Leader Election in a Personal Distributed Environment

D Pearce, J Dunlop and R C Atkinson

Mobile Communications Group
Department of Electronic & Electrical Engineering
University of Strathclyde
Glasgow, UK, G1 1XW
Email: dpearce@eee.strath.ac.uk

***Abstract*— A Personal Distributed Environment (PDE) is the global set of inter-networked communication enabled devices that a user possesses, and replaces the single terminal of the traditional personal communication model. The PDE requires a device to host the local Device Management Entity, which coordinates the operation of these devices. The selection of this device should occur without any user interaction, and should consider the performance capabilities of the various devices. This paper presents an algorithm that selects a leader on the basis of performance capabilities, and is able to change the order of importance of the capabilities according to current circumstances.**

***Index Terms*— Leader election, Personal Distributed Environment, Device Management Entity.**

I. INTRODUCTION

As communications systems continue to evolve, new service provision architectures and management approaches are required. These management architectures present the opportunity to provide new services, but also pose new problems not previously encountered. The Personal Distributed Environment (PDE) [1] is one such service architecture. The PDE is comprised of all the devices that a user may utilise to consume communications services. These devices are managed by a device management entity (DME) that plays the role of a user personal agent, interacting with other communication entities. Some examples (non-exhaustive) of functions that the DME performs include selecting and contracting bearer services to suit individual sessions; combining devices to form a distributed terminal; interacting with local application service providers and filtering incoming calls based on user preferences.

One component of the DME is located at a service provider, providing a persistently contactable point (via a URL) for remote parties wishing to communicate with the user, as well as being a reliable storage point for certain user information. Incoming sessions may then be forwarded (subject to some processing) to one of the user's local devices. This component of the DME is termed the root DME (RDME).

A second component of the DME is placed on one of the user's local devices, and it acts on information in the local environment such as available devices and bearer services. This component is termed the local DME (LDME). The user may have a number of PDE sub-networks, where a sub-network is defined as a set of devices connected by short range wired/wireless technologies. Sub-networks may exist in the home, the office and the user's personal area (PAN), each of which has a LDME reporting to the RDME. All the user's PDE sub-networks are connected through the Internet, using access technologies such as UMTS, DSL, WLAN, etc. The makeup of these sub networks changes over time due to the purchase of new devices, battery failure, or simply user choice as to which device to carry at different times.

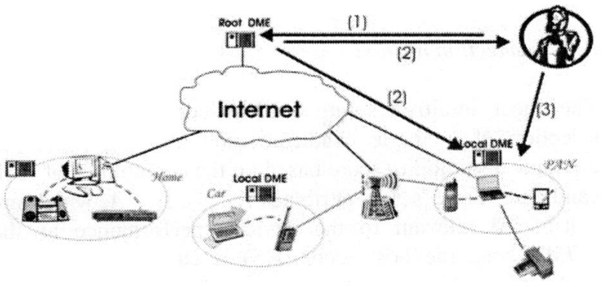

Figure 1 - PDE Concept

Figure 1 PDE Concept

As the LDME performs a number of critical functions, it requires a host device capable of supplying sufficient power, memory, processor and communication facilities. The PDE devices are not homogeneous in their capabilities, ranging for example from smart watches to desktop computers and therefore there exists a need to determine which device is best suited to host the local DME. This is the problem addressed in this paper, specifically

- Selection of a device to host the local DME in a dynamic PDE sub-network.
- Automatic detection of new sub-network formation and merging of sub-networks.

II. Related Work

The local DME coordinates the operation of the user's devices in a sub-network such that they operate as a single distributed terminal. This coordination may include the pooling of communication resources as well as sharing computing resources. The LDME is a software entity that is hosted on one of the devices in the distributed terminal. The LDME will consume resources on the device such as battery power, memory and CPU cycles. As the devices are non-homogenous in their capabilities and, in the case of mobile devices will be resource constrained, it is appropriate to attempt to select the most capable device from amongst those in the distributed terminal. This will prevent undue strain being placed on the resources of the less capable devices. This problem can be thought of as a variant of the leader election problem. The classic leader election problem can be described succinctly [2] as follows; when presented with a number of devices from which to select a leader, (a) eventually there is a leader and (b) there should never be more than one leader .

In this paper a local DME host selection algorithm (LDHSA) is proposed that considers the performance related characteristics of the devices in a PDE. A key strength of the protocol is the ability to consider a variable number of capabilities, based on the currently available devices. In addition, the algorithm is designed to detect the formation of new sub-networks as well as the merging of existing sub-networks.

III. The LDHSA

A. The selection method

The most intuitive solution when confronted with the selection of a single candidate amongst a group is to compute a suitability score based on the capabilities of each candidate. For a set of attributes A_1, A_2, $A_{3.....}A_n$ which are considered relevant to the device's performance as the LDME host, the host score (HS) is calculated as $HS = (w_iA_1 + w_2A_2 + w_3A_3 + ... + w_nA_n)$. The weighting coefficients (w_i) are used to ascribe relative importance to the different device attributes (selection criterion) for device N_i. The problem with this method is that specifying the values of the coefficients is difficult without a target value of *HS*, for which to solve the equation. The algorithm presented in this paper is based on an iterative selection procedure that does not require the determination of weighting factors. A number of desirable parameters that the selected device should possess is first specified. In successive rounds, candidate devices are filtered according to a particular attribute, at the end of which only one device remains and is the local DME host.

Based on the specific application areas, the characteristics deemed important for leader election are different. For the PDE, the criteria proposed are *(1)* The device should contain the necessary (LDME) management software, *(2)* ability to contact the root DME directly or using a proxy, *(3)* should have plentiful energy supply. To have the LDME software installed, a device needs to possess minimum CPU and processor capabilities. This is determined by the specific LDME software utilised. Different PDE service providers may use unique LDME software as a service differentiator. It has been observed that devices with greater CPU capability tend to also have greater memory capacity for executing programs. Newer music players have large hard drive capacities, while not having significant processing power relative to other mobile devices such as a mobile phone or PDA. Based on these observations, the processor capability is used as a criterion, on the assumption that within limits a device with greater processor capability will have greater memory and, if not, the difference will not be significant.

In the event that there is more than one device capable of hosting the LDME from a processor and memory standpoint, it is necessary to consider other characteristics. To maintain the widest range of service delivery options, efforts should be made to maximise the battery life of all devices. PDE devices will be powered in some cases by battery and others by the mains supply. If a mains supplied device that does not have Internet access is selected, it can use a battery-powered device with Internet access as a communication proxy to the root DME. The battery-powered device then only incurs one component of the energy cost associated with hosting the local DME, increasing its expected time to battery failure, thereby maintaining a larger set of devices for a longer period of time. For this reason power source is considered to be more important than Internet connectivity. After power source, the next selection criterion is Internet access. If more than one device has otherwise matching capabilities, the processor speed (MHz) is used as a final differentiator.

The preference order for device attributes when selecting a local DME host is:

- Possess local DME software
- Mains supplied
- Internet access (to the root DME)
- If battery powered, longest battery life

B. Algorithm Execution

Sub-network operation consists of a start-up phase and an ongoing operation phase. The start-up phase is the period immediately after the PDE devices are activated. During this time, devices need to discover each other and select a local DME host. From that point on, devices will be added and removed, new sub-networks may be formed from a subset of current devices and sub-networks may merge. The algorithm execution considers these various possibilities and contains mechanisms to handle these.

As there is no central authority to select the LDME host, the devices must exchange information about their capabilities. At device start-up, a packet is broadcast to all one hop neighbours, indicating the device's capabilities. To reduce message overhead in the network, this announce message is not forwarded beyond the device's one hop neighbours. The receiving device responds with a message indicating its own capabilities. The sending device will wait for a period, equal to the transmitting time for a packet plus a small processing time at the neighbour device. This allows each device to learn all its one-hop neighbour's capabilities. Each node then uses the

978-3-8007-2909-8/05/$20.00 ©2005 IEEE

information about its one-hop neighbours to perform the selection algorithm. If the device finds that it is the most capable local DME host among its one hop neighbours, it will send out a message indicating itself as a LDME host candidate. Unlike the announce message, the LDME candidate message is sent to all devices in the network. Before forwarding an announcement message, a device will record the sender and the message sequence number. If it receives a copy of this message from another device, this will be ignored. The rationale for this approach is as follows:

If a device is not the most capable among its one hop neighbours, there is no reason for the network to incur the overheads involved in sending its information to all nodes in the network. However, though a device can be sure that it is the most capable of the devices in its one-hop neighbourhood, it is not sure if there if a more capable device more than one hop away, and other nodes are not aware of its existence. By propagating only the information from local winners, all devices become aware of the single most capable device in the network, without every device having to send its information through the network.

When a device receives a declaration from a potential leader device, it checks if that device's capabilities are greater than its own or those of a leader announcement it may have previously received. Each device maintains a record of the most capable device that has declared its candidacy thus far. After a suitable delay, the candidate device that has not received a declaration from a more capable device will assume that it has won the election and is the most capable local DME host in the network. At this point it will send out a beacon to all the devices in the network. All other devices should have a record of this device having declared it's willingness to be a local DME host. The beacon will confirm that the election has ended, i.e. all nodes most capable in their one hop neighbourhood have announced themselves, and this device is the best. All devices will then register with the selected device, which returns a confirmation. This ensures that the local DME host is aware of all devices in the network, and the devices are sure that the local DME is aware of its presence.

C. Detecting Lost Devices

The LDME periodically repeats the beacon to indicate to the devices in the sub-network that it is still alive. Each device maintains a timer, which is reset when a beacon is received. If the timer expires, the device tries to contact the LDME and if this fails it initiates the selection procedure by issuing a lost-the-leader message. This sequence of events occurs if the leader device has failed, or has moved away. A second function of the beacon message is to maintain an accurate picture of the devices in the local distributed terminal. The beacon creates a spanning tree rooted at the LDME host. When the leaf devices responds, the message collects information on all devices in the branches of the tree allowing the local DME to determine is any device(s) has failed (or moved away) since it's last beacon. In this way, the movement of any devices away from the sub-network can be detected.

D. Handling Merged Components

The beacon message is also used to detect the joining of two distributed sub-networks. A leaf device in the spanning tree rooted at the local DME will be the first to come into radio range of another sub-network. Upon hearing a beacon from a different LDME, but having the same PDE_ID, it forwards the message by unicast to the LDME host. This occurs when, for example, the user's personal area network (PAN) comes into contact with the home network. The two local DMEs will exchange information and decide amongst themselves, which will control the distributed terminal formed from the merger. The resigning LDME host will indicate to its managed devices that they should register with the winning LDME. This way, a leaf device will not independently change its registration from one LDME to another.

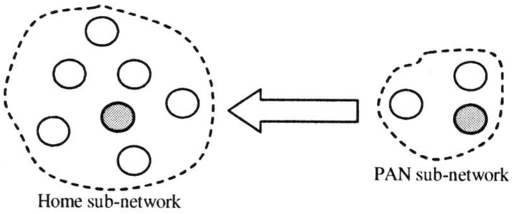

Figure 2 PDE sub-networks coming together

E. Handling New Devices

New nodes send out an announce message, to which the neighbouring nodes that are already registered with the local DME, reply with the id of the device that is currently the local DME host. The new device contacts the local DME host directly. At this stage, the new device and the local DME host compare capabilities. If a handover to the new device is needed is done as is the case with merged components, otherwise the device registers with the current local DME host.

IV. EXPERIMENTS

The primary aim of the experiments was to confirm that by the prescribed message sequence and execution of the selection algorithm, a unique leader is chosen in a PDE sub-network of devices that were previously un-aware of each other. Secondary objectives were to determine if the chosen device is indeed the most capable, according to the specified criteria, determine the time needed to execute sub-network formation and leader selection, and determine the overheads incurred. Also of interest was the ability to handle dynamic PDE behaviour such as the addition/removal of nodes, the formation of new sub-networks and merged networks.

A simulation experiment was conducted using the OMNet++ simulation package, which is a C++ based discrete event simulation package developed at the Technical University of Budapest. All devices used the IEEE 802.11 MAC as it is the most widely deployed wireless MAC in consumer devices, which is the target environment for a PDE.

A. Selection of a Unique Leader

20 devices are deployed in a 30m × 30m area. These dimensions are not intended to define an upper bound on the area covered by a PDE sub-network, but rather are in indication of the typical area a user's local DME sub-network may cover. The most important consideration is that the devices form a connected graph; therefore if for a specific scenario the area under consideration is larger, the transmitter power of the devices can be increased.

At initialisation, each device independently assigns itself a set of capabilities (mains/battery supplied, battery life, internet access, DME software). The characteristics of the devices are then collected by a supervisor module. The supervisor module later uses this information to compare with the decision made by the devices.

The nodes then send their *announce message* and begin execution of the algorithm as previously described. When the selected local DM host sends its first beacon, the supervisor module is prompted to interrogate all devices as to the identity of the node it considers the most capable. In this way the supervisor module can check if a single local DME host was chosen, the identity of this device and whether it is the most capable device. The experiment was repeated 100 times, each time the devices having differing characteristics and different location in the target area. In all runs, a single device was always unanimously selected a single leader. The chosen device was confirmed by the supervisor module to be the most capable leader. The location of the selected device in the network was found to have no impact on the result; neither did closeness in capabilities to other devices.

After the initial selection of the local DME host, 50% of the nodes were instructed to move in unison, away from the initial area, maintaining a connected graph and becoming in effect a new sub-network. The nodes which moved were randomly selected and in some cases included the local DME host. In both the case where the local DME host was stationary and when it was mobile, its absence in the relevant sub-network was detected and a new local DME host selected. When the sub-networks were instructed to move back toward each other, the device which had originally been selected retains leadership of the re-combined group.

B. Timing

The time to complete the algorithm is dependent on the time taken to propagate the information of the leader device, and for all devices to register with the leader device. Using the most robust coding rate of 802.11 (1 Mbps), it takes approximately 0.4ms to transmit a packet (58 bytes). While 802.11 supports higher data rates, the most robust rate was chosen to maximise the likelihood of accurate reception. As each message is contained in a single packet, using this code rate does not exert a significant additional cost in signalling. To account for processing delays, each device was allowed to wait for 1ms after sending out its announcement message to receive a reply. At this point, a device that sends a leader declaration waits a further 10ms to receive declarations from other contenders. After this period, the winning device sends a beacon to which all devices respond with a registration. Consequently the election time is below 1 second (contingent on time settings). In the simulations it was found that this was consistently achievable up to the target of 20 devices.

C. Message Overhead

As the devices utilise scarce wireless channel bandwidth and battery power, it is important to consider the message overhead incurred. If all devices were to distribute their capabilities to all others in the network, the message load per device would increase exponentially. The overhead incurred in performing the selection was measured and it was found that there is a linear relationship between the number of devices in the network and the average number of messages passed by each device (fig 3).

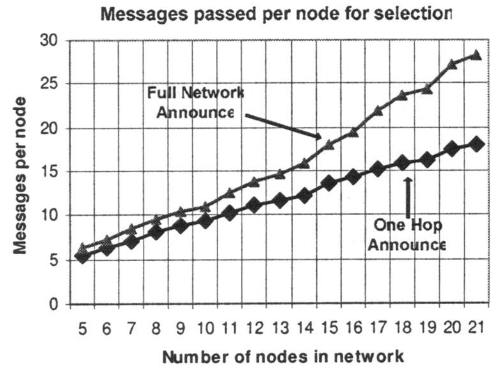

Figure 3 Messages passed per device for selection

D. Startup Messages

While studying the problem of leader election, it was observed that a crucial factor in the successful determination of a leader device is the actual detection of all devices present in the network. It was found that in earlier works based on graph theory, it was assumed that devices in proximity have connection, an assumption well suited to fixed networks. Later works more directly related to wireless broadcast networks, utilise the fact that a message broadcast by a device will be received by all its neighbours within a specific range. In fact [3, 4] make use of the fact that the wireless channel is broadcast to assume that all devices can become aware of their one hop neighbours by each sending a single message. This implies that all devices are awake before any start-up messages are sent. As these are ad-hoc wireless networks, it is unlikely that all devices will know when all other devices are awake. If the start-up message is the sole means of detecting a device's one hop neighbours, the last device to power on will not be aware of any other devices.

This hypothesis was validated via a simulation experiment. A number of devices employing the 802.11 MAC were deployed in a 30m x 30m area. The devices were powered up at one second intervals, at which time they send a start-up message. Each device counts the number of announce messages received to determine its one hop neighbourhood. The ratio of discovered to actually present one-hop neighbours was recorded. For the same area, the number of devices is increased and the experiment is repeated. The results (fig. 4) indicate that it is not possible to determine the complete one hop neighbourhood in the basis of a single announce message. It also shows that increasing device density increases the number of one hop neighbours and therefore the number of neighbours the late starting devices will not discover.

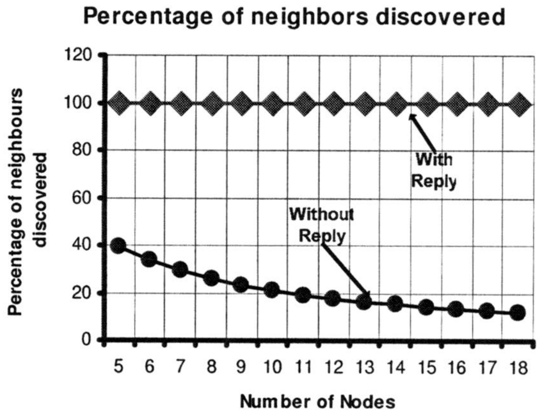

Figure 4 Percentage of devices discovered

To resolve this problem, devices can be required to reply to all received announce message. This allows even late starting devices to detect all one hop neighbours. However the message overhead increases. For *n* neighbours, the first device sends 1+ (*n*-1) messages, the second sends 1+ (*n*-2); eventually the last device sends only one message. The average number of messages per device is therefore $(\sum_{k=1}^{k=n} k)/n$.

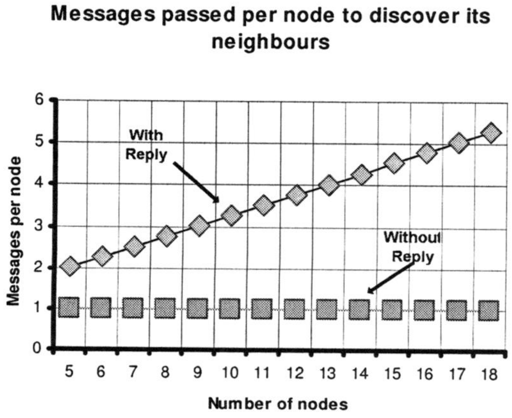

Figure 5 Messages passed per device to discover its neighbours

In the experiment the average number of messages, while linearly related to the number of devices, is lower than the theoretical value as the all devices are not one-hop neighbours of all other devices (fig. 5). These results provide useful insights to developers of protocols which depend on having global knowledge of the membership of an ad-hoc network and are seeking to uses the broadcast nature of the wireless channel to reduce signalling overhead.

V. CONCLUSIONS

This paper presents results in the development of a leader selection algorithm for a dynamic PDE. The leader selection algorithm detects changes in PDE sub-network topology and selects a LDME host, without requiring any user interaction. The experiments also revealed some important effects relating to the minimum signalling required for discovering the total membership of an ad-hoc network.

VI. ACKNOWLEDGEMENTS

The work reported in this paper forms part of the Core 3 Research program of the Virtual Centre of Excellence in Mobile and Personal Communications, Mobile VCE, www.mobilevce.com whose funding support is gratefully acknowledged. More detailed technical reports on this research are available to Industrial Members of Mobile VCE.

REFERENCES

[1] Dunlop, J., Atkinson, R., Irvine, J. and Pearce, D. “A Personal Distributed Environment for Future Mobile Systems”, IST Mobile and Wireless Communications Summit, June 2003.

[2] Attiya, H. and J.L. Welch, “Distributed Computing: Fundamentals, Simulations and Advanced Topics”, pp 31-32, London, UK, McGraw-Hill, 1998.

[3] Deb, B., Bhatnagar, S., and Nath, B. "A Topology Discovery Algorithm for Sensor Networks with Applications to Network Management", DCS Technical Report DCS-TR441, Rutgers University May 2001.

[4] Chandra, R., Fetzer, C. and Hogstedt, K. "Adaptive Topology Discovery in Hybrid Wireless Networks "; Informatics '02.

978-3-8007-2909-8/05/$20.00 ©2005 IEEE