An Evaluation of the Ambient Networks Gateway Selection Architecture

Mikko Majanen and Kostas Pentikousis
VTT Technical Research Centre of Finland
Kaitoväylä 1, FI-90571 Oulu, Finland
Email: firstname.lastname@vtt.fi

Abstract—We study the performance of the Gateway Selection Architecture (GSA), developed within the framework of the EU Ambient Networks Integrated Project, which provides support for gateway discovery, management, and selection for mobile nodes within dynamic routing groups. A routing group (RG) is a cluster of nodes in physical proximity, aware of the group membership, especially in the context of the Ambient Networks architecture, with a common goal of optimizing mobility management and routing functionality in the group. We present two examples on how GSA can be introduced using existing protocols, namely MIP and HIP. Our simulation studies show the benefits gained from group formation when compared to same functionalities implemented in every individual node, and compare the GSA optimized signaling strategy with other competitive approaches.

I. INTRODUCTION

Many anticipate a future wireless world filled by a multitude of user devices and wireless technologies. Effective management of this kind of heterogeneous, mobile, and rapidly changing ad-hoc networks will be a challenging task. The Ambient Networks project (www.ambient-networks.org) is addressing this challenge by developing innovative network solutions based on the dynamic composition of networks providing access through the instant establishment of internetwork agreements. The concept [1] includes e.g. Ambient Control Space (ACS) [2], which provides common control functions to a wide range of different applications and access technologies, enabling the integrated, scalable and transparent control of network capabilities.

Mobility management, a key component of Ambient Networks, can be defined as the set of functions that allow a communications system to adapt itself, seamlessly and optimally, to changes in physical and logical topology of the network. A goal of the Ambient Networks mobility solution is to provide a framework within which existing mobility solutions can be deployed and interoperate, whilst ensuring that new mobility solutions can be added as and when they become available. Novel mobility concepts have been developed to better support moving groups of nodes and users, such as personal area networks and networks formed in mass transport vehicles, such as commuter trains.

In previous work [3], the Gateway Selection Architecture (GSA) was introduced to provide support for gateway identification, management, and selection within a routing group. Emphasis was given to node clustering and gateway selection. The first contribution of this paper is to show how GSA can be introduced using two existing mobility management protocols, namely Mobile IP (MIP) and the Host Identity Protocol (HIP). In §III-IV we explain how GSA enhances their functionality contrasting it in the process with previous work.

Moreover, previous work made a distinction between the terms cluster and routing group (RG) [4]. Take a set of mobile nodes, $U$, and a set of base stations, $B$, connected to the wired network. Each $b_j \in B$ can provide wireless connectivity to all nodes $x \in U$ within its coverage area. A cluster, $S \subseteq U$, is defined as the set of nodes from $U$ that can (i) communicate with each other, (ii) are physically close to each other and, (iii) are likely to stay so. Although (i) and (ii) can be determined using information from layers 1–3, (iii) can be determined only by taking into consideration other situational and context information. Identification and formation of such clusters can enable communication and shared use of applications, while several other optimizations, related to routing and mobility management can be pursued.

In each cluster one node is elected to act as the cluster head. Each cluster head is aware of the cluster topology, including the nodes and their roles. Within each cluster, one or more nodes can act as gateways, relaying packets for other nodes and providing connectivity to other networks. A routing group (RG) is defined as the set of nodes $T \subseteq S$, in which the nodes are aware of group membership. This allows even more possibilities for optimizations than a cluster.

Of course, one might expect that by grouping nodes and delegating mobility management to the cluster head and the gateways certain performance optimizations are possible as discussed in [3]. The second contribution, presented in §V-VI is an attempt to quantify using simulation the benefits from employing GSA in scenarios where several nodes move in a mass transit vehicle. We conclude this paper in §VII outlining future work items.

II. GATEWAY SELECTION ARCHITECTURE—AN OVERVIEW

In the context of mobile/wireless networks three approaches have been followed with respect to gateway discovery. The first is a proactive strategy whereby the gateways broadcast advertisements to the whole network. The nodes requiring gateway services choose the most suitable gateway based on
the advertisements they received. In reactive strategies, the initiative lies with the nodes, which broadcast request messages to the network and select the most suitable gateway based on the replies that are unicasted to them. In hybrid strategies, gateway advertisements are usually broadcasted only to the nodes “near” to the gateway. For instance, the advertisements may have a limited time to live (TTL) value, say, three hops. Nodes farther than this amount of hops use request messages to receive gateway services. That is, if a node does not receive a broadcasted advertisement from any gateway, it will broadcast a gateway request message. [5] includes a performance comparison between proactive, reactive and a hybrid GW discovery approaches; in the scenario used, the proactive approach performed best.

GSA takes a hybrid approach for gateway discovery, introducing a special kind of nodes called gateway selectors (GWS). In GSA, service advertisements and requests are unicasted to the gateway selectors, thus simplifying information dissemination and updates regarding gateway (or more enhanced mobile router) nodes and their capabilities. This should not only decrease the amount of signaling overhead, but also allow the majority of the nodes to have only limited computational capabilities and battery power by keeping the intelligence in the gateway selectors. By introducing GWS nodes, GSA borrows a little from the Service Location Protocol (SLP) [6], with GWSs resembling to Directory Agents, gateways to Service Agents, and other RG nodes to User Agents.

As part of the ACS, GSA is supported by many other ACS functional entities such as triggering [10], [11], context information management [12], and multi-radio resource management [13], which are capable of providing a wealth of information related to gateway discovery and selection. GSA is designed to utilize this extra information aiming at making optimal gateway selections. However, these entities and the gateway selection algorithm and policies are out of scope of this paper and are discussed elsewhere (see, for example, [1]– [4]). The following two sections explain how GSA can be applied to existing MIP and HIP implementations for mobility management.

III. GSA WITH MIP

In MIP [7], [8], base stations act as Home and Foreign Agents (HA and FA, respectively) for mobile nodes. In case of GSA, the gateway nodes act as FAs for all nodes in the RG, just like, for instance in [9]. HAs are still located at base stations. The gateway nodes use base stations as their FAs, so they are working in normal MIP way. GSA utilizes the same messages than MIP, but it extends them by using optional TLV-encoded fields for providing extra information (e.g. QoS parameters). Also, the messages are not longer broadcasted, as depicted in the Fig. 1.

The gateway discovery and selection process starts with the election of GWS. Normally this functionality lies in the same node as the cluster head. The cluster head collects and manages information related to the RG management. Gateway issues are part of this management so it is natural to include GWS functionality in the same node. The elected cluster head (and GWS) node informs the whole RG about its role by broadcasting a role claim (RC) message (see top part of Fig. 1). RG nodes save the address of the GWS.

Base stations (BS) broadcast periodically their own advertisements that contain the address of the BS and optionally some other information about the BS (e.g. QoS parameters). We call these extended MIP BS beacons as Base Station Advertisement (BSA) messages. BSA messages are not forwarded, they have TTL set to one. RG nodes that receive a BSA message and are willing to act as a gateway exploit the information contained in the BSA in forming a Gateway Advertisement (GWA) message describing the gateway service it can provide.

GWA messages are unicasted to GWS by gateway nodes in response to the received BSA message. A GWA message is also sent as a response to a received RC(GWS) message, that is when the RG is formed and the GWS is selected. GWA message includes the address of the GW and its parameters, for instance, bandwidth, battery level, supported radio access networks, connection monetary cost (free of charge vs. charge based on traffic volume or connection duration), to name a few. So, GWA is another kind of MIP BS beacon: extended with some extra information just like BSA, but instead of broadcasting it with TTL=1 it is unicasted to the GWS (with TTL>1).

The gateway selector saves the gateway’s parameters to the list from the received GWA message. The RG nodes make a gateway request (GWR) to GWS when they need gateway service. The GWR message contains requirement parameters for the GW service (same as the GWA message). So, it is a sort of extended MIP solicitation message. When GWS receives a GWR it browses through its list of gateways and selects the most suitable one. GWS replies with a response message (GWRESP) that includes the address of the selected GW and its parameters (yet again, a sort of extended MIP BS beacon). In case the node is not happy with the service chosen/available, it may cancel/postpone the connection or make a new request. Otherwise, it updates its routing table so that the traffic destined outside the RG is routed via the selected gateway. It also sends a registration message (GWREG) to its HA (just like in MIP).

IV. GSA WITH HIP

In the Host Identity Protocol (HIP) Architecture [14] hosts are identified with public keys (Host Identities), not with IP addresses as normally. This helps in mobility and multi-homing issues since the nodes can change their IP addresses and still be reachable via the same Host Identity.

HIP base exchange [15] allows any two HIP-supporting hosts to authenticate each other and to create an HIP association between themselves. It consists of four packets: I1 is the trigger packet sent by the Initiator to the Responder. I1 contains only the Host Identity Tag (HIT) of the Initiator and possibly the HIT of the Responder (if known). The second
packet, R1, starts the actual exchange. It contains a puzzle, initial Diffie-Hellman parameters and a signature covering part of the packet. I2 contains the solution to the puzzle, Diffie-Hellman parameter for the Responder. The packet is signed. R2 is a signed message finalizing the base exchange.

HIP Rendezvous Extension [16] introduces a Rendezvous server (RVS). It serves as an additional initial contact point for its client HIP nodes. Without RVS, HIP nodes need to know each others’ IP addresses in order to initiate the HIP base exchange. With RVS, the initial contact can be made by using RVS’s IP address. RVS’s clients become reachable via RVS’s IP address. This is very beneficial in case of mobile nodes that change their network attachment point, and thus also their IP address, frequently.

The base exchange can also include information about available or requested services [17]. A HIP host capable and willing to act as a service provider includes also REG_INFO parameter in its R1 packets, thus announcing its available services. Also UPDATE packet can be used for this purpose if new services become available after HIP association has been formed. To request registration with a service, a requester includes a corresponding REG_REQUEST parameter in an I2 or UPDATE packet.

There are two ways for HIP nodes to initiate service discovery process [18]. In the so called on-path service discovery a HIP node sends a Service Discovery Packet (SDP) towards the peer node in the Internet (for example its own RVS). Each host on the SDP’s path that provides services responds with a Service Available Packet (SAP). SAP may contain information on all the services it provides, or in case the SDP requested only certain service, only those services are included in the SAP. SAP also includes the R1 parameters. Thus, after receiving SAP, the HIP base exchange can be completed with I2 and R2 messages, SDP corresponds to the I1 packet in this case. If the HIP node wants to search services available only on a certain network region, it may use different multicast addresses instead of the address of the peer node in Internet.

In some cases it is not feasible to use on-path service discovery. The HIP hosts can use so called passive discovery method. In that case the HIP service providing nodes sniff passing by HIP packets. If a packet fulfilling certain conditions is found, a SAP can be created and sent to the HIP node that originated the passing by packet.

GSA can also be used with HIP, with the same principles as with MIP: extending the messages with some extra information and changing the destination of the messages. The signalling is depicted in the Fig. 2. In case of HIP, GWS can be seen as a service that provides gateway selection service. Nodes capable and willing to provide gateway or mobile router service register with GWS. Four-way base exchange extended with service discovery and registration information is needed for that at first time. After registration, gateways can use UPDATE packets as their gateway advertisements. HIP nodes needing gateway services can send their requests to the GWS, which replies with the best available gateway for requester’s needs. After that the requester registers with the gateway service and starts using it. UPDATE packets can be used in case of any changes regarding the service.

V. SIMULATION METHODOLOGY

We use the ns-2 network simulator (version 2.28) [19] to evaluate GSA’s MIP-like mobility management in a commuter train scenario, as illustrated in Fig. 3 (the figure is not to scale). We are interested in quantifying (a) the gains of GSA vs. standard MIP and (b) the advantage of GSA vs. general proactive and reactive strategies. The scenario includes the commuter train (total length=70 m, approx. 3 wagons; wagon width=3 m), and n passenger devices, which are randomly distributed inside the 210 m² area of the commuter train. For the purposes of this study, we configure only one mobile device to act as a gateway. During the first 25 s of the simulation, the mobile devices form a single, stable routing group. At t = 25 s the train starts moving at a constant speed...
of 11 m/s along a straight railway track. From \( t = 25 \) till \( t = 180 \) s the train passes by four base stations located along the railway track. The base stations are placed far from each other so that there is no coverage area overlapping. The first one, BS1, was configured to be the HA for all mobile nodes.

The base stations were connected to each other via wired links and a wired node. The wired links have a bandwidth of 100 Mb/s with propagation delay set to 2 ms. The wired node also acted as a corresponding node to a mobile node, by sending constant bit rate UDP traffic to one of the mobile devices on the train. The IEEE 802.11 MAC layer data rate was set to 11 Mb/s, and it was used by all nodes in the train, using the free space signal propagation model and the DSDV routing protocol inside the RG. For RG formation and management we used the stability-based clustering protocol described in [20]. Unless mentioned otherwise, we run the simulation using the default values and settings in ns-2.

We evaluate the performance of GSA centering on the amount of required control signaling and compare it primarily with proactive and reactive algorithms, but also with the case where every mobile node manages its own mobility using MIP; the case for HIP is analogous. As such, in all results reported below, we consider only the signaling required to provide the same functionality that MIP provides, and we exclude, for instance, routing group formation related signaling and DSDV messages. Further studies of MIP covering, for example, the effect of the velocity to the handoff, throughput and packet loss are presented in [21] and the references therein. The following section presents the mean of ten independent replications, for each of the scenario configurations presented above.

VI. Simulation Results

First, we consider the number of sent messages per mobile node in either of the four alternative strategies. Fig. 4 presents the mean number of sent control messages per mobile node. The error bars indicate the min and max values. The standard deviation \( \sigma \) varies between 9 and 24 with MIP and 0.02 and 0.6 with other cases. Overall, as the routing group size increases, on average, nodes send fewer control messages. Clearly, forming a routing group is beneficial as compared to having each and every node use normal MIP to manage its mobility. The gains are typically an order of magnitude and increase as more nodes are added to the routing group. For example, in the case of \( n = 3 \), on average per node, MIP has to send more than four times the number of messages than GSA. At the other end of the range we explored, with \( n = 98 \) the difference is over 78 times. This is because if a routing group is formed, nodes have to update their location to the HA only once (or every time they change their gateway), but in MIP they do so every time they receive an advertisement from a BS.

Comparing GSA with proactive and reactive approaches (Fig. 5), which also take advantage of group formation, we note that GSA underperforms. When employing GSA, on average, nodes have to send more messages than if they had used a proactive and reactive approaches. This is due to its hybrid strategy. When \( n = 3 \), using GSA nodes transmit 1.5% and 18.7% more signaling packets than when using proactive and reactive approaches, respectively. As \( n \) increases, the proportional difference between the number of messages sent by GSA and PRO increases, reaching 29.1% when \( n = 98 \). Similarly, GSA’s hybrid strategy underperforms REA as the number of nodes increases, although the proportional difference becomes smaller: with \( n = 98 \), GSA nodes send, on average 12.4% more messages.

This underperformance is due to the hybrid strategy GSA adopts. Both gateway nodes and routing group members send advertisements and requests, respectively, to the gateway selector. If a proactive strategy is used, then only the former messages are sent, while if reactive is opted for, only the latter messages are needed. Nevertheless, in GSA all messages are unicasted whereas in the proactive and reactive approaches, all messages are broadcasted, which forces mobile nodes to spend resources processing these messages regardless of whether they are useful in their current state. Broadcasting is a considerably “heavier” operation when compared to unicasting. Moreover, when employing the reactive approach, as implemented in the simulation model, it was not possible to re-select the gateway before the connection was lost (break-before-make handover). On the other hand, in both GSA and proactive approaches the status of the gateways is reported in the advertisement messages periodically, enabling seamless, make-before-break handovers and better load sharing, which may assist avoiding congestion incidents.

Fig. 6 presents the amount of processed control messages and tells a similar story with Fig. 4 with respect to the benefits of forming a routing group as opposed to using normal MIP. We define processed control messages as all sent, forwarded, received and dropped packets handled above the MAC layer. As before, RG management and DSDV routing protocol messages were excluded, and so the average number of processed control messages is a good indicator of the total resource costs needed for gateway discovery and selection.
MIP clearly underperforms the other three approaches. The standard deviation $\sigma$ varies between 159 and 395 with MIP and 1.3 and 6.6 with other cases.

The real gains when using GSA instead of proactive or reactive strategies are illustrated in Fig. 7. First, GSA’s hybrid signaling algorithm overperforms a proactive approach in all configurations. In fact the gains increase with $n$: for $n = 3$, on average, GSA nodes process 13.2% less control messages; with $n = 98$, they process 43.4% less messages. Second, GSA underperforms a reactive approach in small routing groups ($n < 20$). For $n = 3$, GSA nodes process, on average, 14.7% more messages (940.7 vs. 820.4). As $n$ increases, GSA’s relative performance against the reactive approach improves. With $n = 98$, GSA nodes need to process 50.0% less control messages than nodes using a reactive approach.

We note no other significant differences besides those mentioned above. Connection lost time between base stations was effectively the same in all scenarios, with small variations due to the locations of the nodes. The delay introduced by gateway discovery is not studied here and is part of our ongoing work; in this study, gateway selection was triggered well before the GW service was actually needed. Nevertheless, preliminary results (not included due to space limitations) show that GSA has a smaller delay than reactive approaches, because the requesting node has to wait for only one response from the gateway selector; in reactive approach the node has to wait for a certain time in order to gather responses from all possible gateways and do the selection among those. Proactive approach does not have any delay—gateway selection occurs whenever needed among the saved advertisements.

**VII. CONCLUSION**

We evaluated the Ambient Networks Gateway Selection Architecture with respect to sent and processed control messages in a moving commuter train scenario using simulation and found that GSA has a considerable advantage over
other alternatives. In particular, although GSA nodes transmit slightly more but unicasted control messages, as opposed to broadcasted control messages used in reactive and proactive strategies, GSA nodes need to process only 75% or less of the control messages processed using the alternative strategies, for medium-size routing groups. As an aside, we also verify the benefits from forming a routing group as opposed to having each node use Mobile IP independently. Our results show that GSA has more lightweight signaling than proactive or reactive approaches and that it scales much better as the routing group size grows.

The other contribution of this paper is the application of GSA in existing mobility management protocols, namely MIP and HIP. Possible migration scenarios need to be considered and are part of our future work. Our future work agenda also includes a more formal and analytical description of GSA, and the study of the gateway selection algorithm in more detail. As noted in §VI, the delays introduced by gateway discovery is also part of our ongoing work; improvements and optimization are to be pursued in this area.

ACKNOWLEDGMENT

This work has been carried out in the framework of the Ambient Networks project (IST 507134), which is partially funded by the Commission of the European Union. The views expressed in this paper are solely those of the authors and do not necessarily represent the views of their employers, the Ambient Networks project, or the Commission of the European Union. The comments and ideas from people involved in the project’s mobility research and from anonymous reviewers are gratefully acknowledged.

REFERENCES