REQUEST GLOBALIZATION IN AN INFOSTATION NETWORK

Stanimir Stoyanov, Ivan Ganchev*, Ivan Popchev**, Ivan Dimitrov

(Submitted on February 1, 2010)

Abstract

This paper considers the problem of finding an optimal deployment of information resources on an InfoStation network in order to minimize the overhead and reduce the time needed to satisfy user requests for resources. Two formal models are developed for formalizing the problem – a static model and a dynamic model. Ideas of how to use these models for solving the problem of optimal deployment of resources are presented.

Key words: distributed eLearning Centre (DeLC), InfoStation network, request globalization, formal model, optimization task, overhead minimization

1. Introduction. One of the main characteristics of the eLearning systems today is the ‘anytime-anywhere-anyhow’ delivery of electronic content (eContent), which is personalized and customized for each individual user [1, 2]. To achieve this, new types of flexible and intelligent software architectures are required, which are enabled to sense changes in the environment and use this information to adapt their behavior in response to the changing situation. The Distributed eLearning Centre (DeLC) project [3, 4] aims at the development of a context-aware and adaptable software architecture for the delivery of electronic eLearning services in a highly personalized and customized way.

The DeLC communications infrastructure is based on a local area network (LAN) reinforced with information stations (InfoStations) acting as intelligent wireless access and serving points. In the original paradigm [5], the InfoStations

The authors wish to acknowledge the financial support of the Bulgarian Science Fund (Research Project Ref. No. D002-149/2008).
operate as mediators between the user mobile devices and a central server on which a variety of applications are installed. In our architecture we enrich this communications paradigm by extending the role of the InfoStations and deploying services that could be locally activated directly on them. With this distributed deployment of services and other information resources on all available InfoStations, we aim to achieve a good load balancing and better efficiency.

However, the distributed deployment of information resources shows some downsides related to the system overhead associated with the interconnection of information resources during the execution of a service request. This overhead is particularly heavy when the interconnected resources are deployed on different InfoStations. Thus the task of overhead minimization is especially important in the distributed approach.

The amount of overhead depends on the manner of deploying the information resources. A formal model for solving the problem of optimal deployment of resources in an InfoStation network is proposed in this paper. The model is an adaptation of the Overlay-Graph Model case presented in [6]. The Overlay-Graph Model has been used for searching of computer program structures with optimal time behavior.

The rest of the paper is organized as follows. Section 2 presents the InfoStation network and our approach for developing the supporting middleware of DeLC. Section 3 defines (informally) the main notions used in the model. The next two sections present the formal models – firstly the static model and then the dynamic one. The conclusion discusses some initial ideas of using the proposed models for solving the problem of optimal deployment of resources.

2. InfoStation network. The utilized InfoStation-based network architecture provides wireless access to electronic services (eServices) and eContent for users equipped with mobile devices, via a set of InfoStations deployed in key points around a University Campus [7, 8]. The InfoStation paradigm is an extension of the wireless Internet as outlined in [9], where mobile clients interact directly with Web service providers (i.e. InfoStations). The InfoStation-based network architecture consists of the following basic building entities as depicted in Fig. 1: user mobile devices (mobile phones, PDAs, laptops/notebooks), InfoStations, and an InfoStation Center (ISC). The users request services (through their mobile devices) from the nearest InfoStation via available Bluetooth, WiFi/WLAN, or WiMAX/WMAN connections. The InfoStation-based system employs the principles of the distributed control, where the InfoStations act as intelligent wireless access points providing services to users at first instance. Only if an InfoStation cannot fully satisfy the user request, the request is forwarded to the InfoStation Center, which decides on the most appropriate, quickest and cheapest way of delivering the service to the user according to his/her current individual location and mobile device’s capabilities (specified in the user profile). The InfoStation Center maintains an up-to-date repository of all profiles and eContent. The InfoStations
themselves maintain cached copies of all recently accessed (and changed) user profiles and service profiles, and act also as local repositories of cached eContent.

![Fig. 1. The InfoStation-based network architecture](image)

For the development of our system we used a software development approach (presented in [10] in detail) facilitating the use of an InfoStation-based network. The approach uses some fundamental OMG-MDA [11] ideas with additional elements, which take into account the specifics of the InfoStation infrastructure. The core of the system is comprised by a context-aware and adaptive middleware. A flexible development approach is needed in order to create possibilities to examine different development aspects and extend the system architecture step by step. The main idea behind the extended approach is to consider the middleware development as a process of iterations. The term **iteration** – borrowed from the Unified Software Development Process [12] – means a workflow or cooperation between the developers at different levels so as to be able to use and share particular products and artefacts. The iterations in our approach consider different aspects of the development of a context-aware and adaptable middleware. In conformity with the approach, the presented in this paper problem of user request globalization is solving in a particular iteration.

**3. User Request Globalization.** In some cases the user requests for particular services cannot be satisfied fully by the local InfoStation due to resource deficit (e.g. when information needed to satisfy the service request is unavailable in the database of this InfoStation). In these cases the service provision must be globalized in a manner involving other InfoStations (through the InfoStation Center, ISC). The need for globalization depends on the manner in which the resources are deployed on network nodes during the system initialization. Taking into account the fact that each globalization requires an additional overhead, the problem is to find such a deployment, which minimizes the number of global-
izations. The definitions of the main notions related to request globalization are given below.

A request is information sent by a user to the system, which is structured accordingly to the access rules and evokes reaction in the system.

An information resource is a software component, which could be activated and used for processing the requests sent to the system. The resources can be active (e.g. software agents, eServices) or passive (e.g. databases, different data structures). Interconnections between particular information resources may occur during the processing of requests.

A local node for a particular request is the InfoStation, which has received the request and within the service area of which the user is when the request is originally sent.

A global node for a particular request is each InfoStation which is not a local node.

A Request Globalization (RG) is the process of involvement of a global node for the execution of a user request.

The following main principle is used for the development of the middleware: in order to achieve high efficiency in functioning of the system, the execution of requests must be accomplished on the relevant local node as much as possible, i.e. by using only the information resources of the local node, which has received the request originally. This is a kind of an optimization task related to the way in which resources are deployed on the network nodes in order to minimize the number of RGs during requests’ execution.

4. Model for RG-Optimization. Key problems related to the development of the proposed formal RG-Model are treated in this section. This model will be used for the subsequent definition of the task for finding the optimal deployment of information resources in the InfoStation network.

The following sets are defined:

- \( IR = \{ir_1, ir_2, \ldots, ir_n\} \) – a finite set of all information resources deployed in the InfoStation network;
- \( IS = \{is_1, is_2, \ldots, is_k\} \) – a finite non-empty set of all InfoStation nodes.

The way in which the information resources are group and deployed on the InfoStation nodes is called a deployment map, which is formally presented by the mapping \( dm: IR^2 \rightarrow IS \). The set of all possible deployments is denoted as DM where a subset of resources is deployed on a selected node. The subsets are disjoint, so \( dm \) defines a decomposition of IS into subsets.

The existence of large number of possible decompositions is an objective basis for RG-Optimization – a task which could be formulated as follows: from the set DM on IR to find a \( dm \) such as a request to be processed with a minimum number of RGs. The RG-Optimization process can be formal presented as:

- \( RGTime: DM \rightarrow \mathbb{N}^+ \) – a function which gives the number of RGs for each possible deployment of resources.
• select: DM → \{true, false\} – a function which defines the permission conditions. A \( dm \) is called permissible, if select(\( dm \)) = true.

Then the target function of RG-Optimization could be presented as follows: search for \( dm^* \), such as:

• \( dm^* \) is a permissible deployment;
• \( RGTime(dm^*) = \min\{RGTime(dm)\}, \ dm^*, \ dm \in DM. \)

The presented here definition of the possible deployments of resources on individual InfoStations characterizes the importance of the meaningful interpretation of the information resources by taking into account only static factors. Such a static characteristic is sufficient for setting a framework of the optimization problem. However, a practical method for solving the problem cannot be directly derived from it. Some dynamic factors also affect the execution of user requests. Answers to two questions are important for the practical solution of this problem:

• How can we complement the static characterization of deployments?
• As it is impossible to explore the full set of deployments, what criteria can we use to locate a real subset, in which to seek out the optimal deployment?

5. RG-Graph. Graphs are good means for presentation of the structural and functional relationships between the elements of various formulations of tasks, including a large group of optimization tasks. The principle scheme of using graphs for solving the optimization problems includes the following two steps:

• Search for a way for presenting the optimization problem as an appropriate graph;
• Search for optimal elements of this graph (shortest paths, minimum unfolded trees, subgraphs satisfying different conditions, etc.) or generate a set of similar graphs, in which an optimum graph is sought (e.g. different decompositions of one graph).

5.1. Using graphs for local and global optimization. Our primary objective is to search for such a deployment of resources on the individual nodes of the InfoStation network so as a user request to be executed with minimum time consumption. Essential for this in our problem is the number of RGs. In this regard, two types of optimization are possible:

• Local optimization – seeking an optimization of the execution time in regard to a single node. For this, it is necessary to analyze the meaningful (semantic) relationships between the information resources deployed on the same node, which we will call Potential RG (PRG);
• Global optimization – seeking an optimization within the entire InfoStation network, when Request Globalization is required.

Although the global optimization is essential for the solution of our problem, as shown later we (unfortunately) cannot neglect the problems associated with the local optimization (mainly the analysis of the PRGs estimation).
The following base graph is a basic structure we use for (local and global) optimization: BASEGraph = (IR, IRE), where:

- IR – as defined above;
- IRE = IR × IR – a set of graph’s edges; (ir_i, ir_j) ∈ IRE if there exists a meaningful relation between the two information resources (i.e. there exists a PRG at least).

The base graph, however, is not sufficient for solving our optimization problem. Additional information is needed for estimating the nodes and edges of the BASEGraph. An extension of this graph could be defined as follows: EvBASEGraph = (BASEGraph, COST), where:

- BASEGraph – as defined above;
- COST= \{cost_i | i = 1, \ldots, k\}, a set of variety of functions for estimating the nodes and edges of the BASEGraph.

For the purposes of our task we define a cost function prg: IRE → N, which estimates the number of potential RGs for pair of information resources.

EvBASEGraph presents the static deployment of information resources on the InfoStation nodes. In addition, it provides information about (potentially) existed interconnections between information resources. However, during the user requests’ execution, situations may arise which depend on intensity of interconnections and thus may influence the time needed for execution. For accounting these dynamic relationships, we need a graph of a new type – RGGraph-, which is obtained by decomposing the EvBASEGraph, as shown in the next section.

5.2. RGGraph Definition. Now we can define the graph RGGraph = (ISNode, ISE, COST), which we will use for finding an optimal deployment of information resources:

- ISNode = ContIS ∪ NILNodes, such as:
  - if IRGroup_i ∈ ContIS then
    \[\text{IRGroup}_i \subseteq \text{IR} \land \text{IRGroup}_i \neq \emptyset \land \bigcup_{i} \text{IRGroup}_i = \text{IR}, \text{ } i = 1, \ldots, p;\]
  - else if IRGroup_i ∈ NILNodes then IRGroup_i = \emptyset;
- ISE – a set of graph’s edges, such as if (IRGroup_i, IRGroup_j) ∈ ISE, then there exists at least one edge \((ir_i, ir_j) \in \text{IRA}\) and \(ir_i \in \text{IRGroup}_i, ir_j \in \text{IRGroup}_j;\)
- COST – as defined above.

In order to estimate the RGs under the information resources belonging to different InfoStation nodes we define a total for the RGGraph cost function \(rg\_number: ISE \rightarrow N\), such as:

- \(rg\_number((\text{IRGroup}_i, \text{IRGroup}_j)) = \sum \sum prg(ir_i, ir_j), \text{ where } ir_i \in \text{IRGroup}_i, ir_j \in ISE/\text{IRGroup}_i \text{and } \text{IRGroup}_i \neq \text{IRGroup}_j;\)
- \(rg\_number((\text{IRGroup}_i, \text{IRGroup}_j)) = 0, \forall \text{IRGroup}_i = \text{IRGroup}_j.\)

The RGGraph is obtained by decomposing the EvBASEGraph. This graph presents a full characterization of the corresponding deployment map, by
taking into account also the dynamic factors such as the RGs occurring during the execution of a user request.

We would like to demonstrate the two graphs with an example. Let we have a set of information resources $\text{IR} = \{\text{ir}_1, \ldots, \text{ir}_5\}$ with estimated PGRs under them as given in Fig. 2. The corresponding RGGraph is shown in Fig. 3.

**Conclusion.** As could be seen from the above definitions, the set of possible deployments grows very fast with increasing the number of information resources. For instance, if we have only 20 information resources deployed on two InfoStations then the number of possible deployments is more than $10^6$. In general, because of extremely large cardinality of the deployment set, the problem of finding the optimal deployment is not solvable by using acceptable computing resources (CPU cycles, time, and memory). For this reason we decline to search a global optimum (over the whole set). Instead we are going to specify an approach consisting of the next two steps:

- Seeking suitable heuristics for specifying a real subset of possible deployments, where (with high probability) we assume that a satisfying solution exists;
- Seeking an optimum within the new set (local optimum or suboptimum).

Furthermore, we have started the development of an evolutionary strategy for finding a suboptimal deployment map according the proposed two-step approach.

**REFERENCES**


Distributed eLearning Centre
University of Plovdiv
4001 Plovdiv, Bulgaria

*Telecommunication Research Centre
University of Limerick
Limerick 061, Ireland

**Institute of Information Technologies
Bulgarian Academy of Sciences
Acad. G. Bonchev Str., Bl. 2
1113 Sofia, Bulgaria