# myGander: A Mobile Interface and Distributed Search Engine for Pervasive Computing

[1]Jonas Michel, [1]Christine Julien, [2]Jamie Payton, and [3]Gruia-Catalin Roman
[1]University of Texas at Austin, [2]University of North Carolina at Charlotte, [3]University of New Mexico

Email: {jonasrmichel c.julien}@mail.utexas.edu, payton@uncc.edu, gcroman@unm.edu

## TR-ARiSE-2012-001



ARiSE
Advanced Research in Software Engineering

# myGander: A Mobile Interface and Distributed Search Engine for Pervasive Computing

[1]Jonas Michel, [1]Christine Julien, [2]Jamie Payton, and [3]Gruia-Catalin Roman
[1]University of Texas at Austin, [2]University of North Carolina at Charlotte, [3]University of New Mexico
Email: {jonasrmichel c.julien}@mail.utexas.edu, payton@uncc.edu, gcroman@unm.edu

***Abstract*—Motivated by the growing amount of digitally-accessible information in our physical surroundings and the ephemeral nature of that information, there is a profound need to efficiently search information with spatiotemporal underpinnings without *a priori* indexing. Human users in *Personalized Network Spaces* (PNetS), pervasive computing environments connected by opportunistic peer-to-peer connections, need information that is immediate and localized. This tight integration of the user with his immediate surroundings introduces novel search requirements. The requisite support for performing search *of* the here and now *in* the here and now calls for a new paradigm of search that explicitly separates search from advanced indexing of data. In light of this vision, this paper presents *Gander*, a scalable search engine for PNetS, along with *myGander*, a prototype mobile interface for *Gander*. The utility and usability of *myGander* as supported by *Gander* is demonstrated through a practical real-world pervasive computing scenario.**

## I. Introduction

The envisioned pervasive computing spaces of the near future (e.g., the Internet of Things [1]) will be populated with an immense amount of digital information generated at rapid rates by both humans and digital resources embedded in the environment. *Personalized Network Spaces* (PNetS) are pervasive computing environments in which people carry or wear mobile devices that may form opportunistic peer-to-peer connections to each other and to surrounding digital resources. In PNetS, a human user's need to efficiently search the digital space around him, *here* and *now*, is paramount. As our physical surroundings become increasingly digitally accessible, there is a growing demand for mechanisms that help users find the information they need as they move through densely populated and rapidly changing information spaces.

Users in PNetS need information that is immediate and localized. This tight integration of the user with his immediate surroundings introduces new search requirements. An Internet search engine may be used to track the development of a forest fire; aerial photos and emergency alerts may be useful even if a few hours old. However, when fighting the fire, a volunteer firefighter must know the real-time relation between the fire and wind in the immediate vicinity. Using the Internet, one may find available train routes, whereas a person hurrying to board a crowded train may need to search for the closest available second class seat. When planning a trip to an amusement park, one may use the Internet to find directions, hours of operation, etc. On the other hand, visitors at the park may wish to know which rides have the shortest wait right now or where their friends are. These situations demand expressive capabilities for searching about the *here* and *now*; such information is most effectively collected on-demand, directly from the environment. In other words, search *of* the here and now must be performed *in* the here and now.

This material is based upon work supported by the National Science Foundation under Grant No. 844850. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation (NSF).

We believe that developing the requisite support for performing search directly within data-rich, dynamic, and unpredictable digital environments calls for a new paradigm of search that explicitly separates search from advanced indexing of data. This radical shift is motivated by three key factors, which give rise to unprecedented challenges. First, searchable data in PNetS is often supplied by human users, who are likely unwilling to publicly share information (i.e., on the Internet), but willing to share with other nearby, potentially unknown, users [3], [6]. Second, data in PNetS is *ephemeral*, representing changing real-world conditions of the environment, opportunistic social interactions among humans, etc.; information changes on the order of seconds. Moreover, data in PNetS is generated at a rapid rate; there is no expectation it will be consumed. Large volumes of such transient data cannot easily be centrally indexed and associated with a relative spatiotemporal context, as is traditionally done in Internet information retrieval [4]. Search in PNetS demands a richer awareness of context than just location, yet contextually influenced information retrieval is challenging due to an inability to appropriately index information as it is generated. Therefore, explicit computation and comparison of rich contexts is complex even for simple questions in PNetS. Finally, there are situations in which Internet access is costly, inconvenient, nonexistent, or simply unnecessary; local interactions may be better managed by exploiting local connectivity, devices that comprise PNetS may be far from reliable connectivity, and connectivity to the Internet may be sparse or financially unattainable. This, coupled with the small ratio between the data used and the data generated, necessitates a decentralized and scalable approach to localized search.

Motivated by the above vision, we have developed *Gander*, a personalized search engine for the here and now [5]. *Gander* is founded on a novel conceptual model of search in PNetS and is targeted for real-world environments characterized by

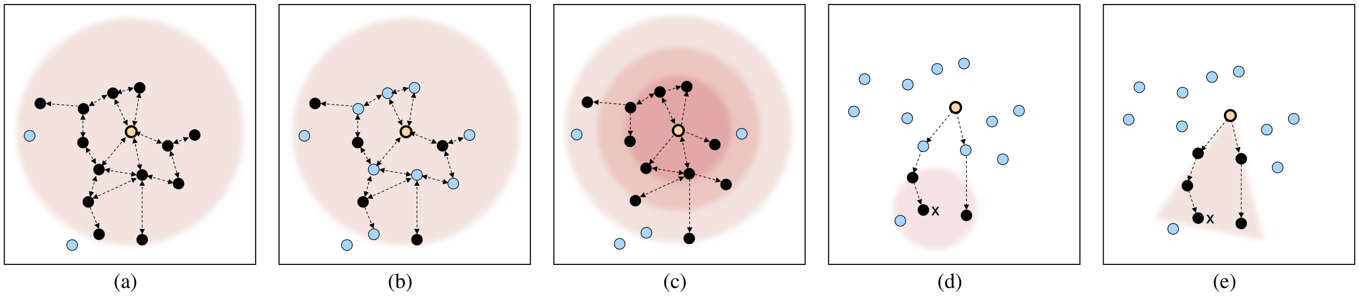


Fig. 1. Query processing styles and sampling. Dashed lines are sent messages. Darkened nodes respond to a given query. (a) Flooding. Every node in a given range (3 hops) retransmits the query; the target area is the shaded region. (b) Random. A receiving node responds to the query with a given probability; a high quality search evenly samples the shaded space. (c) Probabilistic. Every node that receives the query retransmits it with a given probability; the likelihood of reception drops with distance from the query issuer. (d) Regional. The query targets a location centered at "x;" a high quality search has good coverage of the target region. (e) Directional. The nodes in the direction of the target location "x" are sampled; a high quality search follows this search "path."

large volumes of highly transient data. To begin investigating user expectations of such a search engine and the effects of real-world dynamics on search quality in PNetS, we have implemented *myGander*, a prototype user interface and mobile application for *Gander*. The *myGander* prototype enables simple—yet expressive—search of large-scale PNetS.

We overview the design of the *myGander* mobile interface (Section II) and describe a pervasive computing scenario that demonstrates the real-world value of *Gander* (Section III).

## II. Implementation

### A. The Gander Search Engine

In PNetS, people carry or wear mobile devices that may form opportunistic connections to each other and to information and services embedded in the environment, enabling *Gander* queries to be performed using locally available capabilities without the support of an Internet infrastructure. Formal details of *Gander*'s underlying conceptual model are specified in [5], but a brief overview is provided here.

In PNetS, nodes (e.g., smart phones, mobile devices, embedded resources, etc.) may issue queries that are evaluated using information provided by other nodes. A *data item* provides information about the here and now (e.g., a measure of some condition of the environment) and is associated with meta-data that describes its *situation* (e.g., the device(s) that generated it, the location, a timestamp, or even the data's spatiotemporal dynamics, such as volatility or freshness). Data items are generated by and stored at devices distributed in PNetS. Data can be generated, destroyed, changed, and moved arbitrarily; *Gander*'s query model is independent of these processes.

For a given search, every result must "match" the search. A query can also include one or more *constraints*. For example, query resolution may identify data items indicating a *roller coaster*; constraints ensure that the wait times for roller coasters discovered do not exceed a certain threshold. A *relevance metric* compares valid results to each other. A search for roller coasters could favor closer rides or rides near a particular venue. A query can use multiple relevance metrics evaluated independently or using weighted statistics. A *query processing protocol* distributes a query within PNetS. *Gander* can use the query's contents to direct query processing; ultimately the goal is to collect and present only data that is most relevant. From a query's perspective, all data resides in a *global virtual data structure* (GVDS) [7]. Acquiring such a global view of data is infeasible in PNetS; query processing protocols must operate only over locally available data. *Gander*'s query protocols are effectively *sampling* strategies that incrementally construct and adjust a result set by considering additional data items as they are discovered; each strategy efficiently distributes the query (and collects its results) using opportunistic peer-to-peer connections to find the most relevant data. *Gander* provides five styles of spatial sampling, shown in Fig. 1:

**Flooding**: We use a simple protocol constrained by a hop count; every node receiving the query forwards it unless the hop count has been exceeded or the node has forwarded the query before. Every node receiving the query responds.

**Random**: We use the same protocol as in flooding, but nodes respond subject to a probability parameter.

**Probabilistic**: Each node receiving the query forwards it subject to a probability parameter. A node never forwards a query twice. Every node receiving the query responds.

**Regional**: We use greedy forwarding to send a query towards the target location. Any node receiving the query whose location is within a specified radius of the target responds.

**Directional**: We use greedy forwarding with heading and angle parameters; any node receiving the query responds.


Fig. 2. Sample *myGander* Screenshots. (a) Search interface, where a user enters search parameters. (b) *myGander* currently allows the user to select from some built-in relevance metrics. (c) *myGander* stores recently issued queries to simplify their recall. (d) Upon completion of a search, *myGander* shows a list of query results that were collected from the live PNetS, ranked according to the selected relevance metric.

## B. The myGander Mobile Interface

To begin exploring human users' expectations of and interaction with a search engine for PNetS, we developed *myGander*, a prototype front end for *Gander* built on Apple's iPhone Operating System (iOS). Fig. 2 shows sample screenshots of the *myGander* interface. Using *myGander*, users may construct queries by entering a search string, attaching and ordering relevance metrics, and configuring an "appropriate" sampling strategy. Once a query is issued, *myGander* displays the results and the information the user has expressed interest in, per the query's relevance metrics, and dynamically reorders results as they arrive. In our prototype, queries' *ground truth results* (i.e., from a query being evaluated against an omniscient view of the GVDS) may also be viewed in *myGander* to enable users to gauge the quality of the *Gander* results. "Active" queries may be updated (reissued) to potentially acquire fresher results. *myGander* is intended to provide users with the *Gander* search engine in simulated large-scale PNetS.

## C. The Deployment for Simulated PNetS

Experiments with large-scale PNetS are difficult; thus, we integrated *myGander* and *Gander* with the OMNeT++ network simulator [9] to evaluate *Gander*'s capabilities on a large scale (Fig. 3) . We used OMNeT++'s INET framework [2] for networking and SUMO [8] to provide node mobility. Each simulated node executes *Gander*'s query processing logic, which interacts directly with modified versions of INET's MANET routing capabilities to implement our sampling approaches. The PNetS simulator uses unicast routing to send responses. Nodes store their data in a single PostgreSQL database (i.e., the GVDS), which is used to determine the omniscient ground truth result set for a given query. However, when processing queries issued from *myGander*, nodes only have access to their own respective data in the GVDS. Using a wireless connection, any number of *myGander* clients may connect to the PNetS simulator. A proxy server bridges the connection between *myGander* clients and simulated nodes, enabling users to issue queries and receive results from the perspective of a particular mobile node.

# III. Demonstration

To demonstrate *Gander*'s real-world utility, we constructed an amusement park scenario within our PNetS simulator, enabling users to search a realistic pervasive computing space through the *myGander* mobile interface. Our simulated environment uses real data collected about Disney World's Magic Kingdom, including 16 hours of dynamic wait time information, locations and features of attractions and amenities (e.g., ride thrill, restaurant menus, restroom types, etc.), and accurate walking paths. Searchable entities include 30 attractions, 12 restaurants, and 8 restrooms. As simulated visitors walk around the park, their simulated devices collect and carry timestamped data about attractions and amenities that they have recently been near (i.e., within ∼20m). Collected data has a 15 minute lifetime, after which it is discarded.

The demonstration will consist of five iPod Touch devices running the *myGander* interface and a laptop computer running a proxy server and our amusement park scenario within the PNetS simulator. The laptop will display a live map of Walt Disney World and the simulated visitors as they walk around the park (similar to the map in Fig. 3). Attendees will pick a simulated device in the PNetS environment. A provided iPod will be wirelessly connected to the simulated device through the proxy server. Via *myGander*, attendees may then construct queries, attach one or more built-in relevance metrics (e.g., wait time, thrill factor, distance, etc.), tailor a sampling strategy, and issue their queries to the PNetS. When a query is issued, the ground truth result set will immediately be displayed on the iPod using the omniscient GVDS perspective. Attendees may compare this result set with the results dynamically acquired by the *Gander* search engine to gauge the quality of their *Gander* results.

We will demonstrate the expressive search capability of the *Gander* approach by providing several search "campaigns" (tasks) for attendees to perform (e.g., *find the shortest wait time*, *locate the restaurant nearest a ride with a wait time less than 15 minutes*, *plan a valid sequence of three rides nearby constrained by their wait time*, etc.). These campaigns will be designed to illuminate *Gander*'s real-world advantages over Internet-based search in a data-rich, dynamic, and unpredictable environment.

# IV. Technical Requirements

This demonstration will require space for a laptop and two grounded wall outlets to power the laptop and a USB hub.

# References


[1] L. Atzori, A. Iera, and G. Morabito. The internet of things: A survey. *Computer Net.*, 54(15):2787–2805, 2010.

[2] The INET Framework for OMNeT++. http://inet.omnetpp.org/, 2011.

[3] Q. Jones, S. Grandhi, S. Karam, S. Whittaker, C. Zhou, , and L. Terveen. Geographic place and community information preferences. *CSCW*, 17(2–3):137–167, 2008.

[4] C. D. Manning, P. Raghavan, and H. Schutze. *Intro. to Information Retrieval*. 2009.


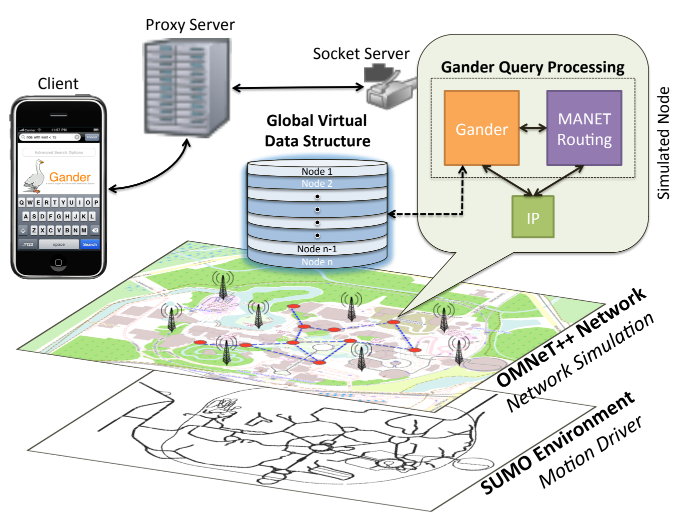


Fig. 3. PNetS Simulator Architecture

[5] J. Michel, C. Julien, J. Payton, and G.-C. Roman. Gander: Personalizing search of the here and now. In *MobiQuitous*, 2011.
[6] D. Mountain and A. MacFarlane. Geographic information retrieval in a mobile environment: evaluating the needs of mobile individuals. *J. of Info. Science*, 33:515–530, 2007.
[7] G. P. Picco, A. Murphy, and G.-C. Roman. On global virtual data structures. In *Process Coordination and Ubiquitous Computing*. 2002.
[8] Simulation of urban mobility. http://sumo.sourceforge.net/, 2011.
[9] A. Vargas. OMNeT++ Web Page. http://www.omnetpp.org/, 2009.