A Survey of Services Placement Mechanisms for Future Mobile Communication Networks

Shahzad Ali
Institute IMDEA Networks
Leganés, Madrid, Spain
Tel: +34-914816962
shahzad.ali@imdea.org

Andreas Mitschele-Thiel
Ilmenau University of Technology
Gustav-Kirchhoff-Str. 1
Ilmenau, Germany
Tel: +49-3677692819
mitsch@tu-ilmenau.de

Ali Diab
Ilmenau University of Technology
Gustav-Kirchhoff-Str. 1
Ilmenau, Germany
Tel: +49-3677692819
ali.diab@tu-ilmenau.de

Amina Rasheed
COMSATS Institute of Information Technology
Abbottabad, Pakistan
Tel: +92-992383591-5
ammina-rasheed@hotmail.com

ABSTRACT

Mobile communication networks experience a tremendous growth. According to the vision of Wireless World Research Forum (WWRF), there will be 7 trillion wireless devices serving 7 billion people by the year 2017. Ubiquitous access to information anywhere, anytime and anyhow at low cost is one of the essential features of future mobile communication networks, which will interconnect a heterogeneity of various systems and be much more dynamic and flexible in terms of changes in access technology, topology, services, etc. Moreover, there will be a need to match resources supply with application demands as these demands are expected to fluctuate over time. This makes an adequate service placement in such networks of a major importance. In this paper we survey the existing service placement mechanisms and present a qualitative comparison of existing mechanisms. Moreover we also highlight the short comings of existing approaches so that the new approaches can remove the short comings of existing approaches.

Categories and Subject Descriptors

C.2.1 [**Network Architecture and Design**]: Distributed Networks.

C.2.4 [**Network Architecture and Design**]: Distributed Systems.

Keywords

Service placement, ubiquitous computing, and future mobile communication networks.

1. INTRODUCTION

Mobile communication networks have experienced a drastic development evident from a plethora of new applications. The basic goal of Ubiquitous access is to provide information anywhere, anytime and anyhow at low cost and therefore it is one of the core features of future mobile communication networks [1], which will interconnect a heterogeneity of various systems. Such networks need to be much more flexible and dynamic in terms of rapid changes in topology, services, service demands etc. Self-organization would be one of the essential features of such networks [2] to be able to cope with the high dynamics of future communication networks while keeping network performance optimized.

One of the essential issues are the service placement mechanisms the network should employ to enable providing the best service possible through an optimized usage of network resources.

The service placement problem is closely related to problems called *k*-median problem and facility location problem [3]. Given a set of facilities (**F**) and a set of clients (**C**) such that for every pair (**i**) and (**j**) Є (**F** U **C**), there is cost associated $c_{i,j} > 0$. The solution to both problems is to find a subset of open facilities **S** ⊆ **F** capable of serving every client **j** Є **C** by a facility **σ** ***(j)*** Є **S** so that costs are minimized.

Traditional solutions to these problems, however, are not applicable to future mobile communication networks because mostly these algorithms are centralized and therefore global knowledge is required for their operation. Therefore, more feasible solutions are distributed solutions. In this work we present a survey of existing services placement mechanisms by categorizing each services placement mechanism into centralized, distributed or hybrid, then we elaborate strengths and weaknesses of each approach.

The rest of the paper is organized as follows. Section 2 presents related work; section 3 presents discussion on the current approaches. Paper is concluded in section 4.

2. RELATED WORK

Currently, there are three basic strategies available for dealing with service placement problem. These strategies are

- Centralized,
- Distributed and
- Hybrid

Centralized placement techniques require information about the system to be collected in a central entity. This centralized entity is responsible for controlling the placement related tasks. For optimal placement the ideal case is that all information about the system should be collected in a central entity. But the major issue with these approaches is that they require global knowledge of service holders, clients, and costs, therefore, they are less scalable and not applicable to future mobile communication networks. A well-known problem of these approaches is the single point of failure.

As opposite to centralized techniques, distributed service placement approaches assume that there is no central entity responsible for controlling the placement operation and each entity has only partial information about the state of the system. In this way, these approaches lead to better scalability and no single point of failure; therefore, they are more applicable to future mobile networks. However, although distributed approaches result in a reduced cost, increased scalability, and reactiviness, the obtained placement of services cannot be expected to be optimal.

The last strategy is to combine both centralized and distributed techniques in hybrid ones. These approaches have some tasks performed in a centralized manner and some others performed in a distributed manner. Scope of knowledge (local or global) required is task-specific; therefore, there are possibilities of single point of failure. These approaches are less scalable than distributed approaches and better than centralized ones. Moreover, they are less applicable to future mobile communication networks as compared to distributed approaches and better than centralized ones.

2.1 Centralized Mechanisms

In this section, we will discuss some of the centralized service placement mechanisms. These algorithms are based on two common assumptions, firstly, network topology is known and secondly, the cost for each connection is assigned. In [4] three algorithms are summarized, random, greedy, and K-min algorithm.

Random algorithm randomly places the replicas without considering client workload. As it uses random approach so the probability of hosting a service for each candidate is the same. It has been shown in [4] that the ratio of the expected maximum client-replica distance between optimal and random placement increases logarithmically. According to [5], as the number of replicas in the network increased, optimal placement increasingly performs better as compared to random placement in terms of expected client-replica distance. When clients are distributed uniformly, good load balancing can be achieved by optimal placement because clients are directed to the closest mirrors.

The Greedy Algorithm solves the service/replica placement problem by placing replicas on the tree topology. All the replicas are placed in a greedy fashion and no replicas are replaced. The complexity of this algorithm is $O(nk)$, where ´k´ is the number of replicas to be placed, and ´n´ is the maximum number of nodes in the tree [6]. This algorithm assumes that the clients always access the replica that is placed nearest to them resulting in selection of replica with the lowest cost.

In [7], the authors consider the problem of placement of web proxies in the internet. They investigate to find the optimal solution for placement of multiple web proxies *(M)* among some potential sites *(N)* under a given traffic pattern. While most obvious solutions employed for this purpose are to place proxies on routers for LANs, ISP gateways, or some strategic locations, the authors model placement of proxies as a dynamic programming problem. The algorithm works by dividing a large tree T into several small trees ti. Then it places ti replicas in the best way in each small tree Ti where $\Sigma i\ ti = t$. The authors were able to find an optimal solution for a tree topology in $O(N^3M^2)$ time.

Although there are many variants of centralized service placement strategies proposed by different authors, most of them are based on prior network knowledge such as topology and client distribution, which is very difficult information to obtain in the dynamic scenario of ad-hoc networks. Moreover getting complete knowledge about topology and client distribution is also costly in terms of message exchanges, especially in case of highly dynamic topology where information id needed to be updated more frequently. Other limitations of centralized approaches are that they are not scalable to dynamic networks and there is a single point of failure as well.

2.2 Distributed Mechanisms

In this section, we will discuss some of distributed service placement mechanisms. Distributed service placement techniques assume that placement and other related tasks are performed locally, therefore, no global knowledge is required.

In [8] an approach called 'REDMAN' is proposed. 'REDMAN' is basically a replication strategy for read only resources in dense MANETs. The authors have proposed a *Replication Manager;* that acts as a controlling entity whose placement is done by considering local topology. The *replication manager* is hosted by a node only if it detects that it is a part of a dense location in the network. A simple heuristic is used for the placement of the read-only resource by the *replication manager.*

In [9] an approach called SONDe is proposed. The basic mechanism is to set a constant upper bound (*h*) for each node. This *h* determines the number of logical hops the client has traverse in order to access the object holder (provider). This makes the communication latency between a node and any provider tunable and predictable. SONDe has also the capability to dynamically adapt the number of providers by taking into account different load variations experienced in localized portions of the system. On the basis of its h-hops neighborhood each node individually makes a decision that whether it should be a provider of a service or not. It is shown theoretically that SONDe self-stabilizes and provides an independent-dominating set of providers. Prior to SONDe in [10], an approach called Sortie was proposed. Sortie deployed the objects on demand and also provided some level of self-organization; the objects were dynamically placed close to the request locations. One of the major issues with Sortie was that it did not take into account maximal hop distance between providers and other nodes.

In [11] a policy for placing a single service in a MANET called Iterative migration of a centralized service based on service demand is proposed. This approach is very much similar to the hill climbing algorithms; iteratively migrating the service from one host to the other neighboring host until you find a host through which traffic accounting for more than half of the SPC is routed. If such a host is not found then the service remains at its current location. This main advantage of this approach is that it incurs no Control Traffic Overhead and in case of change in service demands or network topology, service placement process restarts. In this work, the authors analytically prove that if there is a tree-like network topology their proposed algorithm eventually reaches the globally optimal position. As their algorithm was basically for MANETs so their algorithm yields best result if the routing protocol organizes nodes into minimal spanning trees.

In [12] an approach called Iterative migration and placement of multiple centralized services is proposed. The MagnetOS project as described by [12] implements a distributed operating system for MANETs that makes the entire network appear as a single virtual machine. Simulations results have shown that the proposed algorithms perform well in terms of system´s lifetime by reducing Service Provisioning Cost.

In [13] an approach called Placement of a distributed service based on node mobility is presented. This approach aims at ensuring service availability in the networks where there are frequent network partitions due to node mobility. The basic idea lies in grouping of the nodes with respect to their velocity vectors. Whenever it is detected that a group is moving out of the radio range of other group then if there is a single node that can provide a service to both mobility groups, then a new service instance is created on that node in the mobility group that would leave both of the groups due to mobility otherwise. It is the responsibility of the node that is currently hosting a service to determine that which node in the departing group will host the new service instance. Information related to the velocity and location vectors of the departing nodes is piggybacked on service requests. Redundant service instances are shutdown if they are found in the same mobility group only if a service instance with a higher unique identification number is found. The authors have shown by doing extensive simulations that proposed approach provides full service coverage for networks having nodes correlated mobility patterns. For random movements of nodes, the coverage still remains high (80% to 90%).

In [14], for solving the Uncapacitated *k*-Median (UKM) and Facility Location (UFL) problem a distributed algorithm is proposed. The basic idea lies in representation of a small network by considering a few hop neighborhoods with known network topology and service demands. The same process is mapped for the outer nodes in order to deal with the nodes outside the hop count threshold. The basic idea is to use a centralized algorithm in a distributed way by limiting the overall network topology to small few-hop networks. There is a possibility of some overlapping of service provision, particularly at the edges of these limited networks which could then be merged and considered collectively.

Authors in [15] also consider the UFL problem and came up with a distributed algorithm. The basic idea lies in constant communication regarding the service assignment among nodes which may host services. Each of these nodes calculates and propagates the Service Provisioning Cost based (SPC) on the local topology. Clients may receive notifications from many nodes but select the service with the least SPC. The neighborhood size can be increased exponentially in order to achieve multi-hop placement.

In [16], authors consider the UFL problem and present a distributed algorithm. They make the assumption that each potential service host can communicate bidirectional with all other clients. They prove that facility location problem can be solved in a definite time if it is considered as a fractional linear program. Distributed randomized rounding is used to map the results of the first iteration to the actual UFL problem.

In [14] an approach based on Ant-Based control algorithm is presented. In such a system, a "demon" M_s for each service *s* is created. This demon is called a *service manage.* An overload condition occurs if the service *s* is not yet placed; M_s creates multiple ants also called as agents and sends them out to all it available interfaces. Each ant contains a *service list* containing *s* and the services cooperating with *s*. For each service, the current resource and communication requirements are known. The ant selects the servers on the basis of probability calculated locally during its path. Whenever the ant encounters a server then it assigns one of the services from the list to that server. As this process is done in a distributed way so the path followed by the ant represents a partial solution to the service placement problem. Once the ant assigns all the services in the list, it returns to the service manager from which it was originated and reports its path and terminates. The manager compares the reported paths using the POF with the POFs of the current placement of those services. Finally, M_s decides that whether there is a need for service rearrangement or not. The most important feature of this approach is that it is distributed and it can be still functional if some servers are out of service. The drawback of this approach is the fact that service manager is the single point of failure.

If we summarize distributed service placement mechanisms then such mechanisms lead to scalability and no single point of failure, therefore they are applicable to the future mobile networks. But while this assumption leads to reduced communication and increased scalability, and reactiviness, but the obtained placement of services to resources cannot be expected to be optimal.

2.3 Hybrid Mechanisms

In hybrid service placement mechanisms some tasks are performed in centralized manner while other tasks are performed in distributed manner. Scope of Knowledge (local or global) required is task-specific.

In [17] authors apply facility location theory in context of selection of cluster heads in hierarchical routing. The authors take the basic motivation from the LEACH-C clustering protocol and they propose that a centralized algorithm for the selection of cluster heads should be run on the base station of the network. The major innovation in their work is that they have formulated the problem as an UFL problem instead of UKM problem that is addressed in LEACH-C. They have shown in their simulation results that the UFL-based approach is able to maintain a working structure of clusters for a longer period of time and it also enhances the overall network lifetime.

In [18], authors have proposed a new algorithm called COCOA. The authors have shown that if dynamic and precise information on the current state of the network is provided, COCOA achieves almost similar performance as current centralized algorithms, while maintaining the scalability of distributed approaches.

In short, though hybrid service placement mechanisms combine the features of both centralized and distributed mechanisms, but these approaches suffer from single point of failure and less scalable than distributed approaches and better than centralized ones, less applicable to future mobile communication networks than distributed approaches and better than centralized ones.

3. DISCUSSION

We discussed three different categories of approaches that can be used for service placement in future networks. Centralized service placement mechanisms offer optimal solutions but they are not scalable to dynamic networks [20, 21] and there is a single point of failure as well. On the contrary, distributed mechanisms lead to scalability and no single point of failure; therefore, they are applicable to future mobile networks. But while this assumption leads to reduced communication cost, increased scalability, and increased reactiviness, the obtained placement of services to resources cannot be expected to be optimal. Hybrid approaches suffer from single point of failure as well. They are less scalable than distributed approaches and better than centralized ones. In addition, they are less applicable to future mobile communication networks than distributed approaches and better than centralized ones. Table 1 summarizes the characteristics of all these approaches. One of the most important things to note is that none of the above mentioned categories have well defined self-organization and self-healing capabilities [19].

Table 1: Qualitative comparison between centralized, distributed, and hybrid approaches

	Centralized	Distributed	Hybrid
Scalability	Limited	High	Intermediate
Single point of failure	Yes	No	Yes
Robustness	Weak	Strong	Intermediate
Network knowledge	Global	Local	Task-specific
Self-Organization	No	Limited	No
Self-Healing	No	Limited	No

Distributed service placement mechanisms can be considered as most suitable to future networks. However, there are few issues prohibiting the usage of these approaches directly. First of all, mostly it is considered that all network nodes are homogeneous having similar capabilities, therefore, every node can host the service. This is, of course, not the case in the real world scenarios. Secondly, mostly distributed service placement mechanisms are highly application-specific. For future networks there will be a lot of heterogeneous services running simultaneously, therefore, we need more flexible solutions that can be applied to a wide range of applications. Last but not least, current distributed service placement mechanisms do not have self-healing and self-configuration capabilities, which will be essential for future networks.

4. CONCLUSION AND FUTURE WORK

Future mobile communication networks will offer ubiquitous access to information and services anytime, anywhere and anyhow. Such networks will be much more dynamic and complex in terms of changing of network access technologies, topologies, services, resources, etc. Due to these reasons, such networks will be hard to manage and control and demand.

Traditional service placement solutions are not easily applicable to future mobile communication networks since they are centralized and require global knowledge about facilities, clients, and costs, which restrict the scalability as well as incur the problem of single point of failure. Distributed approaches overcome the problems faced by centralized approaches. They lead to a better scalability and eliminate the single point of failure problem. A major issue with distributed approach is that the obtained placement of services cannot be expected to be optimal. Moreover, most proposed distributed approaches do not have strong self-organization capabilities. Such capabilities would play, however, sheet anchor and crucial role in future mobile networks. Therefore, there is a need for developing new solutions having self-organization and self-healing capabilities.

5. REFERENCES

[1] Alnasouri, E., Mitschele-Thiel, and A., Diab, A. 2009. A New Distributed Dynamic Algorithm For Mobility Patterns Prediction, In Proceedings of the 6th ACM International Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor, and Ubiquitous Networks (Spain, 2009).

[2] E. Alnasouri, A. Mitschele-Thiel, A. Diab, "DWNG: A New Scheme Enhancing the Automatic Neighbor Relation Function for Self-Organizing Networks", submitted to the 29th Conference on Computer Communications, (INFOCOM'10), San Diego, USA, 2010.

[3] R. Lehmann, "Platzierung von Softwarediensten auf Servern in Kommunikationsnetzwerken", diploma thesis, Ilmenau University of Technology, 2008.

[4] E. Yilmaz, Y. Manzano, "Surveying Formal and Practical Approaches for Optimal Placement of Replicas on the Web", April 2002.

[5] S. Jamin, C. Jin, A. R. Kurc, D. Raz, and Y. Shavitt, "Constrained mirror placement on the internet," Proc. of IEEE INFOCOM, March 2001

[6] P. Krishnan, D. Raz, and Y. Shavitt, "The cache location problem," CM/IEEE Transactions on Networking, vol. 8, no. 5, October 2000.

[7] B. Li, M. J. Golin, G. F. Ialiano, X. Deng, "On the Optimal Placement of Web Proxies in the Internet," in proceedings of IEEE INFOCOM, Mar. 1999.

[8] P. Bellavista, A. Corradi, and E. Magistretti, "REDMAN: An Optimistic Replication Middleware for Read-only Resources in Dense MANETs", Journal on Pervasive and Mobile Computing, 1(3):279–310, Aug. 2005.

[9] V. Gramoli, A.M. Kermarrec, E. Le Merrer, "SONDe, a Self-Organizing Object Deployment Algorithm in Large-Scale Dynamic Systems", Dependable Computing Conference, 2008. EDCC 2008. Seventh European, 2008.

[10] Jamjoom, S. Jamin, and K. Shin. "Self-organizing network services", Technical Report CSE-TR-407-99, University of Michigan, 1999.

[11] K. Oikonomou and I. Stavrakakis, "Scalable Service Migration: The Tree Topology Case", In 5th Annual Mediterranean Ad Hoc Networking Workshop, Lipari, Italy, June 2006.

[12] H. Liu, T. Roeder, K. Walsh, R. Barr, and E. G. Sirer, "Design and Implementation of a Single System Image Operating System for Ad Hoc Networks", 3rd Intl. Conf. on Mobile Systems, Applications, and Services, Seattle, USA, June 2005.

[13] B. Li and K. H. Wang, "NonStop: Continuous Multimedia Streaming in Wireless Ad Hoc Networks with Node Mobility", IEEE Journal on Selected Areas in Communications, 21(10):1627–1641, December 2003.

[14] N. Laoutaris, G. Smaragdakis, K. Oikonomou, I. Stavrakakis, and A. Bestavros, "Distributed Placement of Service Facilities in Large-Scale Networks", In 26th Annual IEEE Conf. on Computer Communications, Anchorage, USA, May 2007.

[15] C. Frank and K. Romer, "Distributed Facility Location Algorithms for Flexible Configuration of Wireless Sensor Networks", In 3rd IEEE Intl. Conf. on Distributed Computing in Sensor Systems, Santa Fe, USA, June 2007.

[16] T. Moscibroda and R. Wattenhofer, "Facility Location: Distributed Approximation", In 24th ACM Symp on the Principles of Distributed Computing, Las Vegas, USA, July 2005.

[17] T. Furuta, M. Sasaki, F. Ishizaki, A. Suzuki, and H. Miyazawa, "A New Clustering Algorithm Using Facility Location Theory for Wireless Sensor Networks", Technical Report NANZAN-TR-2006-04, Nanzan Academic Society, Mar. 2007.

[18] J. Coppens, T. Wauters, F. D. Turck, B. Dhoedt, P. Demeester, "Evaluation of Replica Placement and Retrieval Algorithms in Self-Organizing CDNs", In Proceeding of the IFIP/IEEE International Workshop on Self-Managed Systems & Services (SelfMan'05), May 2005.

[19] Ali, S., Mitschele-Thiel, A., and Diab, A. 2009. Self-organization methodologies for services placement in future mobile communication networks. In Proceedings of the 6th international Conference on Frontiers of information Technology (Abbottabad, Pakistan, December 16 - 18, 2009). FIT '09. ACM, New York, NY, 1-4. DOI= http://doi.acm.org/10.1145/1838002.1838062.

[20] Madani, S, Weber, D., and Mahlknecht, S. "Position-based Routing Protocol for Low Power Wireless Sensor Networks", JUCS, Vol. 16, 2010, pp. 1215-1233.

[21] Ali, S., and Madani, S., "Distributed Grid based Robust Clustering Protocol for Mobile Sensor Networks", The International Arab Journal of Information Technology, Vol. 8, No. 3, July 2011.