Vehicular networks and the future of the mobile internet

Mario Gerla *, Leonard Kleinrock

Computer Science Dept. UCLA, 405 Hilgard Ave, Los Angeles California 90024, USA

A R T I C L E   I N F O

Article history:
Available online xxxx

Keywords:
Vehicular network
VANET
Mobile Internet
Wireless network architecture

A B S T R A C T

The first DARPA experiment with wireless mobile Internet – the Packet Radio Network or PRNET – was completely independent of the infrastructure. This model was consistent with DARPA military goals as the PRNET was designed to support tactical operations far away from any wired infrastructure. Beside autonomy, the main challenge was mobility and radio portability. Scarcity of spectrum was not an issue, in contrast with the ARPANET quest to utilize those 50 kbps Telpak trunks more efficiently. Today, the closest civilian descendents of the PRNET are vehicular networks and smart phone based Personal Area Networks. In either case, the wired infrastructure turns out to play a major role. Moreover, spectrum scarcity has now become the most important challenge (while, ironically, the wired Internet has plenty of bandwidth). In this paper we examine this interplay between wired and wireless and extract a message for the design of a more efficient Future Wireless Internet. We focus on the vehicular network since this field is better established and commercially more viable than that of personal, P2P communications among Smartphones. We are confident however that many of our observations will transfer also to smart phone/infrastructure synergy.

Specifically, in this paper we identify the urban Internet infrastructure role in the support of emerging vehicular applications and identify the Core Internet services matching the services in the vehicle grid. As the vehicular applications range from e-mail and voice over IP to emergency operations (natural disaster, terrorist attack, etc.), the type of assistance requested from the infrastructure will vary. A short list includes: (a) addressing (e.g. geo-addressing); (b) directory service, service discovery, mobility management; (c) resource and congestion management; (d) path redundancy; (e) delay tolerant operations; (f) mobile sensor data access and search from the internet, and; (g) anonymity, privacy and incentives.

After the review of vehicular applications and properties, we will offer an Internet history perspective to help understand how the mobile wireless network field has evolved from the early ARPANET and PRNET days. This will reveal trends that can help predict the future of the wireless Internet.

* Corresponding author.
E-mail address: gerla@cs.ucla.edu (M. Gerla).

1. Introduction

Vehicle communications are becoming increasingly popular, propelled by navigation safety requirements and by the investments of car manufacturers and Public Transport Authorities. The essential vehicle grid components (radios, Access Points, spectrum, standards, etc.) are coming into place finalizing the concept of VANET (Vehicular Adhoc Network) and paving the way to unlimited opportunities for car-to-car applications. Safe navigation has now become an important priority for Car Manufacturers as well as Municipal Transportation Authorities. New standards are emerging (DSRC and more recently IEEE 802.11p) and several International Consortia and well publicized testbeds were recently established to promote vehicular communications and demonstrate their feasibility and effectiveness. In this paper, we look at the likely...
trajectory of these emerging applications. We seek to predict how the VANET will evolve in the future and, at the same time, to explain how we got here as this will help corroborate our predictions. Our specific interest is to examine the role of the Internet infrastructure in supporting vehicular applications. As a difference from typical MANETs (Mobile Ad hoc Net works), in the VANET the Wired Internet infrastructure is omnipresent and readily accessible via WiFi, DSRC, WiMAX, 3G, LTE, etc. The wired Internet support will vary depending on the vehicular application. During a Katrina type emergency, the wired infrastructure is likely to be destroyed and the VANET cannot count on it – it must operate autonomously. This is the closest we get to a pure MANET. For content sharing (e.g., advertisements, entertainment) applications, however, the vehicle relies on Internet access to download files from Internet Servers. Peer to peer, “opportunistic” networking among vehicles will be used to distribute the file only after some vehicle has downloaded from the Internet. In fact, in the future, considerable content will be played on mobile terminals (vehicle screens or pedestrian phones). It is thus important to understand how the wired Internet can best accommodate the new requirements and networking styles (e.g., P2P dissemination, geo-routing) introduced by the “mobile” applications at the edge.

The “business as usual” solution of treating mobile, wireless networks as appendages attached via edge gateways and to delegate all interface functions and services to the edge is becoming non economical. There are benefits in extending the services to the Internet backbone itself using mesh networks, overlays and virtualization. We envision that the emerging mobile networking will impact the future Internet design in the following areas: (a) addressing (e.g. geo-addressing); (b) directory service support and service discovery; (c) resource and congestion management; (d) path redundancy and network coding; (e) delay tolerant operations support; (f) mobile sensor data access and search from the Internet (g) anonymity, privacy and incentives.

The following sections review related work and then describe two research directions, namely: (1) emerging applications and (2) vehicle oriented Internet services. Representative vehicular applications have been selected to illustrate how mobile services can be extended into the Internet core, beyond edge gateways, as a first step towards the design of a full fledged mobile vehicular architecture. In particular, since mobility plays a key role in vehicular protocol design and performance, an important effort must go into the development of realistic motion models.

2. Related work

Much of the existing literature on Inter-Vehicle Communications (IVCs) is navigation safety related. A good survey of recent physical layer technologies for IVCs can be found in [18]. For scalable delivery, researchers have proposed geo-routing and further, have focused on exploiting innate characteristics of vehicular networks such as high, but restricted mobility. For example, Urban Multi-hop Broadcast (UMB) [14] features a form of redundant flood suppression where the furthest node in the broadcast direction from a sender is selected to forward and acknowledge the packet. In [39], vehicles collect only the information relative to a given locality (i.e., a road segment); the paper further investigates the influence of broadcast rate on data propagation taking mobility into account and adapting the rate to traffic conditions. In this field, we have proposed a scheme by which each vehicle is able to estimate its transmission range and put it to good use to reduce redundant transmissions. As a result, broadcasting traffic and delays are reduced thus allowing efficient delivery of, for instance, alert messages for traffic safety applications [32,34].

Geo-routing has been extensively investigated. A critical issue is the relaxing of beacon message requirements to eliminate the associated overhead. Füßler et al. [8] has proposed Contention-Based Forwarding (CBF). This scheme does not require proactive transmission of beacon messages for current location advertisements; instead, data packets are broadcast to all direct neighbors and the neighbors themselves decide if they should forward the packet based on a distributed timer-based contention process. A similar approach has been proposed by Zorzi in GeRAF [44] exploiting staggered MAC inter-segment intervals. In [17], the authors have proposed a set of knowledge-based opportunistic forwarding protocols that use geographic information such as motion vectors. Zhao and Cao [42] propose to reduce the delay to a known destination through mobility prediction.

Geocasting services have been proposed to disseminate messages to all nodes within a target region. MDDV [40] aims to support geocast by forwarding a packet along a predefined trajectory geographically. It works even with intermittent connection; intermediate vehicles must buffer and forward messages opportunistically, exploiting mobility. “Abiding Geocast” [26] features a lifetime constraint; namely, it restricts the delivery of messages to all the nodes that are in the geocast region “sometime” during the geocast lifetime.

At the applications level, several cooperative peer to peer type schemes have been proposed. TrafficView [28] disseminates (through flooding) and gathers information about the vehicles on the road, thus providing real-time road traffic information to drivers. To alleviate broadcast storms, this work has focused on data aggregation based on distance from the source. EZCab [43] is a cab booking application that discovers and books free cabs through vehicle multi-hopping. Free cabs are discovered with probabilistic flood search (static or decreasing probability as hop count increases). After discovery, georouting is used to negotiate the fare etc. [41] has proposed an opportunistic resource discovery protocol with a finite-buffer space model. The resource is a spatio-temporal resource, e.g., the availability of parking in a parking lot. A vehicle either “senses” the resources or obtains new resources from passing vehicles. Nodes exchange local databases and each keeps a fixed number (the size of buffer) of relevant resources. In PeopleNet [27], a wireless virtual social network is used to support searching for spatio-temporal information, exchanging resources by random swapping.
Vehicular Information Transfer Protocol (VITP) [4] provides on-demand, location-based, traffic-oriented services to drivers using information retrieved from vehicular sensors. A user “location-aware” query is forwarded to the target location where virtual ad hoc servers (VAHS), i.e., collection of private vehicles, resolve the query. Content sharing in MANETs as well as VANET mainly involves cross-layer optimization of Internet-based protocols. Most protocols have been designed to overcome the discrepancy between a logical overlay and a physical topology of mobile nodes. XL-Gnutella [1] maps the logical overlay neighbors to physical neighbors. The UCLA group has proposed CarTorrent [31], a BitTorrent style content sharing protocol in wireless networks. It uses a proximity-driven piece selection strategy that has proven better than the rarest first piece selection. We have designed CodeTorrent [59] to provide a BitTorrent style content distribution with network coding [10].

To stimulate cooperation among selfish nodes in mobile ad hoc networks, incentives have been proposed. Yet, abuse and forgery must be prevented. For example, with the assumption of tamper-proof hardware on board [48,49], each relaying node earns some virtual credit that is protected by the tamper-proof hardware. Other “secure” incentive approaches make use of reputation-based schemes [25,46,47,52]. Uncooperative nodes are detected and isolated. Researchers have investigated such non-cooperative communication scenarios within a game theory framework [50,53]. By manipulating the parameters (e.g. the amount of gain per forwarding, the designation of charging subject, etc.), those schemes encourage cooperative behavior among selfish nodes. However, as pointed out in [51], if poorly implemented in practice, these incentive schemes themselves have the potential to backfire by offering an incentive to cheat the system in order to gain further benefits.

The advantages of using the infrastructure to enhance the per node throughput capacity of an ad hoc network are well documented in [58]. The asymptotic capacity is found to be $\Theta(\sqrt{N/\log N})$-fold better than in a flat ad hoc network, mainly due to the fact that relaying over APs effectively reduces the mean number of hops from source to destination. Most proposed hybrid (infrastructure + ad hoc) networks are to provide extended coverage of existing services, e.g., wireless LAN and 3G. Re-routing AODV packets over the infrastructure has been proposed. However, flooding of RREQ packets over APs does not scale as the number of mobile nodes and APs increases. To overcome this problem, an Overlay Location Service (OLS) can be used, which maintains geographic locations of APs and mobile nodes, and allows mobile nodes to efficiently utilize geo-routing not only over the vehicular grid but also over the Internet [54]. In addition, OLS provides a “global” view of AP congestion levels, thus leveraging efficient use of communication resources.

A wide range of emerging urban monitoring applications are clear proof of growing interest in the field. Intel Research IrisNet [9] provides large-scale monitoring based on Internet-connected PCs equipped with off-the-shelf cameras and microphones. IrisNet also supports urban monitoring through a group of agents that collect and process raw data to answer queries relevant of the application: for instance, camera data is processed to track available parking in a metropolitan area. CarTel [12] is a distributed mobile sensor computing system. It is composed of intermittently connected database (ICEDB) and carry-and-forward network (CafNet). ICEDB is installed in both clients (mobile user side) and a server (portal), allowing users to send queries. Queries are then sent to mobile users via CafNet. Clients then resolve the query and return the results to the server. ICEDB extends SQL queries to support user friendly interface to a portal and provide efficient management of sensing devices. Our Team has proposed MobEyes [22] a “distributed” proactive urban monitoring solution. Vehicles sense events, process sensed data and opportunistically diffuse meta-data by exploiting vehicles’ mobility.

As we mentioned earlier, safe navigation and intelligent transport have now become important priorities for Car Manufacturers as well as Municipal Transportation Authorities. VANET research is moving out of academia and is now carried out in many national and international projects that bring together government, industry, and academia. These projects have broad reach and go beyond fundamental research to address also development, standardization, field trials and commercialization. Among the ongoing programs that address safe and efficient driving we mention the eSafety framework of the European Union, the Intellidrive initiative in the US, Smartway, DSSS (Driving Safety Support System) and ASV (Advanced Safety Vehicle) in Japan, simTD in Germany and SCOREF in France. Standardization is well under way with the activities worldwide in ISO TC204 and IEEE (802.11p and 1609.x), SAE J2735 in the US, ETSI TC ITS and CEN WG278 in Europe and ARIB T-75 in Japan.

The European Union recently has been very active in this field, sponsoring several projects that deal with the vehicular environment, like CVIS (for the vehicular network architecture definition), Coopers (for the cooperative systems potential for road safety), Geo-Net (for geo-addressing and routing), PRE-DRIVE C2X (for the applications support and V2V and V2I proof of concept as core technologies for field operating tests), iTETRIS (for the modeling and simulation), Sevecom and Preciosa (about security architecture and privacy), EVITA (about the in-vehicle tamper-proof network architecture modules) just to mention a few of the most recent (and still ongoing) projects touching the themes of this paper.

3. Vehicular applications and Internet requirements

Vehicles in a grid are only a few hops away from the infrastructure (WiFi, cellular, satellite, etc.). Protocol and application design must account for easy access to the Internet during normal operation. At the same time, the vehicles are among the few communications nodes that can continue to operate when the Internet goes away, during urban emergency, with enough reserve power to establish a vehicle based emergency network. To this end we examine innovative peer to peer content sharing applications that can still operate with intermittent connectivity and sporadic vehicular traffic and connectivity. Peer to
peer applications have so far been confined to the fixed Internet (e.g., BitTorrent, etc.). The storage and processing capacity of modern vehicles make such applications feasible also on mobile platforms.

In these dynamic scenarios we must understand the role of the Internet in facilitating the smooth transition from full Internet connectivity to full autonomy. This is a radical concept in ad hoc networks traditionally designed for exclusively autonomous operation and thus unable to exploit the interconnection and resource sharing of the wired Internet. In the sequel, we consider a number of emerging VANET applications and study their interdependence with the Internet.

3.1. Content downloading

Probably the application that best illustrates the interaction between wired Internet and vehicle grid is peer to peer content distribution. Beyond conventional file downloads from the Internet (i.e., movies, IPTV), the drivers in the vehicular grid are interested in access to “location dependent and aware” content. This includes not only delay-sensitive content (e.g., emergency-related video streaming [55]), but also delay-tolerant content, namely proximity advertising and marketing segments (e.g., movie clips from the nearby theaters [30,31]). Roadside access points (e.g., at gasoline stations), or digital billboards are used for this purpose. Limited availability of access points and/or limited AP capacity encourage the mobile users to cooperatively assemble the file through BitTorrent-style, P2P file sharing (Fig. 1). A file is divided into pieces and missing pieces are pulled when available from the neighborhood (say up to K hop deep). The car then selects the “best” peer for download. Simulation experiments in [29] show that the best strategy combines closeness and rarity of the “piece” (while BitTorrent generally selects the “rarest” piece). This however involves quite a bit of overhead, namely, peer selection (from several responses) is followed by TCP transfer of the pieces (typically, with several retransmissions).

An architectural feature that can considerably enhance P2P content sharing is network coding [10]. One immediate advantage is the relaxation of the piece selection (i.e., “coupon collection”) problem, a critical issue at the end of the download. Access points generate/distribute coded pieces that are random linear combinations of the original pieces, and the intermediate nodes further mix coded pieces in their buffers. Assuming that each piece is exactly one packet long, a prefix in the packet tells the weights of the linear combination. If the receiver does not have enough packets to “invert the matrix” and recover the file (e.g., some packets may have been lost), it simply requests more random combinations from neighbors, as opposed to requesting specific packets. The network coding scheme offers several advantages. It drastically reduces the number of message requests (seeking a specific piece) and thus the overhead. No TCP is required, just UDP. In fact, TCP will not work with multiple simultaneous sources. Simulation results in [23] confirm that routing overhead in highly mobile/dense network like VANET, is prohibitive, and thus, UDP-based “single-hop” pulling outperforms TCP-based “multi-hop” pulling. Network coding for P2P sharing has been proposed also for the wired network (e.g. Avalanche file sharing by Microsoft [11]). The role of the Internet becomes clear if one considers that, when vehicles download real time streams, they may download from several different APs. Each AP must provide a coded stream consistent with the others. To this end, the “torrent” coordinator in the Internet must anticipate which APs must be served and with what “generations”. Another enhancer of vehicu-

![Co-operative Download: Car Torrent](image_url)
lar content sharing is parallel downloading. Suppose vehicles on the freeway can download with GPRS, at top speed = 100 kbps. This gives very poor video quality. However, if users team up, and the Internet server uses multiresolution codes, vehicles in a neighborhood can exchange packets and combine the multiple streams with quasi linear performance improvement with respect to a single stream. These are all examples of architectural choices (Network Coding and Multiresolution Codes) that tie together vehicles and infrastructure (Fig. 2).

3.2. P2P location significant advertising

Another application that benefits from multiple neighbors downloading is “Ad Torrent” [29]. Say, a driver needs to download previews of movies are playing in a particular neighborhood. And, the same driver wants to dine at an Italian restaurant after the movie. Videoclips, menus, restaurant reviews and addresses must be acquired within tight latency constraints. Driving up to an access point each time is too time consuming and traffic congestion prone. Multihop downloading from a remote access point or from LTE is not practical due to wireless TCP limitations. It may also create excessive data traffic overload on the system. As an alternative, in Ad Torrent the access point feeds passing cars with randomly selected “ad segments”. Next, each car probabilistically disseminates the pieces using an epidemic (“gossip”) scheme (we later review the effectiveness of epidemic dissemination in the vehicular context). As a result the neighborhood becomes populated with ads. Again, there are different neighbor download strategies. One method [30] is to query the neighborhood and selectively download pages that satisfy multi-value queries using Bloom Filters. Another approach is to solicit network coding downloads from neighbors of whatever pieces, they have, that are useful to the driver. The main difference from Car Torrent is that Ad Torrent epidemically disseminates segments (to increase the hit ratio of even not so popular files); also, it downloads from “3rd party” peers who are not trying to assemble the information themselves. This may have impact on tit-for-tat incentives bookkeeping. Again, the Internet plays an important role; for example, the APs will deliver the Ads that are most popular in the area. The learning must be coordinated in the Internet servers since the individual vehicle interests are ephemeral.

3.3. P2P (driver to driver) interaction

There is also content that is generated and consumed entirely in the vehicle grid. A good example is content and messages related to navigation safety. Suppose that a critical traffic/safety situation occurs on a highway, e.g., major traffic congestion, weather condition, natural or manmade disaster or even hostile attack. In such cases, multimedia content, say, video, could be streamed from one or more lead cars to the vehicles following several miles behind – to “visually” inform them of the problem. This will allow them to make a better informed decision (say, whether they should turn around) than if they simply got an alarm text message. Conventional ad hoc broadcast (e.g., via ODMRP or MAODV) may introduce excessive loss in intermittent connectivity and severely impair video reception. In the intermittent situation network coding can greatly enhance stream reliability. Ad hoc Network Coded broadcast, CodeCast, improves delivery ratio as compared to ODMRP reducing at the same time the overhead.
3.4. Sensing the environment

Vehicular networks are emerging as important sensor platforms, for example for proactive urban monitoring and for sharing and disseminating data of common interest. Each vehicle can sense one or more events (e.g., imaging from streets and detecting toxic chemicals), process sensed data (e.g., recognizing license plates), and route messages to other vehicles (e.g., diffusing relevant notification to drivers or police agents). Vehicles can generate much larger volumes of data than traditional sensor networks. They can also store the data, instead of periodically reporting it to sinks [56]. Building upon previously proposed techniques for epidemic information dissemination in mobile ad hoc networks with pedestrian mobility [24,35], a lightweight strategy for proactive urban monitoring is proposed in MobEyes [21]. The basic idea is to exploit vehicle mobility and wireless broadcast to opportunistically diffuse concise summaries (meta-data) of the data stored in cars. The data can be harvested by agents for forensic investigation. It is discarded after timeout.

The use of MobEyes is not limited to vehicular sensors. It can be adapted to embrace sensors embedded in the environments (e.g., fixed video and road sensors, etc.), or in the human body (e.g., pedestrians with medical sensors). Interestingly, these types of sensors are associated with the infrastructure. For instance, video cameras installed in the intersections are connected to Police Headquarters via private wire connections. Medical sensors, e.g., ECG probes, require interaction with the attending physician in case of emergency and exchange messages via the cellular network through the smart phone. An emerging area of research is the seamless integration of heterogeneous sensing environments. This integration must address issues such as application constraints (e.g., latency, coherency of observations, etc.), infrastructure support (deployment of fixed sensors, Internet connectivity), mobility patterns (e.g., group mobility or individual mobility), compatibility of different radio technologies (e.g., WiFi, ZigBee, Bluetooth), etc. In addition, alternative indexing schemes must be explored between the two extremes of metadata uploading to the Internet Server or epidemic dissemination. We envision hierarchical schemes that combine epidemically generated indexes with higher level structured indices (e.g., GHT [36] and DHT [57]), some of which will be stored in the Internet infrastructure (Fig. 3).

4. Network and Internet services

4.1. Routable addresses and position based addressing

Addressing is a major challenge in the management of vehicular network mobility and an important enabler of interconnection to and through the Internet. First, we must distinguish between Unique Identifier (e.g., license plate#: Vehicle-ID#), and; Routable Address (geo-coordinates, or; unique ID (typically IP address) for conventional routing, e.g. AODV). It is becoming apparent that the dominant form of routing in the vehicle grid will be position based routing (e.g., geo-routing). This is because of the emergence of location aware communications, i.e., the need to establish connections and route packets to entities and resources characterized by location rather than a specific ID. More traditional MANET routing schemes, e.g. AODV and OLSR, will also be used in the vehicle grid. These schemes currently use IP address as the routable address to set up/maintain the routes. The IP address is an extremely effective routable address in the static, hierarchical Inter-
net structure (enabling, for example, prefix routing etc.). It is not very helpful in finding (hierarchical) routes in a constantly changing network like the vehicle grid unless it is combined with the Mobile IP construct, with provides the desired redirection. Mobile IP is fine for movable nodes (like laptops), but it does not scale well in very dynamic population like vehicles. Still, IP is widely used as identifier in AODV routing and it is required by most Internet upper layer protocols like for instance TCP. For these reasons, it is important to maintain in the vehicle grid a “unique” IP address for cars, to avoid collisions in TCP connections and routing mishaps in AODV. There have been many proposals for enforcing unique IP addresses in MANETs. A prominent example is the ‘IP address auto-configuration’ of nodes [33]. However, auto-configuration of IP addresses, such that the assignment is unique, requires some extra work in the vehicle grid scenario. Conventional MANET solutions cannot be directly applied due to the high density of nodes; high absolute speed (20–80 mph) (but, low relative speed with respect to other cars traveling in the same direction (3–20 mph)); and practically unbounded network diameter (hundreds of thousands of cars). One possible solution is to implement a leader based approach. Leader approaches generally use a hierarchical structure to configure nodes and perform the DAD (Duplicate Address Detection) procedure only within a cluster [37,38]. A recent solution that is more directly inspired to vehicular traffic envisions leaders proactively organized in a chain (in the same direction of traffic along a highway, say) and operating like DHCP servers to assign (and manage) unique IP addresses to vehicles within their range. Moreover, they maintain uniqueness not only in their cluster but also in the entire chain, exploiting the low relative speed and long contact time of vehicles in the same chain moving in the same direction. This guarantees higher longevity than traditional leader schemes [5–7].

After solving the unique IP address issue, one must focus on “routable address” selection. The goal is efficient packet delivery between cars in the vehicular grid as well as between Internet Servers and mobile vehicles. The routable address clearly goes hand in hand with the routing strategy. Geo-routing fits the bill. With adequate architecture support, geo-routing takes a packet all the way to the “neighborhood” of the target destination. From that point, the receiver unique identifies (e.g. IP address) carried in the packet header will enable correct delivery.

To accomplish the delivery to the “neighborhood”, the source must know with a reasonable accuracy the location of the destination. A critical component of the geo-routing address structure is the Geo Location Service (GLS) - a distributed service that maps a vehicle name to the set of most recent geo locations. If the VANET must be capable to operate autonomously in emergency mode, there must be seamless transition from infrastructure supported mode to autonomous mode. A possible solution is to implement two “parallel” versions of GLS, namely: OLS (Overlay Location Service) and VLS (Vehicle Grid Location Service). OLS is maintained within the Internet infrastructure with overlay technology; VLS is maintained entirely in the vehicle grid, possibly using Geo Hash Table (GHT) technology [36]. The two services are synchronized, but are independently maintained to provide tolerance to total infrastructure failures.

Let us illustrate a possible OLS implementation. An overlay structure is established in the urban Internet. Each car, whenever it passes by an AP, registers its ID (license#, IP address(es), time, owner name, owner IP address billing address, etc.) and the current geo-location. OLS maintains an index of IDs mapped to geo coordinates. The index is managed via DHT (Distributed Hash Tables). Suppose now Host A (fixed or mobile) wants to establish a TCP connection to mobile Host B. Host A first queries OLS with the unique car ID, say: 3MANDC@CA.car starting from the nearest server of the OLS “overlay”. After proper authentication of the query by OLS and verification that this request is authorized by the receiver, Host A gets back the “most recent” geo-locations, the IP address, etc. of Vehicle B. It then extrapolates from past locations the future location of B and the access point AP nearest to B. Host A sends the TCP connection set up message to AP. The message is encapsulated in an IPv6 network envelope that contains the geo-address in the extended header. Routing in the Overlay is based on geo-addresses. Namely, the geo address defines the AP at the end of the Internet path. At the destination, the AP geo-routes the packet through the ad hoc net; vehicle B responds with its own IP address and directs the response (encapsulated in the overlay envelope) to the sender IP. The encapsulation in the geo routed envelope is identical regardless whether the sending Host A is fixed or mobile.

Should the urban Internet infrastructure (or wireless access to it) fail, one must maintain also a VLS totally supported in the Urban Grid. A considerable amount of research has gone into VLS design. The goal is to minimize registration overhead while at the same time minimizing the index search – two conflicting requirements. A possible solution is based on a hierarchal design. At the lowest level, there is a unique (mobile) server – e.g., a CalTran truck that roams in a cell (1 km × 1 km), periodically advertising its coordinates. Vehicles register locally with the truck [43]. At the higher level each vehicle has a permanent Home Agent, say. As the vehicle moves from one cell to another in the Urban Grid, it must update the pointer in its home agent. During normal operations, when the infrastructure is up, home agent updating is done through the Internet overlay, at the same time when the OLS updating takes place. In disconnected operations, the Location Service is supported by the cellular trucks (Fig. 4).

4.2. Routing in the vehicular grid

In general, the vehicle grid will support many routing options simultaneously, the selection depending on the name/address map scheme. The prominent scheme, especially to remote destinations, will be geo-routing. Yet geo-routing in vehicular grids poses research challenges. The first issue is vulnerability to “dead end” traps. Vehicle grids are full of such traps. Once GPSR falls in a trap, the recovery must be done with time consuming graph planarization followed by “perimeter routing”. One open research issue is to investigate schemes that prevent/recover from traps more efficiently than planarization.

One option is to use landmark assisted geo-routing (GeoLanmar). We can detect if geo-routing leads to a trap by comparing the "Euclidian" distance/direction (i.e., the direction "as the crow flies") with the GeoLanmar advertised distance/direction [45]. If the GeoLanmar distance is substantially higher, there is a good probability of a dead end and the landmark path is chosen. This scheme however carries the overhead of periodic GeoLanmar advertising. An option without planarization nor advertising overhead is Geo-Cross, where planarization is NOT performed before perimeter routing [3]. Rather, loops are allowed to occur, and are efficiently detected a posteriori during the routing process.

The above protocols forward packets efficiently when the underlying network is fully connected. However, the dynamic nature of vehicular network, such as vehicle density, traffic pattern, and radio obstacles can create temporary disconnections and networks partitions. To overcome these problems, disruption and delay tolerant geographic routing solutions such as GeoDTN+Nav must be used [19]. As the name indicates, GeoDTN+Nav is a delay tolerant extension of geographic routing that exploits the on-board navigator. GeoDTN+Nav first determines when the network has become partitioned (this is inferred when geo-routing has switched from greedy to perimeter mode and the packet has travelled an unusually large number of hops). With a partitioned network, packet routing is processed in DTN mode (i.e., carry and forward). Delivery latency is improved in GeoDTN+Nav by using passing vehicles’ Virtual Navigation Information (e.g., intended destination, direction, trace of the route covered so far, future route plan, etc.) to select the most appropriate vehicle for passing the packet to. Moreover, the information obtained by neighbor vehicles carries a certain confidence level. As shown in Fig. 5, the path declared by a public bus has 100% confidence, while the path announced by a private car is treated with 55% confidence. GeoDTN+Nav is an example of hybrid routing algorithm. It can switch from conventional greedy geo-routing to delay tolerant routing and back depending on road density conditions.

The presence of the infrastructure has again an important impact on urban routing. In most cases, one must find a route to the closest Access Point, to connect to the Internet, or to remote vehicles. In a city with advanced wireless infrastructure deployment, it will take only a few hops to reach the nearest AP, in day traffic conditions. Routing over the infrastructure instead of the Vehicular Grid to reach remote vehicles will be the norm.

In the previous sections, we have argued that most vehicular applications/services must be supported by a roadside infrastructure with robust Internet connectivity. Navigator based intelligent transport and content distribution (via CarTorrent, say) are representative examples.

Offering an infrastructure that is instantly available to support vehicles is itself a challenge. Roadside Access Points must be placed in special locations, in clear view of the moving traffic, in contrast to Access Points designed for smart phone access in shopping centers, pedestrian malls, bus/train stations, industrial Campuses etc. Ideal AP installations for vehicles are traffic lights, light poles, overpasses and other public structures. Traffic lights in particular are perfectly positioned to act as traffic routers.
They already form a traffic grid; are located where traffic is most intense, and; are equipped with power and directly maintained by local municipalities.

Not all the roadside AP’s will have direct access to the Internet, due to cost and physical limitations. Some AP’s have not even electric power and must be supplied by solar. This calls for wireless mesh connectivity at the periphery of the fixed infrastructure. In the sequel we introduce MobiMESH, a mesh architecture that interconnects vehicular ad hoc segments with the infrastructure to support extended vehicular applications and services [2]. The MobiMESH architecture consists of three main building blocks as shown in Fig. 6.

1. **Mesh Backbone**: a network of MobiMESH routers providing routing, mobility management and Internet connection.
2. **Ad hoc network extension**, exporting MobiMESH functionalities to mobile nodes.
3. **Access network** featuring standard WiFi connectivity.

The Mesh Backbone and the ad hoc extension operate in ad hoc mode, routing is based on OLSR [20] with modifications to account for multiple radios and link quality metrics. The access network operates in the infrastructure mode and can be accessed by standard clients with no specific software as if it were an off the shelf WLAN.
The routers are the main building block of MobiMESH. They are responsible for creating the broadband backhaul, and for providing access to mobile clients. They are equipped with 2 to 4 radio interfaces to be used interchangeably as backbone or local access interfaces. Any radio interface can be tuned to any available channel in the frequency bands 2.4 GHz, 5.7 GHz and 5.9 GHz (via DSRC). In access routers, one of the interfaces is dedicated to WiFi clients.

4.3. Emergency routing when the grid has failed

Consider the aftermath of a major earthquake in Southern California or the destruction of a Katrina class hurricane in the Gulf Coast. Power lines are down and electric service is spotty and intermittent. The Communications infrastructure (wired and wireless) is either out of power or heavily congested. Traffic lights do not work and the major traffic arteries have sustained structural damage. Emergency medical services cannot be contacted or otherwise cannot reach the sites with the most essential means. Fire breaks out in several places because of damaged gas pipelines, and firefighters cannot reach those spots quickly in order to repair damage and prevent further loss of life.

This is when the VANET steps in as the ultimate Emergency Network. The VANET continues to run in spite of natural disasters, and is able to configure itself and route information using the available resources. Queries are sent out to the VANET by Emergency Management Agencies to gather information about vehicle traffic in the city and also about road conditions. The queries are high priority and are instantly relayed by intermediate vehicles to the entire city. Within seconds, detailed “congestion maps” are created and optimal routes are planned and forwarded to the emergency vehicles already on route. If necessary, evacuation routes are configured to lead unfortunate motorists to safety.

The vehicular network also allows paramedics to inform hospital staff in advance about injuries and vital signs of incoming victims. At the same time, hospital bed availability is relayed to ambulances to facilitate patient distribution to facilities. Instant, efficient communications are key to a prompt recovery. The VANET plays a pivotal role in the recovery as it is always “on the ready”. VANET routing protocols are designed to work well even without infrastructure. No extra protocols must be bootstrapped for P2P operation.

5. Historical perspectives and future trends

At the beginning, in the early 70’s, the Internet (then called ARPANET) and the wireless packet switching counter-part (the Packet Radio NET or PRNET) started on different journeys, with different missions. ARPANET was targeting better trunk line efficiency and resource sharing of expensive computer resources. Over the years, this was achieved by building an increasingly powerful backbone with gigabit/s fiber links and terabit/s routers and ground trunks. PRNET was targeting mobility, portability and lightweight. The major customer of the PRNET concept was, and still is, the Military. Over the years, the tactical MANETs have become the building block of the Network Centric Battlefield, with increasingly sophisticated radios (SDR and Cog Radios); more intelligent, self configuring, robust network architectures, and; total independence from the wired Internet.

In the mid 80’s the cellular phone was invented and in the late 90’s the Wireless LAN was established. These have been the most important commercial successes of wireless communications since the PRNET. However, architecturally, they are still a wireless, one hop “extension” of the wired infrastructure. Peer to peer networking with smart phones and WiFi laptops has been limited to research lab experiments.

During the past five years, the vehicular network (VANET) has emerged as the true commercial heir of the PRNET. Peer to peer, ad hoc communications are made possible in VANETs by the ample supply of power on board of the vehicle. Moreover, VANET applications like content distribution and crash avoidance require P2P exchanges.

However, as we have amply illustrated in this paper, the VANET is not a stand alone network. Rather, it is intimately dependent on the wired infrastructure for its applications (e.g. content) and its services (e.g. security, mobility management, routing among remote vehicles, etc.). Let us take another look at the applications we have studied in this paper:

(a) navigation safety applications (e.g., crash prevention, road problem warnings, conditions of the driver, etc.);
(b) navigation efficiency (e.g., Intelligent Transport Sys, road congestion avoidance, personalized navigator, pollution mitigation);
(c) entertainment: download multimedia; multiuser games (some also educational);
(d) vehicle monitoring (OBD, low Carbon emission, green hybrid vehicle management);
(e) urban sensing (congestion, pollution, forensic); also participatory sensing, involving P2P exchanges with other vehicles (e.g., monitoring the “density” of road segments);
(f) social networking; proximity and correlated motion driven acquisition of friends/peers; blog uploading (e.g., places we have seen recently);
(g) emergency (e.g., evacuation for Katrina type disasters);

The diligent reader will note that all the above applications involve a P2P component side by side with an infrastructure support component.

This interaction between wireless P2P communications and wired Internet access in VANETs and more generally in all mobile networked platforms will be the most important trend to watch for in the Future Wireless Internet. This trend will be in part influenced by the new applications that will emerge in future years (a short list is presented above; but, as Internet History has taught us, applications are the most difficult thing to predict in this field). P2P penetration will also depend on our ability to solve the privacy and security risks intrinsic in P2P sharing. There are two more important factors that will influence this trend:
(1) Spectrum shortage: according to the Obama Administration we need 500 MHz more to satisfy the emerging wireless data needs (experts say we actually need close to 800 MHz more bandwidth).

(2) Emergence of Cognitive Radios and FCC white spectrum ruling: new technology and regulations will give an enormous impulse to the creation of new wireless protocols and architectures (some will be P2P) to reclaim the unused bandwidth.

These two factors go hand in hand. Cognitive radios help "create" more available bandwidth. Some of the solutions are P2P. For example, in a future urban scenario we will have LTE (Long Term Evolution) cellular service, with very high bandwidth and broad range, and; WiFi AP and Femtocells for high bandwidth within shorter range. Possibly, WiMAX will also be present. In addition, plenty of TV bandwidth will be available to "scavenge" when not used. The Cognitive radios will allow the user to be "best connected" all the time. For instance, in a shopping mall or in an airport lounge LTE will become congested. The user cognitive radio will disconnect from LTE and will rehome to APs or femtocells. If necessary, it will invade unused TV spectrum. Eventually, when all these resources will become exhausted, Cognitive radios will seek better spectrum reuse via P2P, ad hoc, multihop solutions. Instead of scavenging unused licensed spectrum, they will reuse unlicensed spectrum more efficiently, exploiting the presence of multiple channels (say, 11 channels in IEEE802.11). Recent studies on Cognitive Ad Hoc Networking in urban environments have shown that the efficient "reuse" of residential WiFi spectrum in ad hoc mesh networks (URBAN X) and ad hoc multicast (CoCAST) has led to significant spectrum savings [15,16].

There is another theme that will contribute to shape the wireless mobile trend: privacy (of location and motion patterns) and security (mainly, confidentiality and protection from DDoS attacks). We did not dwell on the privacy/security topic in this paper because we chose to focus on conventional network layer aspects (say single hop to Internet vs P2P multihop). However, two trends are clear. The need of a Certificate Authority (CA) will require efficient connection to Internet Servers. At the same time, to handle protection from bogus attacks in situations when they are disconnected from the Internet, or it is simply too time consuming to consult the Internet CA, the mobile users must organize in P2P communities and use majority rules and/or elect proxy mobile CAs to resolve security issues [13].

In summary, the journey of ad hoc networks started on the pure ad hoc, no infrastructure path in the early 70's. Vehicular networks have reaffirmed the importance of the infrastructure. But, emerging need for better spectrum efficiency and stronger security make P2P ad hoc networking very desirable, with the help of advanced Cognitive Radio Technology.

6. Conclusions

This paper has addressed advanced wireless network architectures and applications focused on vehicle networks. The legacy for these designs comes from the early ARPA studies of Packet Radio networks in the 1970's when many of the core multi-access technologies were developed, namely, ALOHA, CSMA, and related extensions. That early focus was on analytical models as well as testbed implementations and experiments that led to ALOHANET, PRNET, etc. Among the lessons learned from those early studies is that multi-hop wireless networks are extremely inefficient due to collisions, lost packets, long paths, ephemeral connections, hidden terminals, etc. Moreover, the technology of the 1970's produced large, heavy and power-hungry radios. In the 1990's, DARPA revisited the packet radio design issues since, by then, the technology permitted smaller, lighter, and power-efficient radios. This surge of interest, around the time when the mobile cellular revolution was taking place, has led to significantly renewed research and development in MANET systems. The growth and development of peer-to-peer systems provided the technology and experience to extend P2P to the car-to-car networks described in this paper. Today the P2P paradigm is playing an increasingly important role in efficient spectrum reuse and in robust operation during intermittent connectivity episodes.

In this paper, we have identified the need for routing and topology maintenance as being critical for these vehicle networks. The introduction of advanced digital and networking technology into automobiles is expanding dramatically and is in the process of enabling the kinds of applications we have discussed; this is a significant direction of expansion for the automobile industry. A representative example is the use of parked automobiles as stationary, yet ephemeral, content stores and providers. Basically these stationary vehicles have the ability to form a MANET without mobility in which routing is far more easily enabled, although updates to network routing are required when the parked automobiles leave their parking spaces. Another extension is to exploit the use of public buses as reliable roving servers on a known route.

On a broader scope, we anticipate that the future of vehicular networks will be intertwined with that of smart phones and wireless sensor networks. For example, vehicular networks will interoperate with smart phones and possibly airborne sensors to map the environment. In the vehicle itself, body area sensor networks will monitor the driver's vital signs and will cooperate with VANET alert mechanisms to avoid accidents in case of driver collapse. Vehicle networks will leverage the broad gamut of services available on modern smart phones to support geo-location aware applications and social networking. Naturally, the type of data that can be fed to a pedestrian is quite different from what can be fed to a driver since one "must not distract the driver"!

In summary, plenty new applications are emerging in the vehicular arena and anyone is betting on one or the other. Unfortunately, Internet history has taught us that it is absolutely impossible (and useless) to predict the Killer Application. One thing is sure: vehicles will need to use the Internet infrastructure more and more for advanced services. And, the infrastructure must be enhanced and expanded to support mobile users more efficiently.

References


Mario Gerla is a Professor in the Computer Science at UCLA. He holds an Engineering degree from Politecnico di Milano, Italy and the Ph.D. degree from UCLA. He became IEEE Fellow in 2002. At UCLA, he was part of the team that developed the early ARPA NET protocols under the guidance of Prof. Leonard Kleinrock. He joined the UCLA Faculty in 1976. At UCLA he has designed network protocols including ad hoc wireless clustering, multicast (ODMRP and CODECast) and Internet transport (TCP Westwood). He has lead two advanced wireless network projects under ARMY and IBM funding. His team is developing a Vehicular Testbed for safe navigation, urban sensing and intelligent transport. A parallel research activity explores personal communications for cooperative, networked medical monitoring (see www.cs.ucla.edu/NRL for recent publications).

Leonard Kleinrock is considered a father of the Internet, having developed the mathematical theory of packet networks, the technology underpinning the Internet, while a graduate student at MIT. This was in the period 1960–1962, nearly a decade before the birth of the Internet which occurred in his laboratory when his Host computer at UCLA became the first node of the Internet in September 1969. He wrote the first paper and published the first book on the subject; he also directed the transmission of the first message ever to pass over the Internet.

Dr. Kleinrock received his Ph.D. from MIT in 1963, serves as a Distinguished Professor of Computer Science at the University of California, Los Angeles (UCLA), and has founded a number of Internet companies. He has published over 250 papers and authored six books on a wide array of subjects including packet switching networks, packet radio networks, local area networks, broadband networks, gigabit networks, nomadic computing, peer-to-peer networks, and intelligent software agents.

Dr. Kleinrock is a member of the National Academy of Engineering, a member of the American Academy of Arts and Sciences, an IEEE fellow, an ACM fellow, an INFORMS Fellow, an IEE Fellow, a Guggenheim fellow and a founding member of the Computer Science and Telecommunications Board of the National Research Council. Among his many honors, he is the recipient of the L.M. Ericsson Prize, the NAE Charles Stark Draper Prize, the Marconi International Fellowship Award, the Okawa Prize, the Dan David Prize and was further recognized when he received the 2007 National Medal of Science, the highest honor for achievement in science bestowed by the President of the United States.