Towards Intelligent Caching and Retrieval Mechanisms for Upcoming Proposals on Vehicular Delay-Tolerant Networks

Bruno M. C. Silva, Vasco N. G. J. Soares, and Joel J. P. C. Rodrigues *

Abstract: Vehicular delay-tolerant networks (VDTNs) are opportunistic networks that enable connectivity in challenged scenarios with unstable links where end-to-end communications may not exist. VDTN architecture handles non-real time applications using vehicles to relay messages between network nodes. To address the problem of intermittent connectivity, network nodes store messages on their buffers, carrying them through the network while waiting for transfer opportunities. The storage capacity of the nodes affects directly the network performance. Therefore, it is important to incorporate suitable network protocols using self-contained messages to improve communication that supports store-carry-and-forward operation procedures. Clearly, such procedures motivate content caching and retrieval. This paper surveys the state-of-the art on intelligent caching and retrieval mechanisms focusing on ad-hoc and delay tolerant networks (DTN). These approaches can offer important insights for upcoming proposals on intelligent caching and retrieval mechanisms for VDTNs.

Index terms: Vehicular Delay-Tolerant Networks; Delay-Tolerant Networks; Vehicular Ad-Hoc Networks; Mobile Ad-Hoc Networks; Caching; Web-caching

I. INTRODUCTION

Mobile ad-hoc networks (MANETs) [1] are autonomous ad-hoc networks where mobile hosts are connected by multi-hop wireless links. MANETs do not rely on common network infrastructures and services since nodes perform all the networking functions [2]. Typical applications of MANET include rescue operations [3], military scenarios [4], and in cases where is impossible to establish a wired backbone [5]. This communication technology can also be used for extending the coverage of current wireless networks. Vehicular ad-hoc networks (VANETs) [6] are considered a subset of MANETs and are presented as an important component of intelligent transportation systems (ITS) [7].

The key differences between MANETS and VANETs are the nodes mobility that can be anticipated due to the roads and traffic signalization (traffic lights) and also the rapid changes on the network topology taking into account vehicles velocity and movement [8]. Vehicles provide all network services relaying data between them. VANET networks can also include “infostations” that are wireless access points placed on the roadside. Infostations may be used to provide connectivity to the Internet for passing by vehicles. Then, vehicles can be exploited to propagate and deliver data across the network [9, 10]. Those data may include information about accident notifications, road conditions warnings and also Internet connectivity for several applications [11, 12] such as, electronic mail, Web browsing, audio and video streaming, among others. VANETs enable several other applications such as notification of traffic condition, advertisements [9], and cooperative collision avoidance [13].

VANETs are characterized by high mobility of vehicles, which is responsible for a highly dynamic network topology and to short contact durations. Moreover, limited transmission ranges, physical obstacles, and interferences lead to disruption, intermittent connectivity, and partition. Conventional routing protocols proposed for VANETs assume end-to-end connectivity. Thus, they were not designed to handle conditions caused by network disconnection, partitions, or long delays [14-17]. To answer these challenging connectivity problems, delay tolerant network (DTN) architectural concepts can be used [18].

DTNs are presented as a solution to challenged environments characterized by any or all of the following characteristics: intermittent connectivity, long and/or variable delay, asymmetric data rates, high error rates, and even no end-to-end connectivity [19]. DTNs introduce store-carry-and-forward functions by overlaying a bundle protocol layer above the transport layer, which provides internetworking on heterogeneous networks (regions) operating on different transmission media. The bundle protocol [20] is end-to-end, strongly asynchronous, message (i.e., bundle) oriented. The store-carry-and-forward paradigm avoids the need for constant connectivity. The idea behind it is to exploit node mobility to physically carry data between disconnected parts of the network. This approach circumvents the lack of an end-to-end path, enabling non real time (i.e., delay-tolerant) applications. [21]. DTNs have been applied to various scenarios such as interplanetary networks [22], underwater networks [23], wildlife tracking sensor networks [24], and networks for developing communities [25].
Vehicular delay-tolerant networks (VDTNs) [26] appear as new network architecture based on the concept of DTN and also gathering contributions from other network technologies that use out-of-band signaling and aggregation of IP datagrams at the core of the network (such as optical burst switching [27]), which is aimed for sparse or partitioned opportunistic vehicular networks.

VDTN architecture is inspired by the store-carry-and-forward paradigm of DTNs. Nonetheless, the VDTN layered architecture considers an IP over VDTN approach, placing the VDTN layer under the network layer, and assumes out-of-band signaling, with separation between control and data planes. This separation is a contribution of VDTNs in order to reduce the network complexity and management. Signaling messages are used when a contact opportunity is discovered. They carry information such as, but not limited to, node type, geographical location, route, velocity, supported link technologies properties (e.g., range, data rate), energy status, buffer status, bundle sizes, delivery options, routing state information, and security requirements, among others. The control plane uses this information to perform functions related with configuring, reserving and maintaining the necessary data plane resources. This approach is used to optimize data bundle transmission, performed at the data plane level. VDTN architecture has been successfully evaluated and validated through a VDTN Lab prototype [28].

VDTN networks allow delay-tolerant data traffic from a variety of vehicular applications to be routed over time, exploiting the physical movement of vehicles and the opportunistic links they establish with each other and with other network nodes. An example of a VDTN network applied to a rural connectivity scenario is shown in Figure 1. It shows three possible VDTN node types - terminal nodes, relay nodes, and mobile nodes. Terminal nodes act as access points to the VDTN network providing connection to end-users. It is assumed that at least one terminal node has direct access to the Internet. Mobile nodes (e.g., vehicles) are responsible for physically carrying data between terminal nodes. Relay nodes are devices with store-and-forward capability allowing bypassing mobile nodes to drop and pickup data bundles. These nodes increase the number of contact opportunities in sparse scenarios, contributing to improve the delivery ratio and decrease the delivery delay [29]. As stated above, in order to address the problem of intermittent connectivity and network partition, network nodes store bundles on their buffers, carrying them while waiting for transfer opportunities. At the same time, routing protocols may replicate bundles to discover more possible paths, and thus increasing the bundle delivery rate and decreasing the delivery delay [30]. If application protocols used in VDTNs use request-response or publish-subscribe paradigm, and its contents are relevant for various users (e.g., Web pages), then the combination of bundle storage during large periods of time and their replication can be further explored. If data bundles contain hints describing their content, this enables functions of resource retrieval with caching and distributed resource storage in conjunction with retrieval.

Such an approach minimizes end-to-end interactions and increases application performance [31].

This paper surveys the state of the art on intelligent caching schemes for ad-hoc and DTN networks. It reviews the state of the art on these topics gathering contributions for upcoming storage and retrieval proposals for VDTNs.

The remainder of the paper is organized as follows. Section II reviews the related work on Web caching mechanisms and inherent concepts. Section III addresses caching algorithms in ad-hoc networks including cache protocols for MANETs and VANETs. Available caching approaches for DTNs are surveyed in Section IV. Section V overviews the main contributions gathered from the presented projects and proposals including several conceptual base ideas of future content storage and retrieval mechanisms for VDTNs. Finally, Section VI concludes the paper and points further research directions.

II. WEB CACHING MECHANISMS

Caching mechanisms have been applied in several network architectures and communication technologies in order to reduce latency, bandwidth usage, and server load. Web caching principles may be also applied on rural vehicular networks. Then, this topic should be addressed in order to gather contributions for content storage and retrieval mechanisms for VDTNs that may be used on rural areas. By definition, a cache is basically a fast and temporary storage buffer. Cached data can be accessed in a future rather than the original data, saving computational resources. An important example of such technology is a proxy server [32]. A proxy server acts as an intermediate element for a client request and seeks the response in other servers or in a local cache. A caching proxy server improves the service providing a fastest response by retrieving content stored in previous requests [33]. Figure 2 illustrates a generic Web caching system with proxy cooperation. As may be seen, clients interact directly with proxies and they also cooperate with other proxies to provide a fast answer to the user’s request.
The next base station. Afterwards, mobile terminals are contacted from the closest to the farthest one, until receiving an acknowledgement of a mobile terminal that is storing the requested data, or there are no more mobile terminals to contact. To minimize network load and energy consumption, this approach uses timeouts instead of missed messages to notify mobile terminals that a requested object was not cached.

A cooperative cache protocol for ad-hoc networks was proposed in [45]. It presents an efficient approach to save energy consumption and minimize network load. The proposal presents the following three distinct cache protocols: CacheData, CachePath, and HybridCache. In the CachePath protocol, a node records only the data path instead of recording the information of all passing data. This procedure only occurs when the data is closer to the caching node than the data source. In the CacheData protocol, instead of paths, router nodes frequently cache accessed passing-by data for future requests. The HybridCache protocol use both CacheData and CachePath protocols. As demonstrated in [45], caching data locally or its path improves node performance by saving and improving storage space.

A cooperative caching algorithm, called Hamlet, is proposed in [46]. This caching strategy works independently on each network node deciding which contents are cacheable. These caching decisions are taken according to a probabilistic estimate of what nearby nodes may have stored in their caches and with the intention of differentiating its own stored content from the others. Hamlet can be applied to any content distribution system where network nodes can overhear query/response messages, estimate its distance from the query source node and the responding node in hops, identify the version of the transmitting information, and associate a specific dropping time to each item on its cache.

MANETs [1] are autonomous ad-hoc networks where mobile nodes interconnect each other’s by multi-hop wireless links. A MANET may operate in isolation or may have a gateway connection to a fixed network. A cooperative caching service for MANETs is presented in [47] and illustrated in Figure 3. This proposal addresses two important aspects in cooperative caching - cache resolution and cache management. Cache resolution concerns the decision mechanisms of the mobile device to find the requested data from the user. For this purpose, historical profiles and forwarding nodes are used to induce less communication costs. Cache management determines which data will be placed on/purged from the local cache. This management decreases significantly the number of cached copies between nodes, allowing a bigger distinctive data storage and improving the overall network performance [47].

Another cooperative caching scheme for MANETs is proposed in [48]. This proposal includes a distributed application software that implements ad-hoc cooperative Web caching among mobile terminals. This approach aims to improve Web latency among mobile terminals, optimizing their energy consumption and also increasing data availability. It introduces a local caching strategy that is adapted according to the capacity of the terminal. Every cached document is weighted according to its probability of being retrieved in the future and the energy cost associated with getting remotely the document. Documents with the lowest weights are removed from cache. The value of a document weigh is computed

Fig. 2. Illustration of a Web caching scheme with proxy cooperation.

III. CACHING MECHANISMS IN AD-HOC NETWORKS

Ad-hoc networks are wireless networks where nodes cooperate to forward traffic among them until the destination node is reached. Ad-hoc networks topology raises several fundamental research issues that are still being debated. One of these issues is the most appropriate caching strategy. The difficulty of solving this question appears from the fact that network nodes have limited storage capacity and also suffer from energy constraints [43]. In [44] it is proposed a caching solution to improve mobile performance, focusing on terminal’s latency, connectivity and energy constraints. Upon a request, each node uses local statistics of every interaction that identify mobile terminals that have most likely stored the requested data. It also uses a list of known mobile terminals and a routing table to find all nodes that are in lower range to the next base station. Afterwards, mobile terminals are
according to its Popularity, AccessCost, and Coherency values. Popularity value indicates the probability of future access, according to the number of times the document has already been requested since it has been cached. AccessCost value estimates the energy cost of getting the document remotely in case of being removed from cache. Coherency value balances the lifetime of a document and the energy remaining on the terminal.

Vehicular ad-hoc networks (VANETs) [8] are a subset of MANETs where vehicles deliver and propagate data throughout the network. Ott et al. presents the Drive-thru Internet project architecture [49]. This VANET project uses IEEE 802.11 [50] access points that provide several Internet services for moving vehicles in challenged environments with intermittent connectivity. To solve the disconnection problem this approach introduces the concept of performance enhancing proxies (PEPs) [51]. PEPs are usually network agents designed to improve end-to-end communication, breaking one connection into multiple connections. Drive-thru enhancing proxies are placed in a network managed by a Drive-thru service provider and are used as fixed points to relay mobile node requests to and from the Internet. Afterwards, they cache the requested data (e.g., mail services or Web browsers) waiting for a new communication opportunity to forward it [11].

A distributed caching scheme for DTNs is proposed in [31, 53]. It includes application hints to messages performing intelligent caching. The authors present scenarios for DTN applications like hypertext transport protocol (HTTP) or electronic mail (eMail), considering both content retrieval and caching scenarios as well as content storage and retrieval applications [25, 54]. Figure 4 illustrates a content retrieval and caching scenario where a user acts as a requester. The request is assumed to be a Web resource. A user sends a request (Req) for a resource (U), which travels through mobile nodes until it reaches a server. Then, the server will send a response (Resp) that travels all the way back to the user.

Following Figure 4, to retrieve the response (Resp(U)), the resource will be cached on every intermediate node until it reaches the requester. The request or the response may be replicated in various nodes depending on the routing algorithm being used. The replica stays stored on the node until the time to live (TTL) expires or the response gets to the requester (end-to-end DTN acknowledgement functionality) or a more recent response replaces it. If another user issues a request for the same resource (Req2 (U)), this request may cross nodes that have already a replicated response (Resp). In that case, an intermediate node that has that resource in its buffer can create a new response (Resp2 (U)) from its queued (Resp (U)) and send it as a reply to the user, acting as an implicit cache. Therefore, nodes must be able to identify the resource matching both requests (Req2 (U)) and (Resp1 (U)) and determine if the response is still the most recent one. This principle is illustrated in Figure 5.

Fig. 3. A cooperative caching scheme for MANETs.

**IV. CACHING MECHANISMS FOR DTN NETWORKS**

Delay-tolerant networks are characterized by intermittently connected links and long variable latency. The bundle protocol uses persistent storage to solve most of these issues [18]. A DTN node stores a bundle while waits for a future proper connection opportunity. When a communication opportunity occurs, the bundle is forwarded and this process is repeated with bundles relaying hop-by-hop until eventually they reach their destination. A scenario where DTNs can perform a perfect role is in highly disconnected networks such as proving connectivity to remote villages.

Cooperative caching and data prefetching techniques for such a connectivity scenario are presented in [52]. This proposal addresses the usability of limited computer networks in remote villages regarding Web pages request and specialized education applications. This approach uses collaborative caching allowing computers within a village to access Web pages stored on other connected machines. It also introduces a predictive prefetching technique that delivers Web pages that have not yet been requested from a client in a village, or even to clients in other villages, assuming that they will be requested in a near future. The advantage of this approach is that storage functions do not rely only on one single proxy server, thus eliminating single points of failure.

A distributed caching scheme for DTNs is proposed in [31, 53]. It includes application hints to messages performing intelligent caching. The authors present scenarios for DTN applications like hypertext transport protocol (HTTP) or electronic mail (eMail), considering both content retrieval and caching scenarios as well as content storage and retrieval applications [25, 54]. Figure 4 illustrates a content retrieval and caching scenario where a user acts as a requester. The request is assumed to be a Web resource. A user sends a request (Req) for a resource (U), which travels through mobile nodes until it reaches a server. Then, the server will send a response (Resp) that travels all the way back to the user.

Fig. 4. Illustration of request and response operations between user and server.

Following Figure 4, to retrieve the response (Resp(U)), the resource will be cached on every intermediate node until it reaches the requester. The request or the response may be replicated in various nodes depending on the routing algorithm being used. The replica stays stored on the node until the time to live (TTL) expires or the response gets to the requester (end-to-end DTN acknowledgement functionality) or a more recent response replaces it. If another user issues a request for the same resource (Req2 (U)), this request may cross nodes that have already a replicated response (Resp). In that case, an intermediate node that has that resource in its buffer can create a new response (Resp2 (U)) from its queued (Resp (U)) and send it as a reply to the user, acting as an implicit cache. Therefore, nodes must be able to identify the resource matching both requests (Req2 (U)) and (Resp1 (U)) and determine if the response is still the most recent one. This principle is illustrated in Figure 5.

Fig. 5. Illustration of a request and a response between an user and a network node.
In a content storage and retrieval scenario, a user requests storage (streq) of a resource (U) to a server (Figure 6). The storage request (streq (U)) travels through the nodes until it reaches the server. On its path, the resource (U) will be replicated along the traversed nodes. While stored/queued, the resource (U) is available to any requester that is interested in it. The nodes must identify the local copy of the resource and create the response from previous storage requests (streq (U)). This mechanism may act as data backup for stored resources in case of any hardware loss or failure.

![Fig. 6. Illustration of storage operations on network nodes, which can perform backup services in a server failure scenario.](image)

These scenarios clearly present a distributed caching strategy in DTN nodes. In order to improve this distributed caching scheme, in [55], a logical module for storage and retrieval operations is proposed. This module allows cache lookups for stored resources in the queue and cache retrieval for passing requests. Such procedures require additional information about bundle payloads. For this purpose, application-hints extension blocks for the bundle protocol were presented [53]. The application-hint illustrated in Figure 7 contains several cache control fields. Among others, the application protocol field shows the used application protocol (e.g. HTTP, FTP) to perform proper resource matching. A node may only give support and preferential treatment for several applications. Resource identifier identifies the carried resource. Operation type indicates whether the resource is a request, a response, an unconfirmed event, or an unknown operation type and if a bundle contains a resource. In case of bundle fragmentation, the application hint must be copied to all fragments.

![Fig. 7. Illustration of an Application-Hint extension block for the bundle protocol.](image)

**V. DISCUSSION TOWARDS CONTENT STORAGE AND RETRIEVAL PROPOSALS FOR VDTNs**

Both ad-hoc and VDTN networks may assume that network nodes cooperate among them to forward data traffic. The cooperative caching scheme presented in [45] uses the requester nodes ids and node paths to easily cache and retrieve contents, saving energy consumption and minimizing network load. In a VDTN context, this caching and forwarding solution that identifies previous requester and receiver nodes could improve the forwarding of specific cached contents, reducing network load significantly, contributing for energy, time, and bandwidth saving. The above-described proposals presented in [47, 48] furnish several mechanisms to find a specific requested data and to determine which data will be placed or purged from a local cache. These caching schemes can also be adaptive according to the storage capacity of the terminal. Such approaches can contribute to a similar upcoming proposal for VDTN, decreasing significantly the number of cached copies between VDTN nodes, allowing a bigger distinctive data storage and improving the network performance. The Drive-thru Internet project [49] includes the use of performance enhancing proxies (PEPs) [51] on VANETs and this approach offers insights on caching and forwarding procedures that analogously could be applied to VDTN relay nodes.

VDTN networks gather important contributions from mobile DTNs, therefore the above-mentioned DTN caching schemes also offer further directions to an upcoming VDTN storage and retrieval solution. The above-described proposals address important aspects that are equally treated in VDTNs (congestion problems, content storage and retrieval, content retrieval, and caching). The collaborative caching proposal presented in [31] offers insights on similar techniques that can be perfectly adapted and implemented on VDTN relay nodes that can be used as backup in a server failure scenario. In [53], it is proposed to use application-hints extension blocks for the bundle protocols. These application-hints contain several cache control fields allowing different content storage policies. Applying a similar technique to the VDTN architecture implies a new approach designed to the specifications of the VDTN bundle protocol. A further storage and retrieval proposal for VDTNs based on the concept of application-hints must consider that, in a VDTN, a bundle protocol data unit is formed by aggregating several IP packets, based on an aggregation strategy. Then it is suitable to consider that in a VDTN-like proposal, bundles and its contents must have different storage and retrieval mechanisms.

Table I summarizes the above-described main caching and retrieval issues in VDTNs and the relationship and possible contributions with the reviewed solutions in web caching, ad-hoc networks and DTNs.
TABLE I
POSSIBLE CONTRIBUTIONS AND SOLUTIONS FOR CACHING AND RETRIEVAL ISSUES IN VDTNs.

<table>
<thead>
<tr>
<th>CACHING AND RETRIEVAL ISSUES IN VDTNs</th>
<th>WEB-CACHING CONTRIBUTIONS</th>
<th>AD-HOC NETWORKS CONTRIBUTIONS</th>
<th>DTN CONTRIBUTIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENERGY CONSUMPTION AND MINIMAL NETWORK LOAD</td>
<td>- Typical Web caching schemes with proxy cooperation.</td>
<td>- The Cooperative caching scheme presented in [45] uses the requester nodes' ids and node paths to easily cache and retrieve contents.</td>
<td>- The use of application-hints extension blocks [53] and its specific cache control fields that allows a resource matching for a bigger distinctive and selective data storage.</td>
</tr>
<tr>
<td>CACHING AND RETRIEVAL OF SPECIFIC CONTENT</td>
<td>- Cooperative caching that basically introduces a cache on every network node.</td>
<td>- In [45], the identification of previous requester and receiver nodes contributes to improve the forwarding of specific cached contents.</td>
<td></td>
</tr>
<tr>
<td>CACHING AND FORWARDING CONTENT IN A CHALLENGED ENVIRONMENT WITH INTERMITTENT CONNECTIVITY (DISCONNECTION PROBLEM)</td>
<td>- Not available</td>
<td>- Proposals presented in [47, 48] furnish several mechanisms to find a specific requested data and to determine which data will be placed on/purged from a local cache.</td>
<td></td>
</tr>
</tbody>
</table>

VI. CONCLUSIONS AND FUTURE WORK

Vehicular Delay Tolerant Networks (VDTNs) are a DTN-based disruptive network architecture where vehicles act as the communication infrastructure. This paper surveyed available research proposals on cache-based technologies focusing on ad-hoc and DTN networks.

Since mobile nodes (e.g., vehicles) store and carry data on their buffers waiting for contact opportunities, the buffer space availability has a direct effect over the network performance. Therefore, it is important to incorporate suitable network protocols using self-contained messages to improve communication that supports store-carry-and-forward operation procedures. Content storage and retrieval mechanisms appeared as a promising solution, fully automated, that manages the nodes buffer and decides which data is to be stored, sent, or eliminated.

The above-surveyed caching and forward solutions include related technologies (and architectures) to VDTNs. Thus, such approaches may be analogously applied on VDTNs. This study gathers contributions and conceptual approaches from the related literature towards upcoming solutions for storage and retrieval mechanisms on VDTNs. It may be considered for future works.

REFERENCES


Joel J. P. C. Rodrigues is a professor in the Department of Informatics of the University of Beira Interior, Covilhã, Portugal, and a researcher at the Instituto de Telecomunicações, Portugal. He received a PhD degree in informatics engineering, an MSc degree from the University of Beira Interior, and a five-year BS degree (licentiate) in informatics engineering from the University of Coimbra, Portugal. His main research interests include sensor networks, high-speed networks, delay-tolerant networks, e-health, e-learning, and mobile and ubiquitous computing. He is the editor-in-chief of the International Journal on E-Health and Medical Communications. Prof. Rodrigues has served as General Chair, Technical Program Committee Chair, and symposium Chair for many international conferences. He is a member of many international TPCs and several editorial review boards. He has authored or coauthored over 150 technical papers in refereed international journals and conferences, a book, and a patent. He is a licensed professional engineer (as senior member), and he is member of ACM SIGCOMM, a member of the Internet Society, an IARIA fellow, and a senior member of IEEE.

Bruno M.C. Silva received the BSc and MSc degrees in Informatics Engineering from University of Beira Interior, in 2008 and 2010, respectively. He is currently a PhD student on Informatics Engineering at the University of Beira Interior under supervision of Prof. Joel J. P. C. Rodrigues. He is also a PhD student member of the Instituto de Telecomunicações, Portugal. His current research areas include Delay Tolerant Networks, Vehicular Networks, Mobile Computing and Ubiquitous Computing. He authors or co-authors 8 international conference papers and two journal publications.

Vasco N. G. J. Soares received his 5-year B.S. degree (licentiate) in 2002 in Informatics Engineering from University of Coimbra, Portugal. He is an Assistant Professor in the Informatics Engineering Department at the Superior School of Technology of the Polytechnic Institute of Castelo Branco, Portugal. He is currently a PhD student on Informatics Engineering at the University of Beira Interior under supervision of Prof. Joel J. P. C. Rodrigues. He is also a PhD student member of the Instituto de Telecomunicações, Portugal. His current research areas are Delay Tolerant Networks, and Vehicular Networks. He authors or co-authors more than 12 international conference papers, three journal publications, and one book chapter publication. He participates on several Technical Program Committees.