Adaptive User Anonymity for Mobile Opportunistic Networks

Milena Radenkovic
School of Computer Science
University of Nottingham, UK
mvr@cs.nott.ac.uk

Ivan Vaghi
EarlyMorning
20127 Milano, Italy
ivan@earlymorning.com

ABSTRACT
Current mobile opportunistic networks often use social routing protocols to transfer messages among users and to the services. In the face of changing underlying topology, mobility patterns and density of users and their queries, fixed algorithms for user anonymisation cannot provide sufficient level of user anonymity, and adaptive mechanisms for achieving user anonymity are needed. This paper describes a novel flexible and adaptive approach, AdaptAnon that is suitable for dynamic and heterogeneous mobile opportunistic networks. Our approach is multidimensional and combines multiple heuristics based on user profiles, analysis of user connectivity and history of anonymisation in order to predict and decide on the best set of nodes that can help anonymise the sending node. In our demonstration, we show that AdaptAnon achieves high quality of anonymisation in terms of both the number of nodes and the diversity of nodes in the anonymisation layer for varying query intensity and over live San Francisco cab mobility traces while neither decreasing success ratios nor increasing latency. We also compare AdaptAnon to other state of the art single dimensional anonymisation approaches and do real time visualization of performance parameters.

Categories and Subject Descriptors
C.2.2 [**Network Protocols**]: Routing Protocols

General Terms
Algorithms, Measurements, Security, Experimentation

Keywords
Mobile opportunistic networks, Anonymity, Adaptive networks

1. INTRODUCTION
There has recently been an intense research on how to design data dissemination protocols within mobile opportunistic networks [1,3]. These protocols are typically based on the assumption that encounters between mobile devices are more likely to occur between people in the same social network than between random strangers[2], or between cars that share the routes than between those that do not[7]. This paper addresses the problem of anonymity in mobile opportunistic routing schemes and argues that using only social network or label similarity for forwarding can be damaging for maintaining sender's anonymity due to routing protocols repeatedly using similar nodes for forwarding and anonymising their messages. In particular, we propose to design a new adaptive anonymisation overlay in mobile opportunistic networks that aims to maximise the quality of anonymisation while maintaining high success ratios of answered queries and low delays. As the underlying density of network may change dramatically and the user interests may also vary, it is important that the anonymisation overlay is responsive both to the underlying topology as well as to the users' interests. We design a multidimensional K-anonymity overlay that hides the senders' identity from the service the sender aims to use. Our approach manages to dynamically and adaptively balance the trade-off between quality of anonymisation, and success ratios and delays of answered queries. We define the quality of anonymisation as a combined measure of the number of nodes and the diversity of the nodes used in the anonymisation path for a given sender and a given service.

Emerging research [4] on characteristics of mobile advertising shows that almost all advertisements are selected based on the users' profiles created over time or recent environmental context and that advertising traffic volume is significantly higher than that of the application traffic.
This paper aims to address this by making users' context and long term profiles less predictable. Our demonstration that runs AdaptAnon over live streamed mobility trace shows that AdaptAnon manages to extend the length of the anonymisation path and to increase the diversity of the nodes in it while not degrading the quality of service.

CHANTS'12, August 22, 2012, Istanbul, Turkey.

2. RELATED WORK
This section gives a brief overview of anonymity approaches in social and mobile opportunistic networks.

[2] consider social network routing that is based on disseminating information about the social network and describes the privacy concerns this introduces. [2] propose two methods for enhancing privacy in social network routing by obfuscating the social network graphs used to inform routing decisions and show that it is possible to obfuscate the social network information without significantly decreasing routing performance.

[10] proposes AnonySense, a privacy-aware system for realizing pervasive applications based on collaborative, opportunistic sensing by personal mobile devices. However, they assume that the nodes who wish to participate in the AnonySense have to register with the registration authority as well as that the IP addresses and certificates of the task service (TS) and the report service are installed on the mobile nodes.

[11] describes SMILE, a mobile social service in which trust is established solely on the basis of shared encounters and anonymous users' ability to prove to each other that they shared an encounter in the past. SMILE uses standard cryptographic primitives that assume existence of trusted third party.

[12] propose constructing a Privacy Analytics framework that uses the Dataware framework [13] to enable querying and

measurements of large public datasets without leaking intermediate results and potentially compromising privacy. [12] aims to verify the query code, and then send it to the user community to perform measurement tasks, collect variable statistics, and perform aggregation and fuzzing while remaining within the community. The proposed Privacy Analytics framework builds privacy from the ground up and enables the user to exercise meaningful choice over participation and what personal information they reveal.

[14] proposes middleware, CAMEO, that uses predictive profiling of a user's device, network and usage context to anticipate the advertisements to be sent, and then modulates their delivery mechanism to enable effective and low cost mobile advertising. [14] reports on the empirical results on advertisement-related traffic characteristics observed during in-lab testing of a selected set of popular advertisement supported free Android applications. Majority of analysed mobile applications were shown to exhibit recurring behaviour i.e. they displayed advertisements every few minutes. CAMEO manages to cache appropriate advertisements in advance because it is able to predict the range of contexts that the user will encounter in the future.

3. ADAPTANON PROPOSAL OVERVIEW

Our novel, flexible, multi-dimensional approach to K-anonymity (AdaptAnon) enables opportunistic identification and selection of the overlay anonymisation nodes in order to allow for better trade-off management between the length of the obfuscation path and the diversity of the nodes on it while not degrading success ratio and delays. We dynamically combine three types of implicit fully localized heuristics that enable better prioritisation of nodes based on their 1) connectivity patterns, 2) user and interest profile similarity, and 3) anonymisation history. The choice of which connectivity, user profile and anonymisation history heuristics to consider and how to combine them determines the effectiveness of our proposal. .

In order to allow responsiveness to changes in the topology, each node performs analysis of node's past interactions and consists of two locally evaluated components: node's social similarity with the next hop and recency of the contact with the service.[1,2,3]. In order to be responsive to changing users' interests and avoid using nodes that are not interested in a certain content, each node analyses the degree of interest and user profile similarity it shares with the nodes that it meets based on the number of matched profile attributes versus the number of total attributes [5].

In order to counterbalance the potential predictability of the previous two heuristics, particularly for the use in social opportunistic networks, we propose the third type of heuristics that is driven by the anonymisation history analysis performed by every node and for every potential anonymising node. This heuristics allows AdaptAnon to increase the diversity of the nodes in the K overlay in order to improve their utilisation. Our aim is to avoid overuse and underuse of some nodes in the anonymisation layer. For example if the source frequently uses the same node(s) for anonymisation, all the nodes become more predictable and more easily profiled (e.g. only 8 messages are sufficient to decide on who the source is [2]) and thus the effectiveness of the anonymisation overlay is significantly decreased. However, when there are multiple sources that are repeatedly utilising the same overlay in such a way that a single source's usage forms a small fraction of the other sources' usage, the effectiveness of using the same nodes by the same source without being easily profiled is higher. Finally, even if one node alone uses the same node(s) repeatedly, but uses them for different services in an unpredictable manner, the quality of such an anonymisation overlay can also be high as it is less predictable and more difficult to profile.

In order to manage the dynamic trade-offs between these different dynamic anonymisation criteria, each node keeps track of the following primitive heuristics: 1) how often a potential next hop has been on the anonymisation path for any source node and for all services i.e. the more popular the node is, the more desirable it is as many other nodes have been using it; 2) how often the potential next hop has been on the anonymisation path for a given source i.e. the more often it has been used by a particular source, the less desirable it is for further usage by that source; 3) and how often the potential next hop has been on the anonymisation path for a particular service i.e. the more often it has been used for a particular service, the less desirable it is for further usage for that service. Based on these primitive heuristics, each node, performs statistical analysis to keep track of the three complex heuristics.

1) the ratio between the number of times the next hop has been used by the given source and by all other sources in order to be able to make less greedy decisions i.e. the lower this ratio is, the more desirable this next hop is as a particular source node is less predictable as shown in formula (1).

$$\frac{AnonymisationCnt_{ThisNode}}{AnonymisationCnt_{Total}} \quad (1)$$

2) the ratio between the number of times the next hop has been on the anonymisation path for the particular Service and for all other services i.e. this is important in order to enable less greedy decisions as shown in formula (2)

$$\frac{AnonymisationCnt_{ThisService}}{AnonymisationCnt_{Total}} \quad (2)$$

3) the ratio between the number of times the next hop has been used to anonymise a particular source for the particular service, and the number of times it has anonymised other nodes for this service as define in formula (3)

$$\frac{AnonymisationCnt_{ThisNode} \cap AnonymisationCnt_{ThisService}}{AnonymisationCnt_{ThisService}} \quad (3)$$

These three ratios allow for adaptive reuse of anonymisation nodes that keeps balance between reusing the same nodes and using nodes that already have experience in providing anonymisation. Each node chooses the next hop node with the maximum total utility (the sum of all heuristics) compared to other encountered nodes, and sends the query to it for anonymisation. We assume that the same respective heuristics without diversification is used for the return path to the source.

4. DEMONSTRATION CONFIGURATION AND PRELIMINARY RESULTS

For our demonstration, we use live updated coordinate based data stream describing the movement of mobile nodes (SF taxi cabs)[6]. Using the Cabspotter API, we retrieve the taxi IDs and initialize them as nodes in ONE. For all nodes we request position and occupancy updates every minute.

The updates are requested via a HTTP GET call and the response for a taxi position update has the following format:

```
<cab id="eoswshy" minutes="1">
       <point cab="eoswshy"
       lat="37.75309"
       lon="-122.39918"
       status="E"
       time="1329437401"/>
</cab>
```

where *id/cab* is the taxi ID, *lat* and *lon* are the geographical coordinates reported by the taxi GPS unit, *status* is the taxi occupancy status, and *time* is the time of the position update recorded as UNIX time stamp.

We then parse the updates, convert them into ONE-specific input format and feed them into ONE via a modified version of the External Reader Module where they update the corresponding nodes positions. Having recreated a real time representation of the taxis movements in the ONE virtual world, we run the AdaptAnon router with 50 meter range WiFi network interface on these nodes in the face of increasing query load by changing numbers of senders to several services. We assess our proposal across a range of metrics that includes success ratio and latency of anonymised answered queries, and anonymisation path length and its diversity factor. We visualize all the performance parameters in real-time. We demonstrate and compare the performance of live AdaptAnon over the live stream of SF taxi cabs movement patterns to three other comparative algorithms: (social connectivity only, label only and mixed social and label) across two other different traces (Sassy[8] and Infocom2006[9]).

Our demonstration shows that live AdaptAnon does not decrease success ratios of answered queries compared to social and label approaches. AdaptAnon marginally increases latency when compared to the social and mixed (combined social and label) approaches as it includes randomization factor that can delay the selection process for the next hop nodes. AdaptAnon has longer anonymisation path than the social and label approaches. AdaptAnon achieves more than two times better higher diversity of the anonymisation layer than either other protocols. Figure 1 summarises the comparison of the quality of anonymisation and diversification across three real traces Infocom2006, Sassy, and live SF Cabs. We observe that AdaptAnon achieves twice as good diversity factor compared to the label, social and mixed approaches while maintaining comparable levels of K.

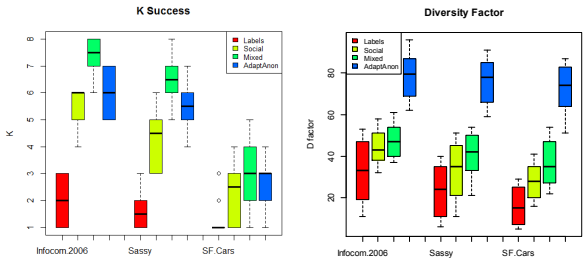

Figure 1. K and D factors

5. CONCLUSION

This paper has two contributions: First, AdaptAnon can achieve higher K (number of nodes in the overlay) compared to non-adaptive K-anonymity approaches with a wide range of query patterns over live mobility data stream while keeping similar success ratios and delays to the state of the art approaches approaches[2][5]. Second, AdaptAnon achieves higher diversity of the nodes in the anonymisation overlay compared to the current single dimensional approaches due to our multidimensional heuristics. We show that AdaptAnon achieves better utilisation of nodes and higher quality of anonymisation even for low Ks (when K is equal to 2 and 3) that are the most realistic achievable Ks for realistic connectivity traces.

Our results could inform the decision on the number of placement of servers that allow different levels of anonymity while providing good services. For our future work, we plan to investigate the performance of AdaptAnon over heterogeneous realistic connectivity traces and the suitability of different models of weightings between the heuristics.

6. REFERENCES

[1] Milena Radenkovic, Andrew Grundy, Efficient and adaptive congestion control for heterogeneous delay-tolerant networks, Ad Hoc Networks, Volume 10, Issue 7, September 2012, Pages 1322-1345, ISSN 1570-8705

[2] I. Parris, T. Henderson: Privacy-enhanced social-network routing. Computer Communications 35(1): 62-74 (2012)

[3] E. Daly and M. Haahr, “Social network analysis for information flow in disconnected Delay-Tolerant MANETs,” IEEE Trans. Mob. Comp, 2009

[4] .A. J. Khan, V. Subbaraju, Archan Misra, S. Seshan, “ Mitigating the true cost of advertisement supported "free" mobile applications”, HotMobile 2012 1

[5] S Zakhary, M Radenkovic, "Utilizing Social Links for Location Privacy In Opportunistic Delay-Tolerant Networks", In the Proc IEEE ICC 2012, Ottawa, Canada

[6] Cab mobility traces, Exploratorium - the museum of science, the cabspotting project, http://cabspotting.org/

[7] Fei Ye; Roy, S.; Haobing Wang; "Efficient Data Dissemination in Vehicular Ad Hoc Networks," *Selected Areas in Communications, IEEE Journal on* , vol.30, no.4, pp.769-779, May 2012

[8] G. Bigwood, D. Rehunathan, M. Bateman, et al, CRAWDAD data set st_andrews/sassy (v. 2011-06-03)

[9] S R Gass and J Crowcroft, P. Hui, et al, “CRAWDAD” dataset cambridge/haggle (v. 2009-05-29)

[10] Cory Cornelius, et al.. Anonysense: privacy-aware people-centric sensing. In Proc. The MobiSys. ACM, New York, NY, USA, 211-224., 2008

[11] J Manweiler, R,Scudellari, and Landon P. Cox. 2009. SMILE: encounter-based trust for mobile social services. In Proc. ACM CCS '09. ACM, New York, NY, USA, 246-255.

[12] H Haddadi, R Mortier, S Hand, I Brown, E Yoneki, D McAuley and J Crowcroft: “Privacy Analytics”. ACM SIGCOMM Computer Communication Review, April 2012

[13] D McAuley, R. Mortier, J. Gouldling, “The Dataware manifesto”, COMCNETS 2011, pp 1-6