OPTIMAL RESOURCE DEPLOYMENT IN AN INFOSTATION-BASED NETWORK

Ivan Dimitrov, Stanimir Stoyanov, Ivan Popchev

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. The RG-Optimization problem and an approach for its solving are presented as well.*

Keywords: resource optimization, graph models, Distributed eLearning Centre (DeLC), InfoStation network, request globalization.
2010 Mathematics Subject Classification: 97R40

1. Introduction

The Distributed eLearning Centre (DeLC) project [1,2] 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 [3], the InfoStations 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. 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 [4,5]. The InfoStation paradigm is an extension of the wireless Internet as outlined in [6], 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 Figure 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.

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. In [7] the problem is presented in more detail.

The rest of the paper is organized as follows. Section 2 presents the optimization problem. Section 3 describes briefly our approach for optimization of resource deployment. A formal model for solving the problem of optimal deployment of resources in an InfoStation network is proposed in section 4. The model is an adaptation of the Overlay-Graph Model case presented in [8]. The Overlay-Graph Model has been used for searching of computer program structures with optimal time behavior.

2. 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 globalizations. 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.

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.

3. RG-optimization approach

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 then 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 which: seek suitable heuristics for specifying a real subset of possible deployments, where (with high probability) we assume that a satisfying solution exists and seek an optimum within the new set (local optimum or suboptimum). In this way the approach prescribes the following four steps (Fig.1.):

- RG-problem presentation (RG-model generation) – the problem has been presented as an abstract graph network, given in next section more detail.
- RG-model transformation – the model has to be transform into a more abstract graph (called Abstract Deployment Graph) in order to be prepared for solving by an evolution strategy.
- Optimal Deployment Graph (ODG) generation – by help of an evolution strategy an optimal graph has been generated as solution of the optimization problem.
- ODG interpretation – the problem solution (ODG) is presented as a real deployment map of the information resources.

4. RG-optimization abstract model

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 groped and deployed on the InfoStation nodes is called a *deployment map*, which is formally presented by the mapping *dm:* $\mathbf{IR}^2 \rightarrow \mathbf{IS}$. 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*: $\mathbf{DM} \rightarrow \mathbf{N}^{+}$ - a function which gives the number of RGs for each possible deployment of resources; *select*: $\mathbf{DM} \rightarrow$ {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**.

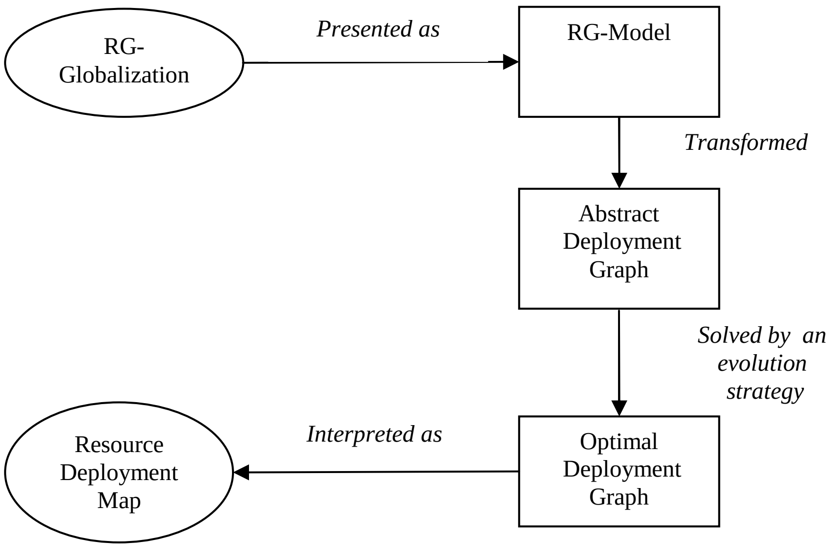

Fig.1. Approach diagram

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? and 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?

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 presentating the optimization problem as an appropriate graph and 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).

Our primary objective is to search for such a deployment of resources on the individual nodes of the InfoStation network so as an 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** x **IR** – a set of graph's edges; $(ir_i, ir_j) \in$ **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**= $\{cost_i \mid i=1,\dots,k\}$, 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** $\to$ **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.

Now we can define the graph **RGGraph** = (**ISNode**, **ISE**,**COST**), which we will use for finding an optimal deployment of information resources: **ISNode** = **ContIS** $\cup$ **NILNodes**, such as: **if** $IRGroup_i \in$ **ContIS** then $IRGroup_i \subseteq$ **IR** $\wedge$ $IRGroup_i \neq \emptyset \wedge \cup IRGroup_i =$ **IR**, $i=1,\dots,p$; **else if** $IRGroup_i \in NILNodes$ then $IRGroup_i = \emptyset$; **ISE** – a set of graph's edges, such as if $(IRGroup_i, IRGroup_j) \in$ **ISE**, then there exists at least one edge $(ir_i, ir_j) \in IRA$ and $ir_i \in IRGroup_i$, $ir_j \in IRGroup_j$; **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** $\to$ **N**, such as: *rg_number*($(IRGroup_i, IRGroup_j)$) = $\sum \sum$ *prg*(ir_i, ir_j), where $ir_i \in IRGroup_i$, $ir_j \in$ ISE/ $IRGroup_i$ and $IRGroup_i \neq IRGroup_j$;

rg_number(($IRGroup_i$, $IRGroup_j$)) = 0, $\forall$ $IRGroup_i = IRGroup_j$. 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 occuring during the execution of a user request.

5. Conclusion

In DeLC an eLearning environment, supporting mobile service provision, is under construction [9,10]. According to the proposed approach an optimizer is being implemented. In this way the eLearning environment will be extend with new services.

6. Acknowledgements

The authors wish to acknowledge the financial support of the Bulgarian Science Fund (Research Project Ref. No. Д002-149/2008).

References

[1] Stoyanov, S., I. Ganchev, I. Popchev, M. O'Droma, From CBT to e-Learning, Journal "Information Technologies and Control", No. 4/2005, Year III, pp. 2-10, ISSN 1312-2622.

[2] Ganchev, I., S. Stojanov, M. O'Droma. "Mobile Distributed e-Learning Center". In Proc. of the 5th IEEE International Conference on Advanced Learning Technologies (IEEE ICALT'05), pp. 593-594, Kaohsiung, Taiwan. DOI 10.1109/ICALT.2005.199. ISBN 0-7695-2338-2. 5-8 July 2005.

[3] Frenkiel R.H. and T. Imielinski, "InfoStations: The joy of 'many-time, many-where' Communications", WINLAB Technical Report TR-119, April 1996.

[4] Adaçal, M. and A. Bener. "Mobile Web Services: A New Agent-Based Framework". *IEEE Internet Computing*, vol. 10, No. 3, pp. 58-65. 2006.

[5] Stoyanov S., I.Ganchev, I.Popchev, M.O'Droma, An Approach for the Development of InfoStation-Based eLearning Architectures, Compt. Rend. Acad. Bulg. Sci., 61, No 9, 2008, Pp.1189-1198.

[6] http://www.omg.org/mda/ (to date).

[7] S.Stoyanov, I.Ganchev, I.Popchev, I.Dimitrov, Request Globalization in an InfoStation Network, Compt. Rend. Acad. Bulg. Sci., 63, No 6, 2010, Pp.901-908.

[8] S. Stojanov, Eine Methode zur Analyse und Optimierung des Zeitverhaltens von grossen Programmen, Dissertation, Humboldt-Universitaet zu Berlin, 1986.

[9] S. Stoyanov, I. Ganchev, M. O'Droma, H. Zedan, D. Meere, V. Valkanova, Semantic Multi-Agent mLearning System, In: "Semantic Agent Systems: Foundations and Applications**,** Springer Verlag, A. Elci, M. T. Kone, M. A. Orgun **(** Eds.), Springer Verlag, 2010.

[10] S. Stoyanov, I. Ganchev, I.Popchev, M.O'Droma, V. Valkanova, Agent-Oriented Middleware for InfoStation-based mLearning Intelligent Systems, 5th IEEE International Conference on Intelligent Systems IS'10, 07.07. – 09.07.2010, London, IEEE Catalog Number: CFP10802-CDR, ISBN:978-1-4244-5164-7, Library of Congress:2009934065, pp. 91-95.

Ivan Dimitrov, Stanimir Stoyanov
Faculty of Mathematics and Informatics
University of Plovdiv
236 Bulgaria Blvd.
4003 Plovdiv, Bulgaria
e-mail: ivandimitrov@uni-plovdiv.bg, stani@uni-plovdiv.bg

Ivan Popchev
Institute of Information and Communication Technologies
Bulgarian Academy of Science
Acad. G.Bonchev Str., Bl.2
1113 Sofia, Bulgaria
e-mail: ipopchev@iit.bas.bg