

Location-Based Notification as a General-Purpose Service

Jonathan P. Munson
IBM T.J. Watson Research Center
19 Skyline Dr.
Hawthorne, NY 10532
jpmunson@us.ibm.com

Vineet K. Gupta
Department of Computer Science
University of Illinois at Urbana-Champaign
Urbana, IL 61801
vkgupta@uiuc.edu

ABSTRACT

An often-discussed mobile commerce application is proximity-based coupon delivery. In a typical scenario a merchant is notified when a valued customer is within some distance of a retail outlet, upon which the customer is delivered a coupon or some notice of a special promotion. We believe that this basic mechanism of location-based notification has application far beyond commercial promotion, and is also of interest for tourism, traffic information, public service, and public safety. Furthermore, we believe that a general-purpose location-based notification service that can serve all these applications would be highly useful. However, as we will show, current architectures deployed by wireless carriers to service location-aware applications cannot handle the load of positioning requests implied by such a general-purpose service. We motivate the need for such a service, discuss the limitations of current architectures for providing location information, and detail the requirements for an architecture which would make such a service possible.

Categories and Subject Descriptors

C.2.4 [Computer-Communication Networks]: Distributed Systems—Distributed applications

General Terms

Performance, Design

Keywords

Pervasive computing, mobile commerce, wireless notification

1. INTRODUCTION

An often-discussed mobile commerce application is proximity-based coupon delivery. In a typical scenario a merchant is notified when a valued customer is within some distance of a retail outlet, upon which the customer is delivered a coupon or some notice of a special promotion. This is an example of what we term *location-based notification*, which we define as the act of sending a text or multimedia message to wireless subscribers when they are

determined to be in a particular geographical area. The most popular scenario involving location-based notification seems to be delivering coupons, but we believe that it actually has a wide range of applications. For example, it can be used to:

- Notify a consumer as they enter a shopping center that an office supply store's back-to-school sale is over in two hours.
- Send tourists brief multimedia descriptions of the monuments in the Washington, D.C. mall as they enter each monument's surrounding area.
- Communicate to fair-goers the current events inside the venue they are approaching.
- Notify drivers who enter a certain section of highway that construction two miles ahead is causing a backup and they should take a detour.
- Alert drivers that because of severe fog conditions ahead they should reduce speed immediately.
- Send a message to all subscribers in a certain area that their water supply will be cut off in half an hour to replace a section of pipe.
- Warn a game player that they are entering a "target zone" and are in danger of being abducted by aliens.
- Inform lottery players that they are close to the "pot of gold at the end of the rainbow" and that they should look for someone dressed as a leprechaun.

Thus we see a range of uses, from public safety and public service, to commercial promotion and tourism, to pure entertainment.

1.1 Location-Based Notification as a General-Purpose Service

It would be possible to build a separate, tailored notification system for each one of these applications. This approach, however, is feasible only for those applications that are in and of themselves important enough to warrant a separate system. It is not feasible for most of the applications mentioned above. Therefore, we are researching a general-purpose notification system open to all organizations and businesses who could use such a mechanism to serve their promotional or informational needs. Such a system could attract clients large and small, and the cost of developing and operating it would be spread over many applications. The investment risk is also lowered because no one application is expected to carry the profitability.

To support the applications cited above, a notification system should offer:

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

WMC '02, September 28, 2002, Atlanta, Georgia, USA.
Copyright 2002 ACM 1-58113-600-5/02/0009...\$5.00.

1. The ability to precisely locate subscribers. We assume the use of wireless handsets with ALI (Automatic Location Identification) capability that will be deployed for the Enhanced 911 (E911) emergency location service mandated by the FCC [1]. The system can also be extended to use the precise location ability in the automobile navigation systems now becoming popular.
2. Precisely defined and arbitrarily located notification areas. In the first scenario above, the notification area is defined to be the shopping center's parking lot, so that commuters on a nearby highway do not receive the promotion. In the Washington, D.C. tour guide case, the areas should be such that tourists approaching the Lincoln Memorial do not receive the description of the Washington Monument.
3. The ability to detect entrance to a notification area within a short time period, such as one minute. This implies monitoring the positions of (a.k.a., tracking) the subscribers to whom notifications may be sent.

An important assumption behind our work is that, as the number of applications expand, the number of subscribers who may receive notifications and thereby must be tracked increases, to the point where it is necessary to track each and every subscriber. Thus a general-purpose location-based notification system must also offer:

4. The ability to track each and every wireless subscriber at an interval that will satisfy requirement #3.

As we will later argue, current system architectures for providing subscriber location to applications cannot meet this last requirement. We first discuss some work related to location-based notification. We then present a high-level model that will demonstrate our performance requirements, and will then outline an architecture for proximity detection that we believe will scale to meet our requirements. Finally, we describe the end-to-end architecture of our system.

2. RELATED WORK

The concept of location-based notification is itself not new, and is often generically referred to as “geofencing.” A search on the Web will turn up many companies offering this capability as a fleet-tracking service, typically employing GPS with vehicle-mounted receivers.

Cell-Loc [3] provides location-based notification services known as GeoFence™ and GeoLasso™. GeoFence is a virtual geographic boundary such as one surrounding a factory. GeoFence events are generated whenever one of a specified list of devices are detected entering and/or exiting the area. GeoLasso™ is a one-time specification of an emergency notification area. Once an area has been defined, a list of cell phone users who are active in the area, or have recently been active, will be generated.

In [5], there was a mention of a system developed by TruePosition [11], which “will allow cellular telephone companies to locate their customers and warn them of possible dangers in the area.” Cingular Wireless has signed on to test the service. We were not able to learn more about this service.

Our work exists in the larger context of *location-aware* computing, which itself is part of the field of *context-aware* computing. Space unfortunately does not permit discussion of our work's relation to this body of work.

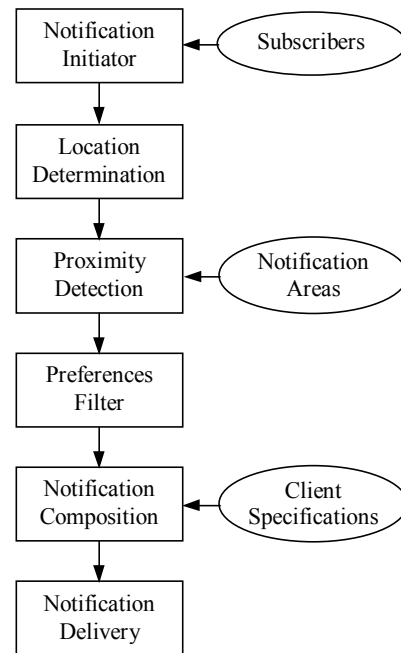


Figure 1. High-level model of location-based notification

3. HIGH-RATE PROXIMITY DETECTION IN WIRELESS SYSTEMS

To discuss the performance requirements of a general-purpose location-based notification system, we decompose the system into the high-level model shown in Figure 1. As mentioned above, we assume the use of ALI-enabled cell phones as the means of determining location.

In Figure 1, “Subscribers” is the set of subscribers to whom some notification may be sent. The Notification Initiator begins the process by sending a subscriber ID to the Position Determination function, which computes the position of the given subscriber using the wireless carrier's position-determining technology—the same technology used for providing E911 service. “Notification Areas” is the set of geographic areas that have been created by clients of the location-based notification service. These would include, from the scenarios above, the shopping center's parking lot, the areas surrounding the Washington Monument and Lincoln Memorial, the known area of dense fog, etc. These are specified as polygons in geographic coordinates. Proximity Detection determines whether a subscriber is in a notification area. The Preferences Filter blocks notifications from being sent to subscribers when their preferences indicate they do not want notifications from a particular client. Notification Composition composes a notification to be sent to a subscriber, based on what the client specified, and the Notification Delivery system actually delivers the notification. This may be the carrier's SMS system, for example.

We apply the model to widely used system architecture for making subscriber positions available to applications. Since we do not wish to focus on any particular wireless carrier's system, we discuss the architecture defined for GSM (Global System for Mobile Communications).

3.1 GSM Location Services Architecture

Figure 2 shows the logical architecture defined by GSM for its LCS (LoCation Services) network feature [2]. To obtain the position of a mobile subscriber, the LCS Client sends an LCS Service Request to the Gateway Mobile Location Center (GMLC), supplying the subscriber's IMSI (International Mobile Subscriber Identifier) or MSISDN (Mobile Station ISDN Number). The GMLC needs to know which MSC is currently serving the subscriber, and so obtains this information from the Home Location Register (HLR). The request is then forwarded to the appropriate MSC. The MSC forwards the request to the Serving MLC, either directly or through the BSS (Base Station Subsystem), depending on how the SMLC is connected.

The SMLC encapsulates the location determination technology, which may be any of the technologies discussed in Section . Handset-based or handset-assisted technologies will make use of the Radio Resource Location Protocol (RRLP) this architecture defines. This will be used to obtain measurements from the handset (handset-assisted) or the position itself (handset-based).

Two observations of the GSM LCS architecture that we would like to make are: (1) it is designed to serve location requests (one-time or tracking) for individual subscribers; (2) it is designed around the requirement that applications have no prior knowledge of a subscriber's location, but know only subscriber's IMSI or MSISDN. Given these two design points, a gateway-based architecture is the logical choice. The GMLC performs the role of determining which MSC/VLR is currently serving the given subscriber (by obtaining this information from the HLR), and forwards the location request accordingly. As long as the request load does not become too high, the GMLC will not be a bottleneck.

3.1.1 Evaluation of Architecture for High-Rate Proximity Detection

The GSM LCS architecture serves the most common applications using location, such as emergency services, navigation applications, field-force automation, and fleet tracking. In these applications the number of subscribers whose position is obtained in a given time period is small relative to the entire subscriber base. Our concept of location-based notification, however, presents an entirely different application model. We postulate that location-based notification has such wide application that the number of subscribers who may be eligible for a notification will be large compared to the number of subscribers that are not eligible. That is, those subscribers who may opt to not receive notifications of a commercial nature, may yet choose to receive notifications about traffic conditions or emergencies.

Under this assumption, the number of requests at interface Le in Figure 2 becomes quite large. To illustrate, suppose a wireless carrier has 20 million subscribers. Suppose half of these have phones that are ALI-capable. Further, we assume under requirement #3 in Section 1.1 that the tracking period is 100 seconds (to use a round number for convenience). Then the rate of requests over Le is 10 million position requests every 100 seconds, or 100,000 position requests per second. This is far beyond a sustainable load with today's technology.

This request rate is not easily distributed over multiple servers because of the dependence on the HLR to direct the request to the appropriate MSC. The actual load on the HLR may be somewhat reduced by caching its responses, but the essential bottleneck

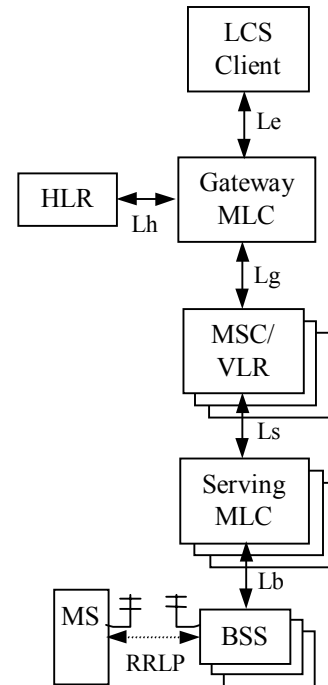


Figure 2. Logical Architecture for GSM Location Services.

induced by a global server for individual subscriber positions remains.

If our assumption holds, that the number of subscribers who may be eligible for a notification will be large compared to the number of subscribers that are not eligible, then the stream of location requests over Le , and the requests to the HLR over Lh , are unnecessary. Given that the MSC/VLR knows what subscribers it is currently serving, it can conceptually generate the request stream itself. This observation is the basis for our high-rate proximity detection model.

3.2 High-Rate Proximity Detection

Our model for high-rate proximity detection is based on the assumption that the number of subscribers who may be eligible for a notification is large compared to the number of subscribers that are not eligible. Thus we initiate location requests at the base station subsystem (BSS) level rather than at the network level. Furthermore, since the notification area definitions are relatively static compared to the subscriber location data, we lower the communications load on the carrier network by pushing this information down to the BSS level as well. The model is shown in Figure 3. For the sake of discussion we assume the context of a GSM network, but our model applies equally to other network types.

We assume an SMLC attached to the BSS. There is a Proximity Detection (PD) node associated with each SMLC. A Query Distributor (QD) node is associated with the MSC. An MSC may serve tens to hundreds of BSCs. A PD node has a list of subscribers that it has received from, and is continuously updated by, the QD. The QD examines the set of currently served subscribers in the VLR and updates the PD nodes according to which BSC the subscriber is served by. For each

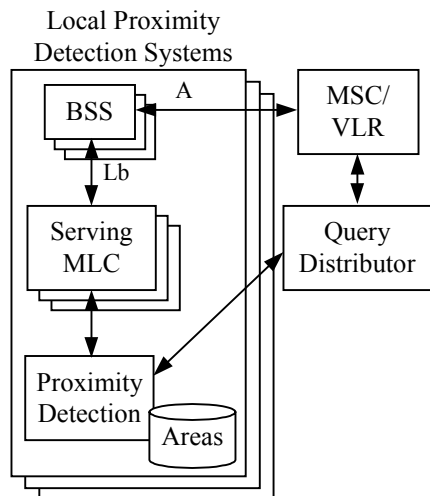


Figure 3. Model for High-Rate Proximity Detection.

subscriber served by its BSC, a PD requests the subscriber's location from the SMLC at periodic intervals.

This scheme bypasses the carrier's network for delivering location requests to the SMLC. The carrier network may not be completely offloaded, however. Depending on the location technology used, there may be messages exchanged by the mobile station and the SMLC. The E-OTD and A-GPS technologies require measurement data to be exchanged. Other technologies rely on passive measurements and so would not require such messages. Further investigation is needed to determine the effects of high-rate proximity detection on this part of the system.

In the next section we describe a high-level, end-to-end architecture for a general-purpose location-based notification system. The high-rate proximity detection model is a component of this system.

4. A GENERAL-PURPOSE LOCATION-BASED NOTIFICATION SYSTEM

In this section we provide a high-level description of a general-purpose notification service. The service has two sets of clients: notification publishers—merchants, the highway department, municipal governments—and wireless subscribers. We assume a simple model where publishers pay to publish notifications and subscribers pay nothing to receive them. In fact, we can imagine incentives to subscribers like a small credit to their wireless phone bill for every notification they receive.

The design of the service is based on several requirements.

- **Specification-based notification.** The service should enable publishers to submit a specification for a notification that defines the notification area and other parameters. Publishers should not be required to provide systems support for the service.
- **Notification in multiple forms.** The system should support specification of the form of the notification, with support for the various forms available within a carrier's system. The forms may include SMS, WAP push, e-mail, or a custom application.

- **Subscriber preferences.** A subscriber should have the ability to block notifications according to the publisher of the notification. He/she should also have some control over the rate of notifications that are sent to him/her.
- **Identity protection:** The notification service provider must not reveal the identity and precise location of the subscriber to the notification publishers. Subscribers may choose to reveal their location only in their response to a particular notification.

4.1 High-Level Architecture

Our architecture for location-based notification (shown in Figure 4) encompasses three primary processes: (1) defining the notification "campaign," (2) detecting subscribers within a notification area, and (3) composing and delivering the notifications to subscribers. shows the logical components of our architecture.

4.1.1 Defining the Notification Campaign

The process of defining a notification campaign proceeds as follows. A merchant or other notification publisher defines a Notification Campaign Specification, including: (1) area of notification (polygon of geographic coordinates), (2) who should receive notifications (perhaps as an identifier of a program a subscriber must have opted in to), and (3) the form of notification (e.g., WAP push, SMS, or other). The Notification Service creates a Notification Campaign ID and returns it to the publisher.

The Notification Service sends Campaign ID and area definition to appropriate Proximity Detection nodes, depending on the location of the notification area. The Proximity Detection node stores the Campaign ID and area definition in its spatial database.

4.1.2 Proximity Detection

We have discussed the proximity detection process in Section 3.2. When a subscriber is determined to be in a notification area, the PD node sends a proximity event to the Notification Processor (NP). The event contains the subscriber ID and the Campaign ID.

4.1.3 Composing and Delivering Notifications

The notification delivery process accommodates different delivery methods and allows publishers to personalize the notifications that are sent.

Before proceeding with the notification process, the NP consults a database of subscriber preferences. (We do not here discuss how subscribers manage their preferences.) We offer a simple, "opt-in/opt-out" model of preferences. Users may choose to receive selected *classes* of notifications (current classes include Emergency, Traffic, and Commercial) or particular notification *programs*. A program is a set of campaigns, and is analogous to an application. The scenarios of the Washington, D.C. tour guide, the action game, and the location-based lottery would be examples of programs. Users may choose *not* to receive notifications from selected publishers. The NP processes a notification if the campaign is in a class or program that the user has opted-in to and is not from a publisher the user has opted-out from.

The NP retrieves the corresponding notification campaign specification for the content and delivery method of the notification (e.g., WAP push [10] or SMS [1]). The campaign specification includes the content (e.g., WML for WAP push or

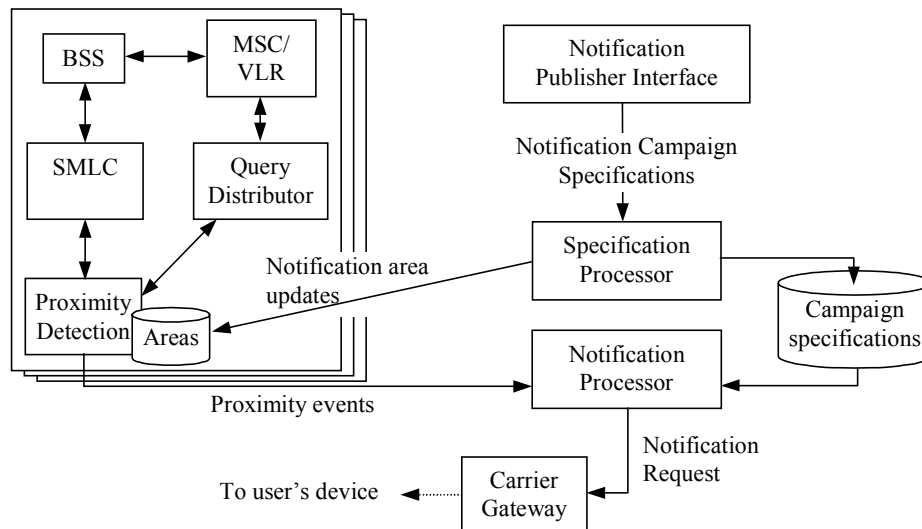


Figure 4. High-Level Architecture of General-Purpose Location-Based Notification System

text for SMS) of the notification. The NP composes the notification according to the publisher's content specification and then sends a request to the appropriate carrier gateway to deliver the notification.

4.2 Implementation

We have developed a first prototype general-purpose notification service. The prototype serves as a testbed for performance evaluations, and as a vehicle for exploring the systems-level issues involved in our distributed proximity detection system.

Initial performance studies, in which we are now engaged, are focused on determining the performance characteristics of our architecture relative to the GSM LCS architecture. Carrier components such as the MSC/VLR are simulated, as is the position-determining technology. For the latter, we are using City Simulator [4] to simulate large numbers of users moving throughout a city. For high performance in proximity detection computations we use the IBM DB2 database with Spatial Extender [9]. For the Notification Processor we are using the Intelligent Notification System [2], a commercial version of which is part IBM's WebSphere family [8].

5. FUTURE WORK

Once our initial performance evaluations are complete, we plan to extend our testbed to allow us to model the resources used by the position determination technology. These resources include not only the processor cycles of the position-determining compute server, but also the radio resources used to communicate with the mobile terminal, as well as the mobile terminal's own resources of processor cycles and battery power. This will allow us to better evaluate the impact of high-rate proximity determination in an actual wireless system.

We have identified certain ways in which our system can be improved with respect to flexibility. One of these is the need for personalization of notifications. For example, a merchant may wish to offer the consumer a promotion tailored to the consumer, or may wish to initiate an interactive session with them. This will

require some interface with the notification system, and we are studying options for doing this.

We also wish to evaluate user response to the system. It will be critical to give users mechanisms to avoid a level of annoyance that will cause them to simply turn off the system altogether. Controlled user studies involving simulated notifications is a technique we are considering.

6. REFERENCES

- [1] 3GPP TS 24.011: "Short Message Service (SMS) support on mobile radio interface".
- [2] V. Bazinette, N.H. Cohen, M.R. Ebling, G.D.H. Hunt, H. Lei, A. Purakayastha, G. Stewart, L. Wong, D.L. Yeh, "An Intelligent Notification System," *IBM Research Technical Report TR 22089*.
- [3] Cell-loc Location Technologies. http://www.cell-loc.com/how_tech.html
- [4] J. Myllymaki and J. Kaufman, "City Simulator: A Spatial Data Generator," IBM alphaWorks emerging technologies toolkit, April 2002.
- [5] CTIA Daily News, May 29th, 2002.
- [6] Enhanced 911. <http://www.fcc.gov/911/enhanced>
- [7] GSM 03.71: "Digital cellular telecommunications system (Phase 2+); Location Services (LCS); Functional Description; Stage 2".
- [8] IBM WebSphere Product Family, <http://www.ibm.com/software/webservers>
- [9] IBM DB2 Spatial Extender, <http://www.ibm.com/software/data/spatial>
- [10] Push Message Specification", WAP Forum™. WAP-251-PushMessage-20010322-a. URL: <http://www.wapforum.org/>
- [11] TruePosition Location Based Services Solutions. http://www.trueposition.com/sol_over.html