A Framework for Data Roving in Ubiquitous Computing Infrastructure

Richard E. Gunstone and David Newell
Computing and Informatics, School of Design, Engineering and Computing
Bournemouth University, Poole, Dorset, United Kingdom
rgunstone@bournemouth.ac.uk, dnewell@bournemouth.ac.uk

***Abstract*—In future Ubiquitous Computing Infrastructure, a growing data demand from users particularly of mobile devices will create further pressures on infrastructure to deliver information at the right time. We introduce the concept of "data roving", where smart infrastructure predicts the likely information needs and replicates data to parts of infrastructure in the vicinity of the user concerned. We propose this concept to reduce the impact of congestion, reduce pressure on finite infrastructure resources, and improve data access performance for the mobile user. Unknown quantities at this stage include the efficiency of anticipating information needs of users, and whether this offers a performance advantage over more straightforward demand-based caching. The main contribution of this paper is a description of how a network infrastructure could be used in predicting information needs and to support replication of data to enhance performance.**

***Index Terms*—mobility; middleware for mobile environments; ubiquitous computing; mobile content delivery networks**

I. Introduction

The concept of *ubiquitous computing* is acknowledged as beginning with the work of Mark Weiser at Xerox [1]. Weiser can be considered something of a visionary, describing a future paradigm of computing that eschewed then-accepted approaches in favour of a highly distributed, interactive and pervasive world. Similarly, we view ubiquitous computing as a move toward an environment where technology diffuses into the background and where software systems are used that adapt to user needs autonomously. A prevalent example of ubiquitous computing to date has been the emergence of *converged devices* such as smartphones and tablet-class computers, such devices being indicative of a widespread trend toward more ubiquitous architectures.

Contemporary converged devices and mobile devices (MDs) in particular have placed unprecedented demand on communications infrastructure. Current-generation MDs are pushing the data bearing capacities of even recent innovations to 3G cellular networks, necessitating widespread deployment of data service upgrades. Growth areas such as mobile video, and increasing device capabilities and consumer need, are likely to expand these demands further (and are spurring on the development of new technologies such as Cognitive Radio [2]). Increasing demands on finite radio spectrum for wireless networking indicate the introduction of smart techniques into the management of data across networks is of benefit.

We introduce *data roving* (DR) as the 'anticipation of data consumption needs in advance, by either a client or infrastructure, and the reproduction of data locally to the client to satisfy data consumption needs in sufficient time and to required quality'. We consider this to be an evolution of *data staging* (DS), which has its origins in business intelligence, data warehousing, and dynamic infrastructure provisioning. DS in these contexts refers to the provision of a duplicate or replica of an information source suitable for processing, that is then moved back into an storage library after processing is completed. (Similarly, where the replica is not moved back into a storage library, DS is equivalent to *caching*, for instance the service offered by Content Delivery Network (CDN) nodes or HTTP caches.) DR has relevance for mobile users particularly in the development of middleware for mobile environments and mobile content delivery networks.

The concept of DR, much like DS, offers benefits in terms of the reduced latency of access for the user. Rather than the mobile user performing I/O operations across a potentially large network of varying performance levels, I/O can be performed with a nearby replica of the information required. This is achieved by providing information locally to the client, which makes use of a computer system local to the MD with sufficient capacity to vend data to mobile users (in many respects similar to a CDN). Employing DR may also reduce the exposure of the mobile user to congestion effects on busy network links between the user and the source, particularly where the hop count from the client to the DR replica is shorter. Performance may also be higher if network links to the replica have equal or greater reliability than the non-DR equivalent.

The DR concept becomes important when evaluated against a backdrop of increasingly high data consumption demand by consumers, and the exchange of increasingly large file sizes especially for multimedia. The use of MDs with limited local storage also creates further pressure for data to be frequently in transit over a fixed or wireless network link. Therefore, the ability to intelligently move data around is becoming increasingly important. While concepts such as Quality of Service (QoS) provide alternative options, these are typically focused on real-time traffic management according to the type of service required.

In section 2, the paper the DR architecture is described, including a high-level process and a discussion of how information can be selected for DR. This is followed by conclusions in the final section, including consideration of related topics.

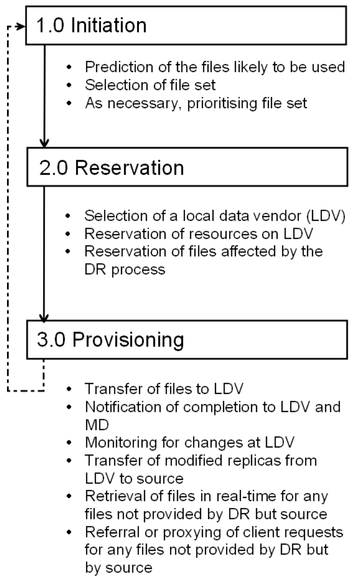

Figure 1. The three steps involved in *Data Roving*: initiating the process, reserving resources within the UCI, and provisioning of the replica(s) to a nearby LDV that also includes committing changes back to the original information store.

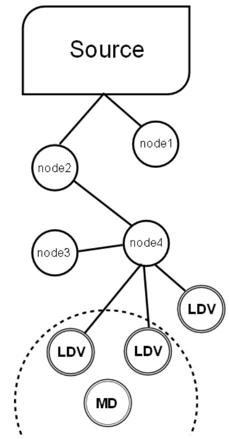

Figure 2. The infrastructural aspects to DR: the most optimal node is designated as the LDV, to provide data to MD without recourse to the source unless the predicted data needed does not match actual needs for MD. More than one LDV may be selected depending on actual and anticipated mobility modelling of MD and other relevant performance factors.

II. Data Roving Architecture

Figure 1 shows the three steps in the DR process, namely *initiation*, *reservation* and *provisioning*, and Figure 2 shows a sketch of an example Ubiquitous Computing Infrastructure (UCI) supporting DR. We define at this point several terms: the Mobile Device (MD) is a converged device that has data provision managed via DR; a Local Data Vendor (LDV) is a node in the UCI that is presently or anticipated to be in proximity (physically, topologically, or both) to the MD, of which there may be more than one at any time; and the "source" is the principal and authoritative manager of the data for the user (and MD).

The initiation phase (1) is concerned with the management of the user data in preparation for distribution to LDVs. The reservation phase (2) is concerned with the selection of LDVs and the reservation of resources both at the source and the LDV. The provisioning phase (3) is concerned with the transfer and synchronisation of files between the source and the LDV. The provisioning phase may also refer or proxy requests from the client for any files that are not available at the LDV but are provided by the source.

In the initiation phase (1), a prediction should made as to what data (files) the mobile user is likely to require. Key elements of this phase include the selection of a set of files that the mobile user is likely to require, based upon a prediction function. As necessary, this stage should include prioritisation of the files into a sorted list based upon a prioritisation rule. Potential algorithms for selecting files for DR include: (1) Selection of all files (in the case of small data stores); (2) The most recently accessed files by the user; (3) The most recently created files by the user; (4) Inspection of recently modified file content and derivation of queries to select further content; (5) A more elaborate file selection based on logical query/policy (e.g., for selecting media by attributes such as genre, artist, year, etc.) developed by the user.

At this point in the DR process, knowledge of a candidate LDV may not be available, and thus a prioritisation of the files after prediction may be suitable for later use. A variety of approaches can be employed for predicting the files a user is likely to need, and in a sense this can be parametrised based on policy settings (including for instance user preferences).

In the reservation phase (2), the process moves onto the selection of the LDV(s). This can identify more than one node in the UCI that can be used to host data nearby to the MD. The selection of candidate LDVs should integrate as many relevant factors as possible; probable factors include the present geographical location of the MD relative to the candidate nodes, the proximity of the MD to the candidate nodes based on number of network hops, quality of link types, applicable policies at the LDV, available LDV resources, etc. This may also involve a historical analysis of the characteristics of interest, of both the LDV and the MD, in an effort to predict a set of LDVs that could be of use to the MD in the future (and not just the present). This could be achieved by integrating the outputs of a mobility modelling process on the UCI and MD. The LDV chosen should be within reach or likely reach of a wireless (or fixed) links from the mobile user, and therefore be in a position to vend data to the mobile user as necessary.

The reservation phase should also reserve resources (storage space, etc.) on the LDV and update the source to maintain awareness of the present state of DR for files identified in phase I. Moreover, appropriate resources should be reserved on the LDV to cater for contention and availability.

After the reservation phase is completed, the process moves into the provisioning phase (3). This part of DR is concerned with the provisioning of data from the source to the MD, via the LDV. Files identified in phase I should in this phase be transferred to candidate LDVs. Taking into consideration the available storage capacities of, and link qualities to, the LDV, a suitable quantity of files are then transferred. Once files are transferred, both the LDV, source and MD should receive a

 ISBN: 978-1-61208-164-9

notification of recent changes.

The provisioning phase is also responsible for monitoring for changes to the files transferred to the LDV, and these changes should be propagated back to the source at the earliest opportunity. Finally, the provisioning phase should provide a function for referral of clients to the source-managed files if these are not available on a local LDV and/or provide a proxy service to retrieve these files on behalf of an MD. The mobile device should be notified by the UCI that a nearby LDV is available for data access needs. The provisioning phase also encompasses the access from the MD to the data provided by the LDV.

Research related to phase 3 includes the work of Flynn et al. [3], which explored the use of data staging (not to be confused with the more generic use of data staging used in the introduction) through the use of the Coda filesystem. They address issues that are beyond the scope of this work at this stage, though the issues are of relevance, and in particular investigate the issue of using untrusted nodes in a network as a vehicle for DS, an aspect that requires further investigation in this framework.

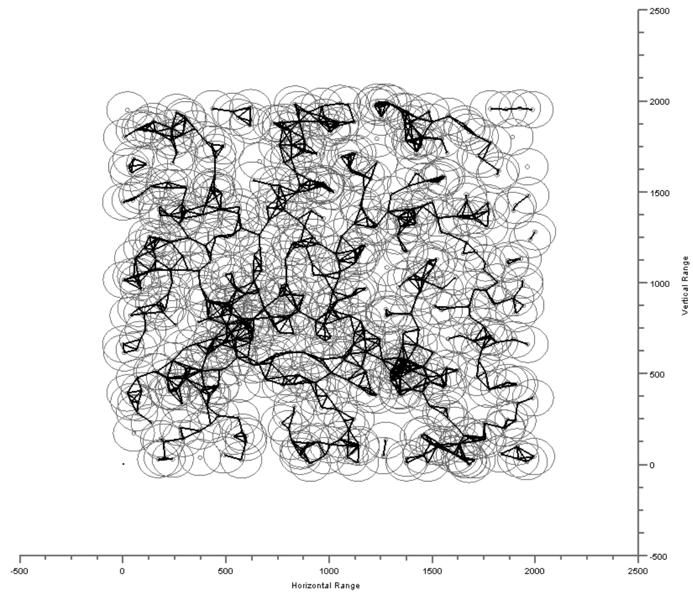

Figure 3. Initial ad-hoc wireless simulation over a 2km/square area.

III. Conclusions

This work-in-progress paper has provided a description of a framework for data roving for ubiquitous computing infrastructure. Three phases in the DR process have been described. Against a backdrop of increasing data demand requirements, relatively low quantities of storage on MDs compared to desktop counterparts, and an increasingly congested communications infrastructure, the concept of DR provides a means through which efficiencies may be feasible.

A number of experimental steps are necessary to validate the DR concept. Firstly, an implementation of DR is necessary to evaluate feasibility characteristics (an early version is shown in Fig. 3). Whereas the work in [3] uses a filesystem, we initially envisage a simpler model to transfer individual files based upon a unique identifier at this stage principally focused on the predictive element of DR. We do not consider heterogeneous traffic at this stage, however this is a future potential area for investigation. Figure 3 shows initial experimental work using a homogenous network type (though we note network selection is a more complex problem more generally, as illustrated in [4]).

A performance evaluation is also required, to assess the performance benefit of particularly the predictive element of DR. That is, the process of anticipating the files needed by a user and transferring this data through the infrastructure to a nearby node. An important parameter in whether the algorithm should consider short- or long-term usage data. Anticipating LDV nodes may lead to several redundant copies of data across the network. While this may be optimised through the use of file synchronisation protocols, there remains a potential inefficiency that can only be justified through the optimal prediction of files the user is likely to require.

Further, and as the introduction noted, employing DR may reduce the exposure of the mobile user to congestion effects on busy network links between the user and the source. In cases where access to the source is lost, DR may offer considerable benefits, and this should be characterised.

An important consideration, briefly mentioned in passing, is a means through which privacy of information transited through the UCI can be ensured. This is not addressed as part of this framework at this stage.

The primary difference between DR and other approaches such as caching, is that the process can be initiated in advance of the anticipated time of need. Particularly in the *reservation* step, an opportunity to make use of low utilisation of UCI, for instance during night time or early morning, when demand network traffic is lower, may create an opportunity to reduce any inefficiencies through the use of DR.

References

[1] M. Weiser, "Some Computer Science Issues in Ubiquitous Computing," *Commun. ACM*, vol. 36, pp. 75–84, July 1993.

[2] S. Hamouda and B. Hamdaoui, "Dynamic spectrum access in heterogeneous networks: HSDPA and WiMAX," in *Proceedings of the 2009 International Conference on Wireless Communications and Mobile Computing: Connecting the World Wirelessly*, IWCMC '09, (New York, NY, USA), pp. 1253–1257, ACM, 2009.

[3] J. Flinn, S. Sinnamohideen, N. Tolia, and M. Satyanaryanan, "Data Staging on Untrusted Surrogates," in *Proceedings of the 2nd USENIX Conference on File and Storage Technologies*, (Berkeley, CA, USA), pp. 15–28, USENIX Association, 2003.

[4] D. E. Charilas and A. D. Panagopoulos, "Network Selection Problem Multiaccess Radio Network Environments," *IEEE Ve*, pp. 40–49, December 2010. Digital Object Identifier 10.1109/MVT.2010.939107.