sensitive information inference. In a recent work, Georgiou et al. [11] studies the inference of sensitive attributes in the presence of community-aware trending topic reports. An attacker can increase their inference confidence by consuming these reports and the corresponding community characteristics of the involved users. This work provides a substantial motivation for the need for the solution proposed in this paper.

## 3 LOCBORG DESIGN RATIONAL

Our proposed approach, LocBorg, applies the following steps to hide user location while maintaining on-line persona (see Figure 1):

(0) LocBorg continuously analyzes the global social media stream to detect topics of similar persona to the user.
(1) When a user posts, LocBorg queues and analyzes all attributes and locations that can be inferred from this post.
(2) LocBorg suggests other topics to write about that are similar to the original post but associated to other locations.
(3) The user is required to write about some of the suggested topics and LocBorg queues all the written posts. These additional posts can be considered the cost that needs to be paid to achieve location privacy.
(4) LocBorg publishes the daily queued posts over a day period but at uniformly random chosen intervals.

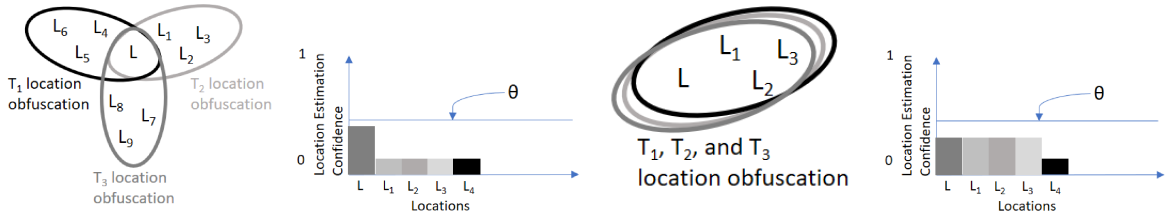
Three fundamental questions need to be answered: *"how does LocBorg identify similar topics to suggest?"*, *"why does LocBorg queue the posts?"*, and *"how does LocBorg choose the locations of the similar topics?"*. We address these specific questions next.

**Similar topic identification:** Users who post about a topic define a multivariate vector of demographics and user attributes. This vector can define a community if most of the users who focus on this topic belong to one community (e.g. *#FreeJustina* was mainly mentioned by female democratic activists who live in Boston, MA). In [12], Georgiou et al. propose a linear scaling algorithm to extract the topics of interest to focused communities. Using the same algorithm, LocBorg uses cosine similarity between attribute vectors to extract the topics of focused communities similar to the user's community but associated to other locations. LocBorg suggests the extracted topics to the user to post about. An alternative is that LocBorg suggests topics similar in the demographics and the user attributes to the topic that the user is currently posting about.

**Timing attack:** if LocBorg allows the user to immediately publish her post $p$ followed by a set of posts, $p_1, p_2, .., p_k$, to obfuscate the location of $p$, an attacker can observe the publishing pattern and cluster the posts into original posts and obfuscation posts. If the attacker can distinguish between the original and the obfuscation posts, the location of the user can easily be inferred by performing location analysis only over the original posts. To overcome the timing attack, LocBorg daily queues all the original and the obfuscation posts and randomly picks a post to publish throughout the day. This prevents an attacker from distinguishing between original and obfuscation posts and hence prevents timing attacks.

**Statistical attack:** assume the user lives in location $L$ and most of her posts are linked to this location. If LocBorg independently obfuscates the location of each post using a set of similar topics associated with a randomly chosen set of locations (as in Figure 2) , an attacker can infer location $L$ using a statistical attack. Figure 3a shows an example of three posts $T_1, T_2$, and $T_3$ linked to location $L$ where $T_1$ is obfuscated using topics linked to locations $L_4, L_5$, and $L_6$, $T_2$ is obfuscated using topics linked to locations $L_1, L_2$, and $L_3$, and $T_3$ is obfuscated using topics linked to locations $L_7, L_8$, and $L_9$. An attacker who analyzes the user's post stream can infer the user's location $L$. As $L$ is a common location for most of the original posts and each original post is obfuscated using a set of posts linked to a randomly chosen set of locations, an aggregate location analysis will report $L$ to be the user's location with a significantly higher confidence than other locations. This happens because $L$ has much higher frequency than the other randomly chosen location. This attack is illustrated using Figure 3b.

To overcome this statistical attack, LocBorg uses a fixed set of *obfuscation locations* $L_1, L_2, ...L_k$ to obfuscate a user's location $L$ as shown in Figure 3c. For every post $T$ linked to location $L$, LocBorg suggests at least k topics, one from each of $L$'s obfuscation locations, to write about. Using a fixed set of locations to obfuscate the user's location allows LocBorg to achieve **k-location-indistinguishability**. Figure 4 shows an example where users from California are obfuscated with users with similar persona in three other states: Montana, Texas and Illinois. k-location-indistinguishability means that an attacker who runs an aggregate location analysis on the user's whole post stream should not be able to distinguish the user's location among a set of k locations. The attacker's confidence about the k different locations should be almost uniform and hence the user's location is hidden among the set of k locations. The attacker's confidence about the user's location is shown in Figure 3d where 4-location-indistinguishability is achieved.




(a) (b) (c) (d)

**Figure 3: Obfuscating the location of each post among an independent set of locations, Figures (3a and 3b), vs. obfuscating the location of all the posts among a fixed set of locations, Figures (3c and 3d).**

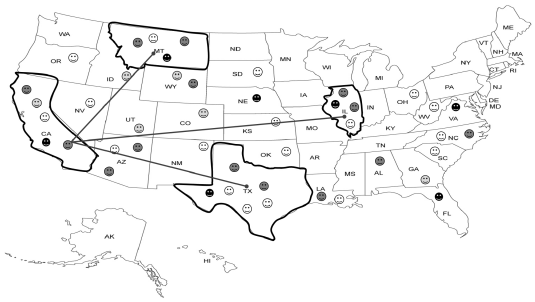


**Figure 4: k-location-indistinguishability**

## 4 CONCLUSION

We envision that social media users will deploy a cyborg to help them in their online privacy battles. In this paper, we presented LocBorg, a vision for a location hiding cyborg. LocBorg is locally deployed on a user's machine and does not require any user group collaboration or any central service deployment. LocBorg hides a social media user's location while maintaining their on-line persona. It continuously analyzes the social media streams and for every user's post, it suggests a set of topics aligned with the user's persona but associated with other locations to post about in order to obfuscate the user's location. LocBorg is designed to protect a user against timing attacks. In addition, to hiding a user's location against statistical attacks, LocBorg uses a fixed set of locations to obfuscate the user's location, thus achieving k-location-indistinguishability.

**Acknowledgments:** This work is partially supported by NSF grant CNS 1649469.

## REFERENCES

[1] 2016. Police are increasingly using social media surveillance tools. https://techcrunch.com/2016/09/23/police-are-increasingly-using-social-media-surveillance-tools/. (2016).
[2] 2017. Tor Browser. https://www.torproject.org/. (2017).
[3] Miguel E. Andrés, Nicolás E. Bordenabe, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. 2013. Geo-indistinguishability: Differential Privacy for Location-based Systems. In *Proc. of ACM SIGSAC (CCS '13)*. ACM, 901–914.
[4] Bhuvan Bamba, Ling Liu, Peter Pesti, and Ting Wang. 2008. Supporting Anonymous Location Queries in Mobile Environments with Privacygrid. In *Proc. of WWW*. ACM, 237–246.
[5] Joshua WS Brown, Olga Ohrimenko, and Roberto Tamassia. 2013. Haze: Privacy-preserving real-time traffic statistics. In *Proc. of ACM SIGSPATIAL*. ACM, 540–543.
[6] Swarup Chandra, Latifur Khan, and Fahad Bin Muhaya. 2011. Estimating twitter user location using social interactions–a content based approach. In *Privacy, Security, Risk and Trust (PASSAT)*. IEEE, 838–843.
[7] Zhiyuan Cheng, James Caverlee, and Kyumin Lee. 2010. You are where you tweet: a content-based approach to geo-locating twitter users. In *Proc.of CIKM*. ACM, 759–768.
[8] Matt Duckham and Lars Kulik. 2005. A Formal Model of Obfuscation and Negotiation for Location Privacy. In *Proc. of PerCom (PERVASIVE'05)*. Springer-Verlag, 152–170.
[9] Cynthia Dwork. 2006. *Automata, Languages and Programming: ICALP 2006, Proceedings, Part II*. Springer Berlin Heidelberg, Chapter Differential Privacy, 1–12.
[10] Cynthia Dwork, Moni Naor, Toniann Pitassi, Guy N. Rothblum, and Sergey Yekhanin. 2010. Pan-private streaming algorithms. In *In Proceedings of ICS*.
[11] Theodore Georgiou, El Abbadi Amr, and Xifeng Yan. 2017. Privacy-Preserving Community-Aware Trending Topic Detection in Online Social Media. In *DBSec*.
[12] Theodore Georgiou, Amr El Abbadi, and Xifeng Yan. 2017. Extracting Topics with Focused Communities for Social Content Recommendation. In *Proc. of ACM CSCW*. 1432–1443.
[13] Mohammad Ghufran, Gianluca Quercini, and Nacéra Bennacer. 2015. Toponym disambiguation in online social network profiles. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 6.
[14] Xiaowen Gong, Xu Chen, Kai Xing, Dong-Hoon Shin, Mengyuan Zhang, and Junshan Zhang. 2015. Personalized location privacy in mobile networks: A social group utility approach. In *(INFOCOM)*. IEEE, 1008–1016.
[15] R. Heatherly, M. Kantarcioglu, and B. Thuraisingham. 2013. Preventing Private Information Inference Attacks on Social Networks. *IEEE TKDE* 25, 8 (2013), 1849–1862.
[16] Shen-Shyang Ho and Shuhua Ruan. 2011. Differential Privacy for Location Pattern Mining. In *Proc. of ACM SIGSPATIAL (SPRINGL '11)*. ACM, 17–24.
[17] H. Kido, Y. Yanagisawa, and T. Satoh. 2005. Protection of Location Privacy using Dummies for Location-based Services. In *(ICDEW'05)*. 1248–1248.
[18] Michal Kosinski, David Stillwell, and Thore Graepel. 2013. Private traits and attributes are predictable from digital records of human behavior. *Proc. of the National Academy of Sciences* 110, 15 (2013), 5802–5805.
[19] N. Li, T. Li, and S. Venkatasubramanian. 2007. t-Closeness: Privacy Beyond k-Anonymity and l-Diversity. In *IEEE ICDE*. 106–115.
[20] Ashwin Machanavajjhala, Daniel Kifer, John Abowd, Johannes Gehrke, and Lars Vilhuber. 2008. Privacy: Theory Meets Practice on the Map. In *Proc. of ICDE*. 277–286.
[21] Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam. 2007. L-diversity: Privacy Beyond K-anonymity. *ACM Trans. Knowl. Discov. Data* 1, 1, Article 3 (March 2007).
[22] Mohamed F Mokbel, Chi-Yin Chow, and Walid G Aref. 2006. The new casper: Query processing for location services without compromising privacy. In *Proc. of VLDB*. 763–774.
[23] Michael J Paul and Mark Dredze. 2011. You are what you Tweet: Analyzing Twitter for public health. *Icwsm* 20 (2011), 265–272.
[24] Pierangela Samarati. 2001. Protecting respondents identities in microdata release. *IEEE TKDE* 13, 6 (2001), 1010–1027.
[25] Pravin Shankar, Vinod Ganapathy, and Liviu Iftode. 2009. Privately Querying Location-based Services with SybilQuery. In *Proc. of UbiComp*. ACM, 31–40.
[26] Latanya Sweeney. 2002. Achieving k-anonymity privacy protection using generalization and suppression. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10, 05 (2002), 571–588.
[27] Latanya Sweeney. 2002. K-anonymity: A Model for Protecting Privacy. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.* 10, 5 (Oct. 2002), 557–570.
[28] Mingqiang Xue, Panos Kalnis, and Hung Keng Pung. 2009. Location Diversity: Enhanced Privacy Protection in Location Based Services. In *Proc. of LoCA*. 70–87.
[29] Mohamed Yakout, Mourad Ouzzani, Hazem Elmeleegy, Nilothpal Talukder, and Ahmed K. Elmagarmid. 2010. Privometer: Privacy protection in social networks. *IEEE ICDEW* 00 (2010), 266–269.
[30] Elena Zheleva and Lise Getoor. 2009. To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles. In *Proc. of WWW*. ACM, 531–540.