Impact of the Mobility Model on a Cooperative Caching Scheme for Mobile Ad Hoc Networks

F.J. Gonzalez-Cañete and E. Casilari
University of Malaga
Spain

1. Introduction

In the last few years more and more mobile devices have been developed and they are now part of our lives. Cellular phones, PDAs and laptops are daily used in mobile environments such as airports or train stations. For many applications these terminals require access to data networks or the Internet. In order to share information among them the mobile nodes have to cooperate to forward and route the traffic from one node to another forming the so-called Mobile Ad Hoc Networks (MANETs). Due to the mobility of the nodes the mobile nodes can enter or leave the coverage area of other nodes forcing to recalculate the routes in order to make possible the packet forwarding. In addition, wireless networks have a more limited bandwidth and a greater error probability than the wired medium, since the radio medium is shared and prone to interferences and packet collisions. Moreover, the mobile devices have to be portable and hence the processing and battery capabilities also have important restrictions.

Due to the limitations of the MANETs and the mobile devices, some caching mechanisms can be implemented in order to reduce the traffic in the network. Reducing the traffic along the network, and hence the number of forwarded packets, also reduces the battery consumption.

Let us suppose a MANET with connectivity to an external network such as the Internet as depicted in Figure 1. The mobile nodes cooperate to route packets to the Access Routers that provide access to the external networks. As all the traffic in the MANET is routed to the Access Routers these devices can turn into a bottleneck. In addition, the Access Routers can become temporally inaccessible because they can be out of the coverage area of any mobile node in the network because of the nodes’ mobility. This situation causes temporal disconnections to the external networks. The caching mechanisms reduce the impact of the temporal disconnection to the Access Routers as the mobile nodes can cooperate to serve the documents they have previously cached to the rest of the nodes. On the other hand, since the mobile nodes also have the capability of serving information the bottleneck produced in the Access Routers is reduced as the traffic does not reaches them.

As the mobility model influences the behaviour of the nodes in the MANET and the cooperative caching mechanisms depend on the connectivity among the mobile nodes this paper evaluates and compares the performance of a caching scheme under different mobility models. Consequently, in this paper we propose a cooperative caching mechanism and evaluate its performance using two different mobility models.
The rest of this paper is organized as follows. Section 2 comments the related work about the mobility models and the caching architectures. Section 3 describes the caching architecture evaluated in this paper. Section 4 illustrates the simulation model. Section 5 describes the performance evaluation of the proposed caching scheme using two different mobility patterns. Finally, Section 6 outlines the main conclusions of this work and proposes future research directions in this topic.

Fig. 1. MANET with Internet connectivity

2. Related work

In order to reproduce the mobility behaviour of mobile nodes in an ad hoc wireless network some mobility models have been proposed in the last few years. These mobility models can be categorized into three groups:

- Unrestricted random models: the next mobile node destination waypoint is decided randomly according to some heuristics depending on the mobility model. The most used models are: the RWP (Random Way Point) (Broch et al., 1998) mobility model simply selects a random destination in the simulation area; the RD (Random Direction) (Royer et al., 2001) mobility model selects a direction that is followed by the node until the simulation boundary area is reached and then another direction is selected; the Markovian Way Point (MWP) (Hyytia et al., 2006a) and Gauss-Markov (GM) (Liang et al., 1999) mobility models select the next destination using Markovian probabilities among waypoints.

- Geographic-based models: the next mobile node destination is decided according to some geographical constraint. In this category we can mention: the Obstacle Model (OM) (Jardosh et al., 2003) which defines a set of obstacles in the simulation area that must be avoided; the FreeWay and ManhattanGrid (Bai et al., 2003) mobility models limit the nodes’ mobility to predefined ways within the simulation area.

- Group mobility models: the nodes’ mobility tries to imitate typical human collective movements. The RPGM (Reference Point Group Mobility) (Hong et al., 1999) includes the possibility of having dynamic groups of mobile nodes with a leader that decides the next target that the entire group must reach. The DartMouth model (Kim et al., 2006)
chooses the destination of the node movements according to real data sets of human behaviour in a simulation area; the Clustered Mobility Model (CMM) (Lim et al., 2006) divides the simulation area into clusters so that the mobile nodes are assigned to the clusters. The nodes move or not between clusters depending on the number of the mobile nodes in the clusters; ORBIT (Ghosh et al., 2007) randomly defines a set of clusters and the mobile nodes are assigned to some of them moving only between the assigned clusters; SLAW (Self-similar Least Action Walk) (Lee et al., 2009) mobility model represents social contexts present among people sharing common interests using fractal waypoints and heavy-tail flights on top of the waypoints.

As it can be observed, each mobility model tries to reproduce some mobile characteristics although none of them is able to be general enough to be considered the most reliable mobility model under any circumstances. Specifically, mobility models such as Freeway and Manhattan Grid are suitable for vehicular networks since ORBIT or SLAW are adapted for human mobility.

On the other hand, some cooperative caching architectures have been proposed in order to reduce the traffic among the mobile nodes in a wireless network and hence the power consumption. The caching procedures aim to reduce the number of requests sent to the network as some of them can be resolved by the caches implemented in the mobile nodes. Moreover, the cooperation among mobile nodes using caching techniques also reduces the traffic in the data servers or the routers to external networks because the requests are replied on its way to the servers.

The cooperative caching strategies can be divided into four categories:

- Broadcast-based: the mobile nodes broadcast the requests in order to find a mobile node to reply with the requested document. The data server is a static node and hence it can also reply to the request.
- Information-based: the mobile nodes interchange or store information about where the documents are located in the network.
- Role-based: Each mobile node has a function in the network which can be organised in clusters. Depending on the architecture some mobile nodes are selected as information coordinators, clients, etc.
- Directed requests: The requests are directly sent to the server and it is expected to be replied in their way.

MOBEYE (Dodero & Gianuzzi, 2006) is a broadcast based caching scheme that proposes implementing a cache with the LRU (Least Recently Used) replacement policy into each mobile node. When a mobile node needs a document (and it does not have a valid copy in its local cache) it broadcasts a request message. If a mobile node receives the request message and it has a valid copy in its local cache, the mobile node replies using an ack (acknowledge) message to the requester. Finally the document is requested to the first mobile node that acknowledges the request.

SimpleSearch (Lim et al., 2006) is another broadcast based caching scheme very similar to MOBEYE. If a mobile node needs a document that is not stored in its local cache, a broadcast request message is sent a limited number of hops away. When a mobile node with a document copy is found it replies with an ack message that stores the path between the node with the document and the requester. Finally, a confirm message is sent by the requester to the node with the document following the inverse path. Three replacement policies were proposed to use with this scheme:
• **TDS_D** (Time and Distance Sensitive – Distance) – The first criteria to evict documents from the local cache is the distance in hops to the server node. Thus, nearest copies are evicted first.
• **TDS_T** (Time and Distance Sensitive – Time) – The documents with the highest time from the last access are evicted first.
• **TDS_N** – Distance and frequency are pondered in order to choose a document to be removed.

SimpleSearch also defines an admission control that avoids to store in the local cache those documents that are served from less than a certain number of hops away from the requester. In that way very popular documents are avoided to be stored in all the caches.

**DGA** (Distributed Greedy Algorithm) (Tang et al., 2008) is an information based scheme that implements for all the network nodes a table informing about the location of the documents in the network. The nodes store which is the closest and the second closest node where the documents are stored. In addition the mobile nodes send `AddCache` and `DeleteCache` broadcast messages in order to inform the rest of the nodes about the insertion and deletion of documents in the local cache so that they can update their information tables. When a mobile node requests a document it first checks if there is a valid copy in its local cache. If not, it checks if the corresponding table includes the possible document locations. If so, the document is requested to the node stored in the table. If this fails, the document is requested to the data server.

Similarly to DGA the GroupCaching scheme (Ting & Chang, 2007) proposes the mobile nodes to implement a local cache and a group table that stores information about the documents stored in the nodes located only one hop away. Every second the nodes send information to their neighbours informing about its local cache changes in order to maintain the group table updated. On the other hand Hello messages are used to know if a node leaves the group.

**COACS** (Cooperative and Adaptive Caching System) (Artail et al., 2007) is a role-based scheme that obliges the mobile nodes to adopt one of two roles: QD nodes caches the requests and the CN nodes caches the documents. The QD nodes maintain a distributed table about where the documents are located. In that way, if a QD node receives a request and it does not know where to find the document, the request is forwarded to the closest QD. The documents are stored in the CN only if they are served by the data server. In that case, the CN informs the nearest QD about this fact.

Another role based schema was proposed in (Denko, 2007), where the mobile nodes create clusters with a cluster head node (CH) (responsible of the communication among clusters), a data source node (DS) (that stores the data about where the documents are located), caching agents (CA) (that implements a local cache) and mobile hosts (MH). When a node needs a document it is requested to its neighbours, to the CA, the DS and the CH respectively. If the document is not found in any of them it is requested to another cluster using the CH.

The above mentioned cooperative caching schemes have been evaluated to measure their performance using only the Random Waypoint mobility model. As the employed mobility model influences the behaviour of the nodes in the MANET and hence their connectivity we consider necessary to evaluate the caching schemes not only using a unique mobility model, but also using at least one more model in order to compare the obtained performance results.
3. Caching scheme proposed

The caching scheme proposed follows the same request-reply model mentioned in the related work. There is one or more static data server in the MANET that stores the universe of documents and the rest of the devices are mobile nodes that periodically request documents to the data servers. When a node requests a document it waits for the reply a certain amount of time. If the document is not received during this time the node will request it again.

3.1 Local caching

Firstly, all the mobile nodes implement a local cache that stores the received documents. Therefore, if the mobile node have to request a document it first searches in its local cache for a valid copy of the document. If the document is found the request is avoided and hence there is no traffic generated in the network and the server load is also diminished.

The local cache has some parameters that have to be taken into account: the replacement policy, the cache size and the document’s expiration. The replacement policy defines which documents have to be evicted of the local cache in order to make room for a new one. The replacement policy objective is to select for eviction those documents that have the least probability to be requested again in the near future. Unfortunately this is not trivial because it depends on the traffic characteristics. As the traffic in real MANETs is not as well known and studied as the Internet traffic only a few replacement policies have been proposed such as the classic LRU; TDS_D, TDS_T and TDS_N proposed for SimpleSearch; and SXO (Size x Order) proposed in (Yin & Cao, 2006). We adopt the LRU replacement policy because of its simplicity.

The cache size is another important parameter that has to be considered because the bigger the cache is, more documents it will store and the probability to find a previously requested document increases. Due to the fact that the mobile devices may have some restricted storage capabilities the cache sizes of actual equipment will not be too large.

Finally, all the documents in the network have an associated expiration time or TTL (Time To Live) that defines when the information contained in the document is considered obsolete and hence the document has to be requested again if needed. On the other hand, the obsolete documents stored in the local cache can be deleted because they are not valid.

3.2 Interception caching

The functionalities of the mobile nodes can be expanded if they are enabled to perform as a proxy for the other nodes. Since the mobile nodes have to forward the requests to the data servers, they can also check the requested document and search for a valid copy in its local cache. If found, the mobile node replies with the document to the requester node instead of forwarding the message to the server. Using this capability the latency perceived by the user is reduced because the document is served by a closer node in the route to the server. Similarly, the network traffic and server load are also decreased because the request is replied before it reaches the data server.

This operation is illustrated in Figure 1. In the ad hoc network snapshot shown in the figure, DS represents the data server (or a node that accesses to an external network which provides the documents), nodes 1, 2, 3, 4, 5 and 6 are mobile nodes and the lines between them symbolise the existing routes. In this situation if node 2 requests the document A to DS the request will pass through node 1 and will reach DS that will reply to node 2 through
node 1. As the document A is received in node 2 it will be cached. If now node 3 requests the same document A, the request will reach node 2 that will search for the document A in its local cache. As it has a valid copy of A, node 2 will reply to node 3 with document A and will not forward the request to the DS. Using the request interception the number of hops and messages has been reduced from 6 (3-2-1-DS-1-2-3) in the case of no interception to 2 (3-2-3). As the number of hops is reduced the latency perceived by node 3 is also reduced. In addition, the server load is also reduced as the request does not reach the DS. This obviously is achieved at the cost of a higher processing load in the MANET nodes, as they are obliged to analyse all the document requests passing thorough them.

![Fig. 1. Example of ad hoc network](image)

### 3.3 Redirection caching

As the mobile nodes in the MANET have to forward requests and replies from other nodes and the data servers, they can use this information to learn where and how far (in number of hops) the documents are stored in the MANET. Using the information collected from the traffic the forwarding mobile nodes can redirect the requests to other node that is known to have the requested document and that is located closer than the original destination of the request.

Let us suppose that node 3 in Figure 1 requests the document A to DS. The request will pass through nodes 2 and 1 to DS and they will annotate that node 3 has requested document A and it will be available there in the near future at one and two hops away respectively. When DS replies with the document A to node 3 through nodes 1 and 2 they annotate the document’s Time To Live (TTL) in order to know the expiration time. Nodes 1 and 2 do not store information about the reply because the document was served by the DS. If node 4 then requests the document A to DS the request will reach node 2, which will realize that DS is located two hops away and node 3 is one hop away and both have a copy of the document A. As node 3 is nearer than DS, node 2 will redirect the request to node 3 instead of forwarding it to DS. When node 3 receives the request it replies to node 4 with the copy of the document A stored in its local cache. Using the redirection feature the number of hops and messages have been reduced from 6 (4-2-1-DS-1-2-4), in the case of no redirection, to 4 (4-2-3-2-4). As the number of hops is reduced the latency perceived by node 4 is also reduced. The server load is also reduced as the request is served by node 3.

In the previous example the information stored was relative to the requester because the first reply was performed by the data server. Let us suppose that node 6 requests the document B to DS and node 2 has a copy of B in its local cache. The request will pass
through nodes 5, 4 and 2, which will annotate that node 6 will probably store the document B in its local cache in the local cache. As node 2 receives the request it will reply to node 6 intercepting the request. The reply will pass through nodes 4 and 5 to node 6 that will annotate that the document B is stored in node 2 and they will also update the TTL information of the request. Under this situation if node 4 requests the document B it will realize that DS, node 2 and node 6 are located 3, 1 and 2 hops away respectively and hence node 4 will request the document directly to node 2, which is the closer node that is known to have a valid copy of the document B.

We have to remark that if the TTL of the information about the document location is not set, the redirection is not allowed. This constraint prevents from redirecting a request to a node that has not received a certain requested document as the TTL is obtained from the reply and not from the request.

Unfortunately, although the TTL assigned to the redirection information prevents from redirecting requests to nodes that have an obsolete copy of the document, this mechanism does not avoid the requests redirection to a node that has evicted the document because of the replacement policy. To cope with this situation we propose that the node that receives a redirected request and it has not a valid copy of the document in its local cache sends a special error message to the requester in order to send the request again. This message will pass through the redirecting node that will update the information about the incorrect redirection.

Let us suppose that after the situation described previously in the Figure 1, node 6 deletes the document B from its local cache and then node 5 requests the document B. Node 5 has stored that nodes 6 and 2 have the document B and they are located at 2 and 1 hops away respectively. As node 6 is closer the request will be redirected to node 6. When node 6 receives the request it realises that there is not a valid copy of the document B in its local cache and replies with a redirection error message to node 5 that deletes the information about the location of the document B in node 6. Then node 5 will proceed to request the document to the node 2. The redirection errors generate more traffic in the network as well as the latency perceived by the requester node because the number of hops also increases.

Aiming at reducing the number of redirection errors produced by the eviction of documents in the local caches we propose to set as validity time for the redirection information the minimum between the document TTL and the mean time the documents are stored in the local cache. This value is easily calculated by each node considering the amount of time since the document has been stored and the instant in which it is evicted from the local cache.

Figure 2 lists the pseudo-code for the redirection mechanism.

4. Simulation model

We have evaluated by means of simulations the performance of the caching scheme described in the previous section. In order to evaluate the mobility model influences we compare the performance results obtained using the Random Waypoint and the Manhattan Grid mobility models. The simulations are based on the network simulator NS-2.33 which is a popular simulator for the researchers on ad hoc networking (Kurkowski et al., 2005). The BonnMotion (Aschenbruck et al., 2010) and the setdest mobility generators were used to create the mobility scenarios for the Manhattan Grid and Random Waypoint models respectively.
Table 1 summarises the main simulation parameters. We will assume a default scenario with 50 mobile nodes distributed in a square area of 1000x1000 meters. The scenarios with 25, 75 and 100 mobile nodes have also been evaluated in order to study the influence of the density of nodes in the network. There are two fixed servers (DS) located at the coordinates (x,y)=(0,500) and (x,y)=(1000,500) respectively. There are 1000 documents (identified by a number) with a size of 1000 bytes equally distributed between the two servers. Thus, documents with an odd identification number will be stored in one server and the documents with an even identification number will be stored in the other server. All the documents have an associated TTL modeled as an exponential distribution with mean of 2000 seconds. Additionally, we have also tested a mean TTL time of 250, 500, 1000 and infinite (the documents do not expire) in order to study the influence of the document expiration time.

The mobile nodes request documents to the servers following a Zipf-like traffic pattern distribution with a default slope of 0.8 although the 0.4, 0.6 and 1.0 slopes have also been tested aiming at studying the influence of the Zipf slope in the caching scheme proposed. The Zipf-like distribution has been chosen as a traffic pattern because it has been demonstrated to properly characterize the popularity of the Web documents in the Internet (Adamic & Huberman, 2002). The Zipf law asserts that the probability \( P(i) \) for the \( i-th \) most popular document to be requested is inversely proportional to its popularity ranking as shown in the Equation 1.

\[
P(i) = \frac{\beta}{i^\alpha} \quad \text{with } \alpha \text{ close to 1}
\]  

Fig. 2. Pseudo-code for the redirection caching mechanism

```plaintext
for each message (msg) to be sent or forwarded

msg.method - Request (GET) or response (RESP)
msg.id - Document identification
msg.hops - Number of hops from the source node
msg.TTL - Document's TTL

switch (msg.method)
case GET:
    redirectNode = lookRedirectionCache(msg.id)
    if (exists(redirectNode) and distanceHops(redirectNode) < distanceHops(server(msg.id))
        redirectMessageTo (redirectNode)
    else
        forwardMessage
    savePassingByInformation(GET, msg.id, msg.hops)
    break

case RESP:
    updateTTL(msg.id, msg.TTL)
    if (msg does not come from a server)
        savePassingByInformation(RESP, msg.id, msg.hops, msg.TTL)
    break
end
```
## Table 1. Simulation parameters

The parameter $a$ is the slope of the log/log representation of the number of references to the documents as a function of its popularity rank ($i$) while $\beta$ is the displacement of the function. Each time a mobile node requests a document it will wait for a timeout to receive the reply. If the document is not received during this time it will be requested again. Once the requested document has been received the node will wait during a certain amount of time modelled by an exponential distribution with a mean of 25 seconds before proceeding to a new request. Waiting times of 5, 10 and 50 seconds have also been tested. Using this wide range of mean time between requests we can explore the influence request load.

The LRU replacement policy has been chosen for the caches with a default storage space of 35 documents. Cache sizes with a capacity of 5, 10, and 50 documents have also been simulated aiming at testing the influence of the cache size.

The simulation time has been set to 20000 seconds. 20% of this time (4000 seconds) has been used to warm-up the caches and avoid cold start influences. Consequently the statistics collected from the simulations are those corresponding to the time after the warm-up.

The 802.1b MAC protocol with the Two Ray Ground propagation model and a coverage radio of 250 meters were used. The popular AODV (Perkins et al., 2003) (Ad hoc On Demand Vector) protocol was selected as the MANET routing protocol.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Default</th>
<th>Tested values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simulation area (meters)</td>
<td>1000x1000</td>
<td></td>
</tr>
<tr>
<td>Nodes</td>
<td>50</td>
<td>25-50-75-100</td>
</tr>
<tr>
<td>Servers</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Documents</td>
<td>1000</td>
<td></td>
</tr>
<tr>
<td>Document size (bytes)</td>
<td>1000</td>
<td></td>
</tr>
<tr>
<td>Timeout (s)</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>TTL (s)</td>
<td>2000</td>
<td>250-500-1000-2000-∞</td>
</tr>
<tr>
<td>Mean time between requests (s)</td>
<td>25</td>
<td>5-10-25-50</td>
</tr>
<tr>
<td>Traffic pattern (Zipf slope)</td>
<td>0.8</td>
<td>0.4-0.6-0.8-1.0</td>
</tr>
<tr>
<td>Replacement policy</td>
<td>LRU</td>
<td></td>
</tr>
<tr>
<td>Cache size (number of documents)</td>
<td>35</td>
<td>5-10-35-50</td>
</tr>
<tr>
<td>Simulation time (s)</td>
<td>20000</td>
<td></td>
</tr>
<tr>
<td>Warm-up time (s)</td>
<td>4000</td>
<td></td>
</tr>
<tr>
<td>MAC protocol</td>
<td>802.11 b</td>
<td></td>
</tr>
<tr>
<td>Radio propagation model</td>
<td>Two Ray Ground</td>
<td></td>
</tr>
<tr>
<td>Coverage radio (meters)</td>
<td>250</td>
<td></td>
</tr>
<tr>
<td>Ad hoc routing protocol</td>
<td>AODV</td>
<td></td>
</tr>
</tbody>
</table>

The Mobility pattern

<table>
<thead>
<tr>
<th>Mobility pattern</th>
<th>Random WayPoint</th>
<th>ManhattanGrid</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min. and max. speed: 1m/s</td>
<td>Min. and max. speed: 1m/s</td>
</tr>
<tr>
<td></td>
<td>Pause time: 0s</td>
<td>Pause time: 0s</td>
</tr>
<tr>
<td></td>
<td>8 blocks</td>
<td>6, 8 and 10 blocks</td>
</tr>
</tbody>
</table>

The parameter $a$ is the slope of the log/log representation of the number of references to the documents as a function of its popularity rank ($i$) while $\beta$ is the displacement of the function. Each time a mobile node requests a document it will wait for a timeout to receive the reply. If the document is not received during this time it will be requested again. Once the requested document has been received the node will wait during a certain amount of time modelled by an exponential distribution with a mean of 25 seconds before proceeding to a new request. Waiting times of 5, 10 and 50 seconds have also been tested. Using this wide range of mean time between requests we can explore the influence request load.

The LRU replacement policy has been chosen for the caches with a default storage space of 35 documents. Cache sizes with a capacity of 5, 10, and 50 documents have also been simulated aiming at testing the influence of the cache size.

The simulation time has been set to 20000 seconds. 20% of this time (4000 seconds) has been used to warm-up the caches and avoid cold start influences. Consequently the statistics collected from the simulations are those corresponding to the time after the warm-up.

The 802.1b MAC protocol with the Two Ray Ground propagation model and a coverage radio of 250 meters were used. The popular AODV (Perkins et al., 2003) (Ad hoc On Demand Vector) protocol was selected as the MANET routing protocol.
The default speed of the nodes is 1 m/s. No pause time is considered between consecutive movements. Speeds of 2 and 5 m/s have also been tested in order to study the speed influence in the caching mechanism.

For the Manhattan Grid mobility model 8x8 blocks have been chosen as a default scenario. In addition, scenarios with 4x4, 6x6 and 10x10 blocks have been also simulated since these scenarios will allow us to evaluate the influence of the connectivity. Figure 3 illustrates the scenario with the Manhattan Grid mobility model with 8x8 blocks. The mobile nodes (represented by small circles) move along the grid using the lanes defined by the blocks. The two servers A and B (represented as big circles in the figure) are located in the middle of the left and right sides of the scenario.

![Fig. 3. Example scenario using the Manhattan Grid with 8x8 blocks](image)

5. Performance evaluation

The goal is to evaluate the performance of a MANET with the proposed caching scheme taking into consideration the speed and density of nodes, the traffic load (mean time between requests), the mean document expiration time (TTL), the traffic pattern (Zipf slope) and the cache size. For all these analyses, the network performance is studied using both the Random Way Point and the Manhattan Grid mobility models.

For the study of the influence of the density and speed of the nodes every simulation scenario has been executed five times using the same TTL for each document, mean time between requests and request distribution but using a different starting point within the simulation area and a different mobility pattern for each mobile node. The simulation of the rest of scenarios have been executed five times using the same TTL for each document, time between requests and mobility pattern for each node but using a different request distribution. The performance evaluation presented is the mean of the results obtained for the five simulations. Again, the presented results are the mean of the measurements obtained for the five simulations.

As performance metrics we use the following measurements:

- Traffic – The amount of traffic that each mobile node in the network has to process because the node generates the packets or because the packets have to be forwarded. This measurement includes not only the traffic corresponding to document requests and replies but also the overhead introduced by the routing protocol.
• Hops - Defined as the number of nodes that a document has to traverse to be served. It includes the request from the requester to the node that serves the document and back again to the requester node.
• Delay - Defined as the time elapsed between a document request and the reception of the corresponding reply.
• Percentage of timeouts - Defined as the proportion of requests that must be retransmitted again because the reply does not reach the destination before the timeout is reached.
• Local hit ratio – It is the ratio between the number of documents served by the local cache and the total number of documents requested by each node. The higher the local cache hit ratio, the lower the traffic injected in the network is.
• Remote hit ratio – It is the ratio between the number of documents served by a node that is not a server (because of an interception or a redirection) and the total number of documents requested by each node. As the remote hit ratio increases, the server load decreases because more requests are served by the mobile nodes instead of the servers.

5.1 Effect of the network load
Figure 4 represents the mean traffic processed by the nodes (a), the mean delay (b), the mean number of hops (c), the percentage of timeouts (d) and the cache hits (e) as a function of the mean time between requests.

Figure 4.a shows that the traffic generated in the scenario using RWP is greater than that using MG. This is caused by the AODV broadcast messages employed to create the routes between the mobile nodes (Saad & Zukarnain, 2009). As the RWP mobility model tends to concentrate the mobile nodes in the centre of the simulation area (Hyytia et al. 2006b), more nodes receive the broadcasted RREQ (Route Request) messages.

In Figure 4.b we can observe that as the periodicity of document requests increases, the delay is also augmented. As the time between requests increases, the number of documents expired in the nodes’ local caches is also increased and the documents in the local caches are less updated. This can be observed in Figure 4.e where the cache hits decreases as the network load decreases. Therefore, the reduction of the cache hits increases the delay as less requests are served by the local or remote caches. On the other hand, the delay perceived by the RWP (Random Way Point) mobility model is slightly smaller than the Manhattan Grid using 6x6 (MG6) and 8x8 (MG8) blocks but greater than the 10x10 (MG10) blocks. This behaviour is due to the fact that the connectivity is improved as the number of blocks increases because the nodes can communicate with more nodes located at adjacent lanes as long as the distance between lanes is shorter.

In addition, the route TTL configured in AODV is ten seconds and hence the network with a mean time between requests less or equal to this time will take advantage of the already created routes while greater time between requests will have to create the routes again. However, Figure 4.c shows that under RWP nodes need less hops to obtain the documents than under MG although the difference declines as the number of blocks increases. This can be explained as before, the probability to find a shorter route with RWP is higher because the nodes move freely along the simulation area so that they are not restricted to move along the lanes defined by the blocks. Finally, Figure 4.d shows that the number of timeouts is diminished as the network traffic decreases (the mean time between requests increases) until 25 seconds between requests but for 50 seconds between requests the number of
Fig. 4. Mean traffic (a), mean delay (b), mean hops (c), percentage of timeouts (d) and cache hits (e) as a function of the mean time between requests
timeouts is increased. This can be explained similarly as in the case of the delay. Obviously, when the document TTL expires, the effectiveness of the local and remote caching mechanisms decreases and hence the probability to have to request the documents to the data servers increases. As the data servers could remain unavailable due to the nodes’ mobility the probability of timeouts is also increased.

5.2 Effects of the TTL
Figure 5 shows the mean traffic (a), the mean delay (b), the mean number of hops (c), the percentage of timeouts (d) and the percentage of cache hit (e) as a function of the mean documents’ TTL.

The TTL defines the time that the documents are stored in the local caches. We have tested the situations from a low mean TTL (the documents expire after a short interval and they are deleted from the caches very soon) to an infinite TTL (the documents never expire). As the TTL increases the percentage of cache hits is also increased from about 10% to 35% as shown in Figure 5.e and then more requests are served by the local caches. This fact causes the progressive reduction of the traffic generated in the network (Figure 5.a), the delay perceived by the nodes (Figure 5.b), the mean number of hops (Figure 5.c) and the percentage of timeouts (Figure 5.d).

Figure 5.a shows that the traffic generated under RWP mobility model is also greater than with MG as in the studies presented in section 5.1. Finally the figures show a similar behaviour as the presented in section 5.1, the mean delay and the mean number of timeouts is higher using MG6 and MG8 than RWP while MG10 obtains the lowest delay values. However, the RWP obtains a better performance in terms of the number of hops as it is able to find shorter routes.

5.3 Effects of the traffic pattern
Figure 6 shows the mean traffic (a), the mean delay (b), the mean number of hops (c), the percentage of timeouts (d) and the percentage of cache hit (e) as a function of the Zipf parameter $\alpha$.

As the Zipf parameter is closer to one the probability to request again a popular document is higher. This fact drastically enhances the number of local hits as shown in Figure 6.e where the local hit ratio evolves from about 3% to 30% for $\alpha$ equal to 0.4 and 1.0 respectively.

The remote hit ratio is also slightly increased as the parameter $\alpha$ is closer to 1.0. The higher cache hits obtained as $\alpha$ is increased causes the reduction of the generated traffic (Figure 6.a), the delay perceived by the nodes (Figure 6.b), the number of hops needed to obtain the documents (Figure 6.c) and the number of timeouts (Figure 6.d).

The mobility models follow the same behaviour as the previous studies. Under RWP, the network performance obtains intermediate results between MG6, MG8 and the best results obtained by MG10 for the mean delay and mean percentage of timeouts. On the other hand RWP mobility generates more traffic than MG although it requires a lower number of hops to obtain the documents.

5.4 Effects of the cache size
Figure 7 depicts the mean traffic (a), the mean delay (b), the mean number of hops (c), the percentage of timeouts (d) and the percentage of cache hit (e) as a function of cache size.
Fig. 5. Mean traffic (a), mean delay (b), mean hops (c), percentage of timeouts (d) and cache hits (e) as a function of the mean document’s TTL
Fig. 6. Mean traffic (a), delay (b) and hops (c), percentage of timeouts (d) and cache hits (e) as a function of the Zipf slope $\alpha$. 

www.intechopen.com
The cache size determines the number of documents that fit in the local cache. As more documents are stored in the nodes’ local cache the probability of a local or remote cache is increased as shown in Figure 7.e. In this figure we can observe that the cache hit ratio increases from about 18% for the smaller cache (10 documents) to about 36% for the larger cache (50 documents). As the hit ratio increases the amount of documents that have to be requested to the servers is decreased and the number of requests served for the mobile nodes is increased. As a consequence the traffic in the network is reduced (Figure 7.a) as well as the mean delay (Figure 7.b), the mean number of hops (Figure 7.c) and the mean number of timeouts (Figure 7.d).

The RWP mobility generates more traffic than MG for all the cache sizes although it obtains the better performance if we consider the mean number of hops. For the rest of the metrics (delay and percentage of timeouts) the RWP mobility model achieves a better performance than MG6 and MG8 but worse than MG10.

5.5 Effects of the density of nodes
Figure 8 illustrates the mean traffic (a), the mean delay (b), the mean number of hops (c), the percentage of timeouts (d) and the percentage of cache hit (e) as a function of the number of mobile nodes in the network.

As the node density increases the probability to find a route between the requester node and the server is also increased. So, the mean percentage of timeouts is reduced drastically (from 80~90% to 25%) as shown in Figure 8.d. For the lowest tested density of nodes (25 nodes) the RWP performs better than the MG because it obtains a better cache hit ratio (Figure 8.e). For node density greater than 25 nodes the difference in percentage of timeouts between the mobility models is reduced and all the scenarios obtain similar results for a network with 100 nodes.

Similarly, RWP obtains a lower mean delay than MG for low density networks as depicted in Figure 8.b, while for higher densities the mean delays are very similar. This fact is produced by the higher cache hit obtained by RWP. On the other hand, the RWP mobility model, as in the previous studies, obtains a lower mean number of hops (Figure 8.c) at the cost of injecting more traffic in the network (Figure 8.a).

5.6 Effects of the nodes’ speed
Figure 9 shows the mean traffic (a), the mean delay (b), the mean number of hops (c), the percentage of timeouts (d) and the percentage of cache hit (e) as a function of the node’s speed.

Figure 9.e shows that the cache performance does not depend on the nodes’ speed as the performance results are the same for the considered values of the speed.

As the nodes’ velocity increases the routes created between them are broken more frequently. Thus, the routes to the servers have to be created again. Consequently, the perceived delay augments as the nodes’ speed increases as shown in Figure 9.b. Due to the same reason, the percentage of timeouts is also increased as the nodes’ speed increases (Figure 9.d). On the other hand, RWP needs less hops to obtain the documents than MG as showed in the previous sections (Figure 9.c) while the required traffic is higher (Figure 9.a).
Fig. 7. Mean traffic, delay and hops, percentage of timeouts and cache hits as a function of the cache size
Fig. 8. Mean traffic (a), delay (b) and hops (c), percentage of timeouts (d) and cache hits (e) as a function of the number of nodes.
Fig. 9. Mean traffic (a), delay (b) and hops (c), percentage of timeouts (d) and cache hits (e) as a function of the nodes’ speed.
6. Conclusions

In this paper we have presented a caching scheme for Mobile Ad Hoc Networks that implements a local cache in each mobile node of the network. The mobile nodes have the capability of intercepting and responding the requests that they have to forward to the data server if they find a copy of the requested document in its local cache. On the other hand, the mobile nodes also implement a cache of document location in order to redirect the received requests to another mobile node that is known to be closer than the original destination of the request. This redirection cache is filled using the information obtained from the requests and replies that the nodes have to forward.

We have evaluated the performance of the proposed caching scheme through simulations using the mean generated traffic, the delay, the number of hops, the percentage of timeouts and the percentage of cache hits as performance metrics. We have compared the proposed caching scheme using the popular Random Way Point and the Manhattan Grid mobility models. The Manhattan Grid model has been evaluated using different topographical configurations (6x6, 8x8 and 10x10 blocks). In addition, we have evaluated the effect of several factors such as the mean time between requests, the documents' TTL, the request pattern (Zipf slope), the cache size, the nodes' density and the nodes' speed.

As main conclusions we can assert that the traffic generated using the RWP mobility model is greater than the traffic generated by the MG for all the parameters evaluated. Similarly the mean number of hops used by RWP is lower than that used by MG for all the performed simulations. If we consider the mean delay, the RWP mobility model performs better than MG when the distance between parallel lanes reduce the node connectivity (6x6 and 8x8 blocks) but worse than MG with a higher proximity of the lanes (10x10 blocks). The same results are obtained if the mean percentage of timeouts is taken into consideration. The cache performance is similar for all the studied parameters except for a low nodes' density where the network using the RWP mobility model obtains a better performance.

As the mobility model defines how the mobile nodes behaves in the network and the cooperating caching schemes depends on the behaviour of the mobile nodes, we can conclude that the mobility model used to evaluate a caching scheme clearly influences the obtained performance results of the network.

As a future research direction we suggest to evaluate the proposed caching scheme using more mobility models as those presented in section 2. On the other hand, the presented caching scheme has to be compared with other caching schemes in order to evaluate its effectiveness.

7. Acknowledgement

We would like to thank Adela Isabel Fernández Anta for revising the syntax and grammar of this paper. This work was supported by FEDER/MCyT funds (Project TEC2009-13763-C02-01).

8. References


Being infrastructure-less and without central administration control, wireless ad-hoc networking is playing a more and more important role in extending the coverage of traditional wireless infrastructure (cellular networks, wireless LAN, etc). This book includes state-of-the-art techniques and solutions for wireless ad-hoc networks. It focuses on the following topics in ad-hoc networks: vehicular ad-hoc networks, security and caching, TCP in ad-hoc networks and emerging applications. It is targeted to provide network engineers and researchers with design guidelines for large scale wireless ad hoc networks.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following: