Towards Efficient GeoSocial Content Dissemination

Ye Zhao
Dept. of Information and Computer Science
University of California, Irvine
yez@uci.edu

Nalini Venkatasubramanian
Dept. of Information and Computer Science
University of California, Irvine
nalini@ics.uci.edu

ABSTRACT

In this paper, we consider the problem of efficient dissemination of "GeoSocial" contents to users who are interested in (a) events and information within their immediate vicinity, and (b) events and information of others who are socially related to them. We refer to the notification applications as "GeoSocial" applications; key examples include online social networks and social media, geomarketing applications, traffic and weather alert and etc. These applications have become a part and parcel of our lives today. Our approaches [17, 26, 25] cover two key research aspects to enable efficient societal scale dissemination of GeoSocial contents: (a) we designed and evaluated efficient dissemination structures to provide personalized contents to users and (b) we designed techniques to optimize content delivery by considering different network environments and devices the users connect to the system with.

1. CONTEXT AND MOTIVATION

The recent years have witnessed the growth of societal scale notification applications that have penetrated multiple aspects of our day-to-day life. Key examples include online social networks and social media (e.g., Facebook and Twitter), location-based social networks (e.g., Loopt and Foursquare), geomarketing applications (e.g., Groupon and ShopAlert [6]), traffic alerts (e.g., BeatTheTraffic [1]), emergency response (e.g., MyWarn [3]), etc. We have observed a commonality among the applications that the information needs of a user are usually geographical correlated as well as socially correlated: (a) people are often interested in events and information within their immediate vicinity, and (b) people are often interested in events and information of others who are socially related to them. We refer to such information or contents published to users as "GeoSocial" contents. The applications that disseminate GeoSocial contents are refered to as GeoSocial notification applications. Many of the systems are large (e.g. Twitter is estimated to reach 500 million users in 2012 and currently handles over 340 million tweets daily) and geographically distributed. Moreover, with the growing popularity of mobile communications, users connect to the applications using mobile devices (e.g. smartphones, tablets) and expect to derive the same experience no matter how and where they connect.

Several challenges arise in enabling efficient societal scale dissemination of GeoSocial contents: (a) *personalization*: it is important for a notification application to deliver only appropriate contents that a user is interested in to ensure user experience; for GeoSocial contents, this means the need to combine the geographical and social network knowledge of users for dissemination decisions; (b) *timeliness*: many applications are event-triggered and information is expected to be delivered to relevant users in a timely manner (e.g., emergency alerts of a campus fire); (c) *dynamicity*: in GeoSocial notification applications users' interests change dynamically based on their locations or context, it is important for the system to adapt its content dissemination based on such context changes; Moreover, compared to traditional content dissemination applications (e.g., movie dissemination on Netflix), GeoSocial applications are examplified by more dynamic content creation and publication (e.g., alert of a nearby traffic accident or a photo uploaded to Facebook by a friend); (d) *scalability*: since GeoSocial notification applications usually target for common users rather than software agents or enterprises, they must be able to scale to serve a very large number of endhosts. Moreover, GeoSocial contents are examplifed by user-generated contents whose publications are usually magnitutes larger than traditional media contents (e.g., over 300 million photos are uploaded to Facebook every day [5]).

Typical approaches for content dissemination includes centralized client/server model, application layer multicasting (ALM) [16, 13, 14], publish/subscribe systems [12, 22], and P2P systems [10, 19]. These efforts did not consider the geographical and social aspects of the receivers and the contents to improve dissemination efficiency. Location-based publish/subscribe [11] and Geo-aware P2P systems [8, 24, 7] have been proposed to efficiently disseminate information to dynamic geographical regions. However, they did not consider the users' need to receive socially correlated contents. There are also studies [21] to exploit geographic information and knowledge from Online Social Networks to improve content caching in Content Delivery Networks (CDNs). The caching problem is orthogonal to the dissemination challenges we consider.

In this paper, we investigate techniques for efficient and scalable dissemination of GeoSocial contents. Specifically, we study and develop approaches in two research aspects:

- **Content-centric**: techniques developed in the content-centric aspect focus on providing personalized information and contents to users. They can be classified into two categories *publish/subscribe based dissemination* techniques where users receive personalized contents by explicitly express their interests as subscriptions to the notification system and *context based dissemination* techniques where the system decides which information to deliver based on the current context of a user (e.g., his/her location and activity).

- **Delivery-centric**: techniques developed in the delivery-centric aspect focus on optimizing content delivery by taking into account different network environments and devices the users connect to the system with. Content delivery strategies for users with stable broadband access might be different from those for mobile users with sporadic network availability causing intermittent access.

2. CONTENT-CENTRIC TECHNIQUES

We designed and developed two approaches to achieve efficient personalized dissemination of GeoSocial contents such that the applications only deliver appropriate information and contents that a user is interested in. The first one is a context based dissemination technique; it explores the users' geographical and social network context to disseminate information to only relevant recipients. The second one is a publish/subscribe based dissemination technique that relies on users' explicit subscriptions to disseminate information of users' interests. Both techniques build distributed notification systems that are highly scalable and reliable.

2.1 Context Based Dissemination

In this section we consider techniques to personalize content dissemination based on users' context. In our work [17] we are motivated by the disaster alerting and warning scenario where the eventual goal of the notification is to deliver appropriate messages to all relevant recipients with high reliability in a timely manner to prevent loss of lives and property. When a disaster event occurs, the notification system is expected to deliver the event message to recipients who are either (a) located in the event area or (b) socially correlated to the event (e.g., relatives/friends of those who are impacted by the event). For example, if a forest fire occurs near an elementary school, parents will need to be informed of the current situation as soon as possible; information on the current status of their children will help them make decisions on how to respond quickly and effectively (e.g. plan details on how families will evacuate).

To achieve this objective in a scalable and efficient design, we built a fully distributed P2P dissemination framework. Specifically, we construct a GeoSocial notification overlay network that is aware of (a) the geographies in which the message needs to be disseminated and (b) the social network characteristics of the intended recipients. Figure 1 depicts the system architecture. The notification system operates over a Global Target Geography (GTG) that defines the global geographical region over which geo-social notifications occur. The system has prior knowledge of the GTG. The eventual recipients of notifications are real users located

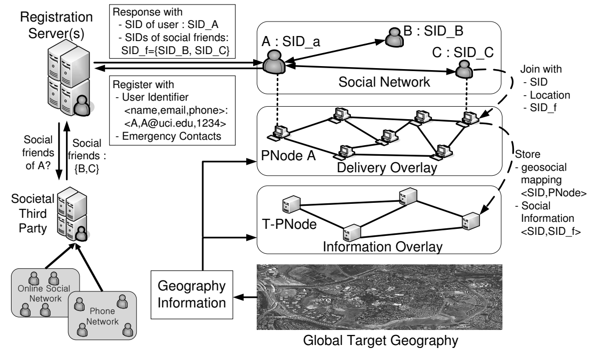

Figure 1: Architecture of the proposed context-based notification system for disseminating GeoSocial contents

inside the global geography interested in receiving the notification and socially connected to each other. Users register the notification application with a registration server and is assigned a unique public Social ID that hides his real identity. At the registration time, the social network information of the user is captured and stored in the overlay network.

The message of an event is tied to (intended for) a specific sub region inside the GTG. We consider the set of users inside the message target region as the geographically correlated recipients and their social friends who are also interested in the message as the socially correlated recipients. To provide personalized notifications we developed two distinct overlays in the system: Delivery Overlay and Information Overlay. The delivery overlay aims to reach geographically correlated recipients efficiently and effectively using a geographical notification technique. On the other hand, the main purpose of the information overlay is to capture and maintain the geo-social information of participating users for efficient lookup of socially correlated recipients. To enable this, we developed a navel peer-oriented geo-aware multicast overlay structure such that: (a) it is self-organizing and fault-tolerant (b) it supports efficient geographical regional multicasting and (c) it supports DHT-style reliable storage of social information.

The delivery and information overlays are created apriori given knowledge of the geography and social networks of individual users. When a message is created, it is conveyed to the geographically correlated recipients by the delivery overlay. To reach the socially correlated recipients, the information overlay explores the stored geo-social information to identify and locate them and forwards the message via the delivery overlay.

2.2 Pub/Sub Based Dissemination

Publish/subscribe has for long been a popular communication paradigm to provide customized notifications to users in a distributed environment. It is nature to consider them as candidates for delivering GeoSocial contents. Pub/sub systems can be classified into two categories: content-based where subscribers receive notifications when the content of the message matches their interest or topic-based when the "topic" of the publication matches their interest. Since the notion of subscriptions for GeoSocial contents are often simple, and instantiated by an average citizen – deals for local shops, traffic alerts for freeways, weather conditions for zipcodes, and location check-ins from friends, we conjecture that a topic-based pub/sub system can form the basis of an efficient architecture to most of the applications.

The challenge to apply pub/sub systems to GeoSocial notification applications lie in the dynamicity of users' interests as subscriptions and the dynamicity of content creation as publications: for instance, the dynamically changing location of a mobile user is a key aspect of what constitutes a GeoSocial subscription and consequently a relevant notification. Existing pub/sub systems [12, 22, 15, 18, 20, 9] assume a relatively static subscriptions from clients and construct the dissemination structure in a self-organizing manner to optimize notification efficiency. Under dynamic subscriptions, the self-organizing approach can negatively impact the system stability and efficiency.

To cope with this challenge we design a dynamic topic-based pub/sub architecture [26] with a different design philosophy where the dissemination structure configuration and reconfiguration are incorporated in a more systematic and planned manner. This is to reduce the reconfiguration overhead from frequent subscription changes that we already know will happen. This is achieved through built-in mechanisms that detect when sufficient changes have occurred and that deal with those changes through planned structure reconfiguration.

Figure 2 shows the pub/sub system architecture. In the proposed system, users connect to brokers for topic subscriptions and content publications. To provide scalable and efficient pub/sub service, the system constructs Rendezvous-based broker network based on structured overlays (e.g., Chord DHT [23]), and maintains independent topic routing trees to route messages of different topics. The centerpiece of the system are two efficient dynamic mapping algorithms in response to users' and brokers' dynamic subscriptions: 1) a simple distributed *similarity-based user placement* algorithm that maps pub/sub users to pub/sub brokers; 2) a *broker network reconfiguration* algorithm running on a Reconfiguration Manager to manage the broker network structure (i.e. mapping of brokers to overlay nodes in logical space).

The aim of the user placement algorithm is to reduce brokers' subscription changes so as to reduce subscription management overhead and alleviate the demand for broker network reconfiguration (described later) when the pub/sub users' subscriptions are dynamic. This is done by dynamically aggregating users with similar subscriptions to be managed together by the same set of brokers. The goal of the reconfiguration is to intelligently adapt the broker network based on updated pub/sub environment to construct efficient topic routing trees. The brokers periodically report their pub/sub states (e.g. number of publications and subscription summaries) to the reconfiguration manager. Based on the received information, the reconfiguration manager performs a cost-driven reconfiguration computation to seek a balance between potential performance improvement and reconfiguration cost from the network reconfiguration. The reconfiguration manager then coordinates the overlay structure configuration through a navel reconfiguration protocol.

3. DELIVER-CENTRIC TECHNIQUES

With growing populairty of mobile communications, more and more users access GeoSocial contents using mobile devices and expect to derive the same experience no matter

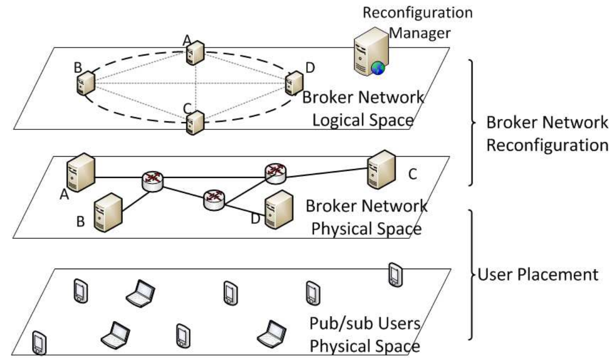

Figure 2: Architecture of the proposed pub/sub system for disseminating GeoSocial contents

how and where they connect. However, several challenges arise in enabling rich GeoSocial contents on mobile devices including: (i) sporadic network availability causing intermittent access (ii) bandwidth limiations in shared wireless media and (iii) high access cost from volume driven data-plans – all of the above result in degraded user experience.

In our work [25], we are motivated by content viewing through social network and social media applications such as Facebook, Twitter and Instagram. They are key example applications for sharing GeoSocial contents. A market study reports that 37% of users accessing rich content, e.g., video, audio, and images, on PCs have switched to mobile devices; moreover, 55% of smartphone users use Facebook on their smartphones [4]. Today, the social media mobile applications, such as mobile Facebook [2], are simple client side implementations that work as dedicated social media browsers. In the absence of connectivity at the time users wish to use the apps, users do not have access to uptodate social media content. We consider the problem of efficient social media access on mobile devices and develop solutions to make the mobile social media experience seamless, personalized, and effective. We achieve this by: (i) ranking the social media streams based on the probabilty that a given user views a given content item, and (ii) investing the limited resources (network, energy, and storage) on prefetching only those social media streams that are most likely to be watched when mobile devices have good Internet connectivity.

A comprehensive solution for these two tasks must account for multiple factores including network conditions, content characteristics, device status, and users' social networks. These factors are typically outside the purview of content providers (e.g., current network conditions for a particular user) or the app at the user side (e.g., characteristics of a specific content item). We developed a mobile middleware system (Fig. 3) that bridges the device OS and network environment to various GeoSocial applications to determine what content to prefetch and when.

Our future work focuses on looking into GeoSocial content dissemination in hybrid network environments (e.g., the coexists of celluar and local WiFi) and collabrative content sharing techniques to improve content reliability and reachability. Moreover, we will conduct further research on content dissemination cross social media networks, i.e., systems support with multiple social networks.

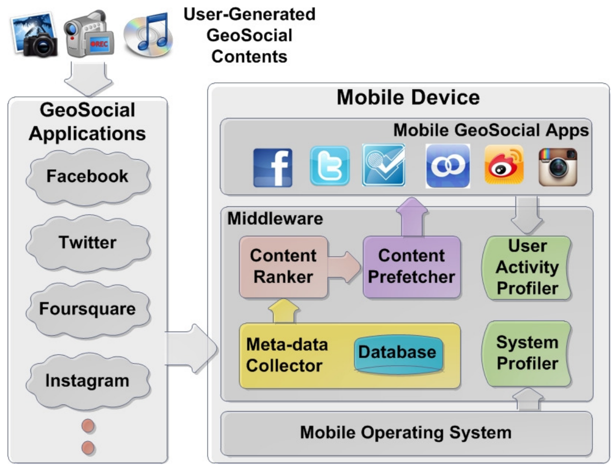

Figure 3: Architecture of the mobile middleware system for GeoSocial content prefetching

4. REFERENCES

[1] Beatthetraffic. `http://www.beatthetraffic.com`.

[2] Facebook mobile app. `https://www.facebook.com/mobile`.

[3] Mywarn. `http://www.mywarn.com`.

[4] Ndp report. `http://tinyurl.com/d3xq7dy`.

[5] Report. `http://www.digitalbuzzblog.com/infographic-instagram-stats/ãĂ`.

[6] Shopalert. `http://shopalerts.att.com`.

[7] J. H. Ahnn, Uichin Lee, and H. J. Moon. Geoserv: A distributed urban sensing platform. In *Proc. of CCGrid*, 2011.

[8] F. Araujo and L. Rodrigues. Geopeer: A location-aware peer-to-peer system. In *Proc. of NCA*, 2004.

[9] R. Baldoni and et al. Tera: topic-based event routing for peer-to-peer architectures. In *Proc. of ACM International Conference on Distributed Event-Based Systems (DEBS'07)*, 2007.

[10] Y. Chawathe, S. Ratnasamy, L. Brreslau, N. Lanham, and S. Shenker. Making gnutella-like p2p systems scalable. In *Proc. of SIGCOMM'03*, 2003.

[11] X. Chen, Y. Chen, and F. Rao. An efficient spatial publish/subscribe system for intelligent location-based services. In *Proc. of ACM International Conference on Distributed Event-Based Systems (DEBS'03)*, 2003.

[12] G. Chockler and et al. Spidercast: A scalable interest-aware overlay for topic-based pub/sub communication. In *Proc. of ACM International Conference on Distributed Event-Based Systems (DEBS'07)*, 2007.

[13] M. Deshpande, B. Xing, I. Lazardis, B. Hore, N. Venkatasubramanian, and S. Mehrotra. Crew: A gossip-based flash-dissemination system. In *Proc. of International Conference on Distributed Computing Systems (ICDCS'06)*, 2006.

[14] M. Ferreira, J. Leitao, and L. Rodrigues. Thicket:a protocol for building and maintaining multiple trees in a p2p overlay. In *Proc. of IEEE International Symposium on Reliable Distributed Systems (SRDS'10)*, 2010.

[15] S. Girdzijauskas, G. Chockler, and et al. Magnet: practical subscription clustering for internet-scale publish/subscribe. In *Proc. of ACM International Conference on Distributed Event-Based Systems (DEBS'10)*, 2010.

[16] K. Kim, S. Mehrotra, and N. Venkatasubramanian. Farecast: Fast, reliable application layer multicast for flash dissemination. In *Proc. of ACM/IFIP/USENIX International Middleware Conference (Middleware'10)*, 2010.

[17] K. Kim, Y. Zhao, and N. Venkatasubramanian. Gsford: Towards a reliable geo-social notification system. In *Proc. of IEEE International Symposium on Reliable Distributed Systems (SRDS'12)*, 2012.

[18] M. Matos, A. Nunes, R. Oliveira, and J. Pereira. Stan: exploiting shared interests without disclosing them in gossip-based publish/subscribe. In *International workshop on Peer-To-Peer Systems (IPTPS'10)*, 2010.

[19] J. Pouwelse, P. Garbacki, D. Epema, and H. Sips. The bittorrent p2p file-sharing system: Measurements and analysis. In *International workshop on Peer-To-Peer Systems (IPTPS'05)*, 2005.

[20] F. Rahimian, S. Girdzijauskas, A. H. Payberah, and S. Haridi. Vitis: A gossip-based hybrid overlay for internet-scale publish/subscribe enabling rendezvous routing in unstructured overlay networks. In *Proc. of IEEE International Parallel and Distributed Processing Symposium (IPDPS'11)*, 2011.

[21] S. Scellato, C. Mascolo, M. Musolesi, and J. Crowcroft. Track globally, deliver locally: Improving content delivery networks by tracking geographic social cascades. In *Proc. of International World Wide Web Conference (WWW'11)*, 2011.

[22] V. Setty, M. Steen, R. Vitenberg, and S. Voulgaris. Poldercast: Fast, robust and scalable architecture for p2p topic-based pub/sub. In *Proc. of ACM/IFIP/USENIX International Middleware Conference (Middleware'12)*, 2012.

[23] I. Stoica and et al. Chord: A scalable peer-to-peer lookup service for internet applications. In *Proc. of SIGCOMM'01*, 2001.

[24] J. Zhang, G. Zhang, and L. Liu. Geogrid : A scalable location service network. In *Proc. of International Conference on Distributed Computing Systems (ICDCS'07)*, 2007.

[25] Y. Zhao, N. Do, S. Wang, C. Hsu, and N. Venkatasubramanian. O2sm : Enabling efficient offline access to online social media and social networks. In *in submission*, 2013.

[26] Y. Zhao, K. Kim, and N. Venkatasubramanian. Dynatops: A dynamic topic-based publish/subscribe architecture. In *Proc. of ACM International Conference on Distributed Event-Based Systems (DEBS'13)*, 2013.