From Location Databases to Pervasive Catalog

Panos K. Chrysanthis *
Computer Science Dept.
University of Pittsburgh
Pittsburgh, PA 15260
panos@cs.pitt.edu

Vladimir I. Zadorozhny †
Info. Sci. & Telecom. Dept.
University of Pittsburgh
Pittsburgh, PA 15260
vladimir@sis.pitt.edu

Abstract

One characteristic of the next generation wireless mobile environments is a high level of interoperability between mobile applications and services. This imposes significant demand on the knowledge maintained by the wireless network about itself (system meta-data). Maintaining a repository of wireless meta-data in a way that location databases are maintained is not sufficient to support adaptable application interoperability. The availability/freshness/precision requirements for the system meta-data are different from the requirements for location data and involve more complex monitoring queries. In this paper we examine the problem of implementing and maintaining a wireless interoperable infrastructure with a comprehensive meta-data repository — a pervasive catalog system. Our proposed pervasive catalog system utilizes advanced data management methods in wireless environment and is structured as a dynamic hierarchy of meta-repositories, consolidating meta-data at different levels of details.

1 Introduction

Recent technology advances have resulted in a phenomenal increase of the deployment of wireless access networks and the growth in the number of subscribers. This trend is expected to continue in the observable future [23]. Typical wireless access networks include analog and digital cellular phone networks, wireless local area networks and wide area mobile data services (e.g., circuit switched data service in digital cellular networks, Metricom's Ricochet service, etc.). Ongoing research extends the capabilities of wireless access networks to provide multimedia service at higher data rates, while efficient data management in wireless networks still remains an important research challenge [1, 6, 9, 12, 14, 15, 18].

One characteristic of the next generation mobile environments is a high level of interoperability between mobile applications and services. Interoperability is the ability of distributed systems to interact and to integrate functionalities, and hence to perform collaborative operations. For example, the next generation wireless environments are expected to support a variety of Mobile Web Services built from data services that are potentially located anywhere in the world [22]. The mobile services should be integrated into future mobile portals and made interoperable. Interoperable Mobile Web Services will allow mobile users to combine apparently unrelated applications, possibly supported by other, peering mobile devices, into composite services.

Principles of interoperability allow users to share data and applications across distributed networks, varying processing platforms and vendor brands. A major component of an interoperable environment is a catalog that stores system meta-data (e.g., information about data mapping, name and type conflict resolution rules, protocols and data formats to marshal/unmarshal application parameters, etc.). These principles which have been developed in the context of heterogeneous and distributed systems, including multidatabase systems [8], assume a static environment with respect to data sources, services and processing sites, and a reliable and high speed communication network. In contrast, the mobile and wireless environment is a highly dynamic environment with respect to all of these aspects. These include wide disparity in the availability of remote services, limitations on local resources imposed by weight and size constraints, concern for battery power consumption and lowered trust and robustness resulting from exposure and motion, and unpredictable network quality of wireless networks. Further, these well-known problems for wireless environments are going to remain in the observable future.

Thus, the principles of interoperability need to be extended in the resource-scarce and constantly changing mobile wireless environment in order to support interoperabil-

*This research has been partially supported by the National Science Foundation under grant ANI-0123705

†This research has been partially supported by the National Science Foundation under grant IIS-0135142.

Proceedings of the 13th International Workshop on Database and Expert Systems Applications (DEXA'02)
1529-4188/02 $17.00 © 2002 IEEE

ity under highly variable and bounded resources. In this paper, we look into these principles which lead to what we call *adaptable application interoperability*: interoperability between wireless mobile applications that adjusts in real-time to match varying computational capabilities and resource constraints of wireless mobile environments.

A key issue for adaptable application interoperability is the efficient monitoring of the wireless environment. To provide a proper level of application interoperability under given conditions of the wireless environment should efficiently monitor itself. It should support an infrastructure for executing complex monitoring queries, such as: *What nodes have enough resources (power, available memory, bandwidth, etc.) to implement certain level of interoperability?* Further, besides the initial selection of resources to be involved in a particular interaction, the infrastructure should facilitate the self-monitoring of the involved application enabling them to adapt to the changing conditions in the wireless environment. The information that support above queries (e.g., location, power, available memory, bandwidth, etc.) represents *wireless network meta-data*. Clearly, the requirements of freshness, precision and availability of these system meta-data are especially high.

The contribution of this paper is the design of a comprehensive meta-data repository — a *pervasive catalog system*, whose functionality goes beyond location management of services which have been proposed to support the tracking of mobile object [21]. It will accommodate the need to integrate meta-data located in a highly distributed and constantly changing wireless mobile networks and support efficient execution of monitoring queries so that it meets the high requirements of freshness, precision and availability of the system meta-data. Towards this, our proposed pervasive catalog is structured as a dynamic hierarchy of meta-repositories, consolidating meta-data at different levels of details and utilizes peer-to-peer (P2P) based service discovery of data, services and operations. It is worth pointing that sharing of resources can be structured as services and hence our proposed pervasive catalog can also be used to facilitate the development of peering relationships between multiple mobile devices so that they can collectively poll their resources and complete their pending tasks in a resource-scarce mobile environment [3].

In the next section we consider a motivating example of an adaptable wireless query processing. In section 3 we discuss location data management and its limitations to support interoperability. Sections 4 and 5 detail the concept of pervasive catalog infrastructure.

2 Motivating Example

We assume a general mobile computing environment in which the network consists of stationary and mobile hosts (MHs) [16]. Certain specialized stationary hosts called *Base Station* or *Mobility Support Stations (BS/MSSs)* are equipped with wireless communications capabilities that enable the mobile hosts to connect to them, and through them, to the high-speed fixed network. The logical or physical area served by a single MSS is called a *cell*. Further, we assume that MHs within a geographical region are able to directly communicate with each other without the support of any MSS or residing within the same cell. That is, MHs in adjacent cells can establish a peer-to-peer communication. It should be noted that current peer-to-peer approaches [10, 19] support a single type of service and assume that the peers execute the same client software.

Consider an example of wireless database queries. Assume that the fixed infrastructure includes several servers with electronic catalogs of products and prices maintained by major software vendors. A group of wireless mobile clients wants to keep track of price increases for products. To achieve this, in the absence of any broadcast push service, the clients submit database queries to the servers in the fixed infrastructure. As a consequence, multiples clients can download the same or overlapping data. Consider the query *find database-related products developed by Oracle or IBM with the price range of less then $2000 for university customers*. We assume that the schemas of the corresponding Oracle and IBM data sources are already integrated (e.g., using wrapper/mediator technology [25]). A wireless client C_1 submits the following SQL expression:

(Select product_name, product_price from Oracle_server where product_price < 2000 and license='university')
union
(Select product_name, product_price from IBM_server where product_price < 2000 and license= 'university').

The alternative data sources that can be used in answering this query include the *Oracle_server*, the *IBM_server*, and other peer wireless nodes; let us assume two such mobile nodes S_1 and S_2 that have already downloaded relevant data. Obviously, these data sources have different and varying capabilities. In contrast to the *Oracle_server* and the *IBM_server*, the two mobile data sources are not expected to support full SQL query processing functionality. For instance, let us assume that data source S_1 does not support SQL queries although it provides data caching services and has downloaded all of the requested data items. On the other hand, S_2 does support partial SQL query processing, but it has downloaded few of the requested data items.

A query plan will attempt to maximize the completeness of the result, taking into consideration the processing capabilities of S_1, S_2 and the client C_1. Assuming C_1 has basic SQL processing capabilities, a possible query plan would be to decompose the query into two subqueries to be executed remotely on *Oracle_server* and *IBM_server* and do

Proceedings of the 13th International Workshop on Database and Expert Systems Applications (DEXA'02)
1529-4188/02 $17.00 © 2002 IEEE

the union on the wireless client C_1. However, in the case that one or both of above servers are not available or overloaded, the other plans might be:

- If the bandwidth and latency are acceptable, download data from S_1 and execute entire query locally on C_1.
- If S_2 has enough power, execute the entire query locally on S_2 by requesting missing data items from S_1 and send the result to C_1.
- If the bandwidth and power are not enough to use any one of the sources S_1 and S_2 or the client C_1 to execute the entire query, try to decompose the query so that subqueries can meet the resource requirements and capabilities of C_1 and each of the sources S_1 and S_2. The result will be aggregated from the subquery results. Note, that although C_1, S_1 and S_2 individually might not fully answer the example query, their aggregated service might.

Clearly, the right choice of the query plan depends on the current and accurate knowledge of data distribution, replication and caching, as well as statistics about the wireless network, such as bandwidth and available power in the above example. To answer the above query the system should not only realize what part of data is cached at S_1, but also possibly tolerate a partial answer that does not include all the product names satisfying the selection condition. (In the case that the query involves some form of aggregation, then the system should also anticipate approximate answers.) Query planner and executor should be able to request the relevant statistics and application meta-data both during query planning as well as during monitoring of the execution. These meta-data queries may be quite complex.

3 Location Data Management and Wireless Interoperability

The meta-data management issue in mobile wireless environments that attracted most attention thus far has been the tracking of mobile objects. Identifying the current location of a mobile device is important to both the delivering of data as well as to support location-based services.

In wireless access networks, communicating with a mobile node and running mobile applications include an extra work on searching and updating their location. With increase of population of the mobile users, an overhead from the additional network traffic for the location processing becomes quite significant. An important part of wireless network architecture is an efficient mechanism for storing, querying and updating the location data. A major trade-off in implementation of such mechanism consists in proper balancing the cost of location lookups against the cost of location updates with different levels of data availability, freshness and precision [21]. Existing approaches to the design of location databases in wireless networks implement reasonable options in the above lookup/update trade-off. Among them are hierarchical location [24] and regional matching [2] that generalize the basic two-tier scheme with Home Location Register (HLR) and Visitor Location Register (VLR). Different optimization techniques, such as proper database placement, caching, and replication were suggested to improve performance of queries over location databases [21].

Maintaining a meta-data repository for interoperable wireless environments in a way that location databases are maintained is not an appropriate solution. Issues of adaptable interoperability impose more significant demand on the knowledge maintained by the wireless network about itself (system meta-data) and the algorithms that use that knowledge [9]. First, location databases are typically used to evaluate simple queries, such as "given an object name find the object location." In principle it is possible to use location databases to execute more complex location-aware queries, like finding services and points of interests based on the requesters' location attributes. However, efficiency of such queries in a wireless environment with millions of mobile devices may not be acceptable. Second, visible location updates (e.g., due to crossing the cell boundary) are not very frequent compared to significant fluctuations in the mobile environment that should be reflected in the wireless network meta-data for interoperable applications. In addition to the traffic related to tracking mobile users, an interoperable wireless access network should carry an extra traffic with meta-data that supports implementation of specific interoperability strategies (e.g., availability of services, device capabilities, user preferences, local network conditions, access bandwidth).

The availability/freshness/precision requirements for that meta-data are different from the requirements for location data: the frequency of updates is much higher, the imprecision is less tolerable, and the availability requirements are stricter. As a result, we are not able to re-use directly the approaches for implementation of location databases to provide meta-data support for interoperability in wireless networks. Thus, our goal is to devise an infrastructure that utilizes the flexibility and performance advantages of location databases while providing the components that implement the extra needed functionality for interoperable mobile services.

4 Towards Pervasive Catalog Infrastructure

To meet high requirements to freshness, precision and availability of meta-data in the interoperable wireless environments, as well as support complex monitoring queries

Proceedings of the 13th International Workshop on Database and Expert Systems Applications (DEXA'02)
1529-4188/02 $17.00 © 2002 IEEE

we develop a concept of *pervasive catalog infrastructure*. The pervasive catalog together with advanced query processing component form a core of fully supported adaptable interoperable infrastructure for wireless mobile networks. The concept of pervasive catalog subsumes the concept of location database and provides the additional functionality to meet high meta-data requirements of interoperable wireless applications. The design and implementation of the pervasive catalog utilize efficient methods of distributed data and meta-data management. Given that the concept of pervasive data management subsumes issues of data management in mobile environments, the pervasive catalog infrastructure will also provide a uniform basis to integrate networked data management with service location and interaction protocols supported by the network layer and mobile IP.

In general, the existing approaches to distributed data management can be classified in one of the following two categories [4]. The *warehousing approach* assumes that raw data is transferred from the wireless devices to a central database. The raw data is structured in the appropriate data model and integrated in the centralized database. Queries are executed at the central database. With a *distributed query processing approach* (e.g., device databases [4], dataspaces [14]) information is stored on the network nodes and some queries, or subqueries can be executed on the storing nodes or on other nodes including mobile ones.

Existing ways to implement distributed meta-data management can also be broadly classified into two categories: (1) *Naming services with fixed structure* in which the discovery of data/services/operations is based on the traditional client-server architecture [7]; and (2) *Peer-to-Peer (P2P)* in which the discovery of data/services/operations is dynamically provided by distributed communities of hosts [10, 19]. As opposed to naming services with a central registry, in P2P the meta-data search requests are passed around in a relay, from one host to the next. Thus, meta-data about current host communities is more accurate in dynamic P2P system compared to the naming services with fixed structure. So, the P2P approach is more suitable in frequently changing distributed environments, such as wireless networks. The disadvantage of the P2P approach is that the network may be over-flooded with the peer-to-peer traffic in the case of big host communities and restricted network resources.

We propose a hybrid approach for the pervasive catalog with *adaptive dynamic distribution granularity* (Figure 1). This approach represents a compromise between the fixed structure client/server and P2P approaches, and combines advantages of the warehousing and distributed query processing techniques. Under this approach, every node (device) in the wireless network is associated with its own (i.e., node-specific) most current and accurate meta-data (e.g., CPU and battery utilization in % and other information about capabilities of the device). Some of the nodes are responsible to monitor other nodes and can additionally host more summarized meta-data about a subnet of devices in their local database (meta-data repository). A cost based decision support system (on-line optimizer) is responsible for the choice of a monitored subnet and the content of the meta-data repositories. The local databases are arranged in a monitoring hierarchy. As the wireless environment changes, the on-line optimizer periodically relocates the local databases and re-arranges the monitoring hierarchy.

Thus, the pervasive catalog is implemented as a monitoring hierarchy of meta-repositories dynamically distributed over the wireless network. The decisions about the relocation of local databases and re-arrangement of the monitoring hierarchy are based on the meta-data stored in the pervasive catalog. To break this loop the on-line optimizer uses a set of basic heuristics (bootstrapping base) to build an initial catalog hierarchy. The heuristics in the bootstrapping base should take into account initial approximations of processing and communication capabilities of the wireless devices to minimize meta-data traffic, as well as energy consumption while providing certain level of data accuracy and freshness. The on-line optimizer can be tuned for either maximum lookup or update performance and levels in between. In this way it is able to implement different trade-offs between data availability/freshness/ precision and overall system efficiency.

Our approach allows us to adjust the trade off between the efficiency of data access and data freshness, as well as to introduce a natural meta-data filtering and summarization scheme. The efficient filtering is important because amount of meta-data collected from constantly changing communities of millions mobile devices can grow to unmanageable size quite fast.

5 Querying the Pervasive Catalog

As mentioned above, one of the tasks of the pervasive catalog infrastructure is to provide access cost distributions to data sources that can be used by tools such as WebPT [11] as part of query planning. This also includes providing answers to network monitoring queries such as what nodes have enough resources (power, available memory, bandwidth, etc.) to implement certain level of interoperability. The catalog should also support accurate network condition report generation, such as current and expected link utilization percentage, current and expected throughput of each device (node), etc.

The catalog queries would be submitted to the highly distributed system of meta-repositories that establish the pervasive catalog infrastructure. The meta-repositories repre-

Proceedings of the 13th International Workshop on Database and Expert Systems Applications (DEXA'02)
1529-4188/02 $17.00 © 2002 IEEE

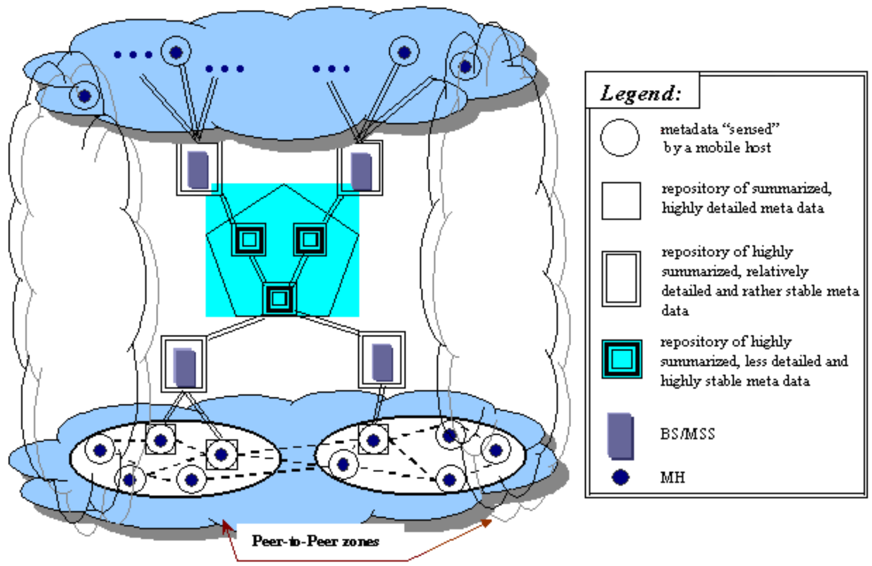

Figure 1. Pervasive catalog infrastructure

sent dynamic data sources with possibly overlapping information and different levels of summarization. We addressed issues of querying highly distributed data sources in the prior research [25]. In particular, we proposed a two-phased approach to query planning and optimization in wide area environments. The first phase utilizes a cost-based pre-optimizer to choose among alternative data sources. In the second phase, we use an optimizer to produce a good query execution plan.

The pre-optimizer constructs a pre-plan, which identifies relevant data sources and their query processing capabilities that can be used to answer a query. The pre-plan is a higher-level abstraction that circumscribes and describes a space of query execution plans. Each query execution plan in the pre-plan space is a physical plan for evaluating the query that has a specific cost. Consider the motivating example of wireless database queries from Section 2. The query pre-optimizer will build a pre-plan including all alternative data sources that can be used to provide information relevant for the query, i.e., *Oracle_server*, *IBM_server*, and two peer wireless nodes S_1 and S_2 that already downloaded relevant data.

To make a right choice the query planner and optimizer should send a *catalog query* for the relevant statistics. Since the pervasive catalog is implemented as highly distributed dynamic system of meta-repositories (i.e., data sources with possibly overlapping information and different levels of summarization), the catalog query itself should be carefully planned and optimized using the same two-phase approach. While planning the catalog query, the optimizer should consider the tradeoff between the accuracy and availability of the wireless meta-data. For example, the first, most general request would be sent to highly available and highly summarized meta-repository on a powerful network server to find out what would be a set of candidate alternative data sources for a given application query. The next, more refined request could be sent to more accurate, current and detailed meta-repositories closely monitoring the relevant data sources, up to sending a request for the most current meta-data to specific sources. The decision on sending more refined catalog requests should be made on the results of previous general catalog requests.

Since planning the catalog queries is based on the meta-data stored in the pervasive catalog itself we have to provide an appropriate bootstrapping base — a set of heuristics, to avoid infinite recursion in catalog queries. As in the case of building the catalog hierarchy, the heuristics in the bootstrapping base should take into account initial approximations of processing and communication capabilities of the wireless devices to minimize data traffic, as well as resource consumption while providing acceptable level of data accuracy and freshness.

We explore the problem of efficient querying the information about current condition of the mobile wireless network environment in terms of its cost and accuracy limits. To provide a highly scalable pervasive catalog infrastructure we explore and utilize advanced query techniques based on methods for querying sensor databases [5, 13], characteristic routing in decentralized network databases [18], network level dataspaces [14] and directed diffusion [15].

Proceedings of the 13th International Workshop on Database and Expert Systems Applications (DEXA'02)
1529-4188/02 $17.00 © 2002 IEEE

6 Conclusion

In this paper, we discussed the need for efficient meta-data management to support adaptable application interoperability enabling the integration of diverse mobile web services and peering of mobile clients. Towards this, we introduced the concept of pervasive catalog infrastructure that monitors the wireless and mobile environment and stores system meta-data in a reliable and distributed fashion.

We described an approach to implement, maintain and query the pervasive catalog to satisfy the high requirements of meta-data availability, freshness and precision. Our approach extends the notion of location databases by utilizing a dynamic hierarchy of summarized meta-repositories consolidating meta-data at different levels of details and adopting P2P-based service discovery of data, services and operations.

Currently, we are working in refining our proposed architecture for the pervasive catalog in conjunction with our two-phase approach to planning and optimization of wireless queries across multiple data sources. Specifically, we are investigating models for describing of services and resources as well as profiles within the pervasive catalog.

Acknowledgements: The authors thank Sujata Banerjee and Mohamed Sharaf for their helpful comments on this work.

References

[1] S. Acharya, R. Alonso, M. Franklin,and S. Zdonik. Broadcast Disks: Data Management for Asymmetric Communications Environments. *ACM SIGMOD Conference*, pp. 199–210, 1995.

[2] B. Awerbuch, and D. Peleg. Online Tracking of Mobile Users. *Journal of the ACM*, 42(5):1021–1058, 1995.

[3] S. Banerjee and P. K. Chrysanthis. Peer Support in a Mobile World, Position paper (invited), *NSF Workshop on Context–Aware Mobile Database Management*, 2002.

[4] P. Bonnet, and P. Seshadri. Device Database Systems. *IEEE Conference on Data Engineering*, pp. 194, 2000.

[5] P. Bonnet, J. Gehrke, and P. Seshadri. Towards Sensor Database Systems. *Mobile Data Management Conference*, pp. 3–14, 2001.

[6] M. Cherniack, M. Franklin, S.Zdonik. Expressing User Profiles for Data Recharging. *IEEE Personal Communications*, pp. 6–13, 2001.

[7] http://www.corba.com.

[8] A.K. Elmagarmid, M. Rusinkiewicz and A. Sheth. *Management of Heterogeneous and Autonomous Database Systems*, Morgan Kaufmann, 1999.

[9] M. Franklin. Challenges in Ubiquitous Data Management. *Informatics*, pp. 24–33, 2001.

[10] http://www.gnutellanews.com/information.

[11] J.R. Gruser, L. Raschid, and V. Zadorozhny, T. Zhan Learning Response Time for WebSources using Query Feedback and Application in Query Optimization. *The VLDB Journal*, 9(1):18–37, 2000.

[12] J.M. Hellerstein, and R. Avnur. Eddies: Continuously Adaptive Query Processing. *ACM SIGMOD Conference*, pp. 261–272, 2000.

[13] S. Iyengar, R. Brooks. *Multi–Sensor Fusion: Fundamentals and Applications with Software*. Prentice Hall, 1997.

[14] T. Imielinski, and S. Goel. DataSpace – Querying and Monitoring Deeply Networked Collections in Physical Space. *ACM MobiDE Workshop*, pp. 44–52, 1999.

[15] C. Intanagonwiwat, R. Govindan, and D. Estrin. Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks. *MOBICOM Conference*, pp. 56–67, 2000.

[16] J. Ioannidis, D. Duchamp, and G.Q. Maguire. IP–Based protocols for mobile internetworking. *ACM SIGCOMM Symposium on Communication, Architectures and Protocols*, pp. 235–245, 1991.

[17] D. Kossmann. The State of the Art in Distributed Query Processing. *ACM Computing Surveys*, 32(4):422–469, 2000.

[18] J. C. Navas and M. Wynblatt. The Network is the Database: Data Management for Highly Distributed Systems. *ACM SIGMOD Conference*, pp. 544–551, 2001.

[19] Software and Information Industry Association. Stretching the Fabric of the Net: Examining the Present and Potential of Peer–to–Peer Technologies. 2001. (http://www.siia.net)

[20] http://www.microsoft.com/data.

[21] E. Pitoura, and G.Samaras. Locating Objects in Mobile Computing. *IEEE TKDE*, 13(4):571–592, 2001.

[22] http://www.sophica.com.

[23] *US News and World Report*, pp. 48–58, 13 Dec. 1999.

[24] J.Z. Wang. A Fully Distributed Location Registration Strategy for Universal Personal Communication Systems. *IEEE Journal on Selected Areas in Communications*, 11(6):850–860, 1993.

[25] V. Zadorozhny, L. Bright, M.E. Vidal, and L. Raschid, T. Urhan. Efficient Evaluation of Queries in a Mediator for WebSources. *ACM SIGMOD Conference*, 2002.

Proceedings of the 13th International Workshop on Database and Expert Systems Applications (DEXA'02)
1529-4188/02 $17.00 © 2002 IEEE