Integrating Advanced Mobility Services into the Future Internet Architecture*

Shreyasee Mukherjee, Akash Baid†, Dipankar Raychaudhuri
WINLAB, Rutgers University, {*shreya, baid, ray*}@*winlab.rutgers.edu*

***Abstract*—This paper discusses the design challenges associated with supporting advanced mobility services in the future Internet. The recent transition of the Internet from the fixed host-server model to one in which mobile platforms are the norm motivates a next-generation protocol architecture which provides integrated and efficient support for advanced mobility services. Key wireless access and mobility usage scenarios are identified including host mobility, multihoming, vehicular access and context addressability, and key protocol support requirements are identified in each case. The MobilityFirst (MF) architecture being developed under the National Science Foundation's future Internet Architecture (FIA) program is proposed as a possible realization that meets the identified requirements. MF protocol specifics are given for each wireless/mobile use case, along with sample evaluation results demonstrating achievable performance benefits.**

I. Introduction

The Internet is fast approaching an inflection point with wireless/mobile devices overtaking wired PCs as the primary end-user device, i.e. mobility as the norm. Since the iPhone was introduced in 2007, worldwide smartphone usage continues to grow at an exponential rate. The Cisco VNI Global Mobile Data Traffic Forecast 2013 [1] predicts that traffic from smartphones alone will account for about 7.5 Exabytes/month in 2017, a factor of 10x relative to 2013. The Cisco report also forecasts that "by 2016, wired devices will account for only 39% of all IP traffic". This fundamental shift in Internet usage presents a unique and timely opportunity to consider the requirements and wireless access challenges from the ground-up and provide protocol solutions to address them.

The current TCP/IP based Internet protocol framework has several limitations when applied to wireless access scenarios with mobile endpoints. IP address assignment and management via protocols such as DHCP and DNS are relatively static while TCP assumes the existence of a contemporaneous end-to-end path. In addition, IP addresses serve the dual roles of end-point identifier and routable network locator, making it difficult to deal with many aspects of dynamic mobility such as disconnection or multihoming. Incremental network and transport layer solutions (e.g. Mobile IP [2], TCP Multipath [3]) aim to tackle only part of the problem, whereas clean slate naming conventions like the Host Identity Protocol (HIP) [4] concentrate primarily on the name-address separation issue. As mobile networks expand to encompass everything around us in a "connected world", 3GPP evolution 5G access aims to provide improved last mile connectivity. It is anticipated that the future 5G access standard will support gigabit bandwidth and millisecond latencies to meet the requirements of the diverse set of services, application and users [5]–[7]. However networking solutions for cellular mobile data service continue to involve both 3GPP and IP protocols with all the limitations of multiple protocol architectures and associated gateway processing.

In this paper we discuss the wireless access challenges and mobility service requirements and propose an integrated network architecture ("MobilityFirst") which meets these needs. Section II and III enumerate the key mobility requirements from an Internet architecture point of view. In Section IV, the MobilityFirst protocol now under development under the National Science Foundation's Future Internet Architecture (FIA) program [8], is introduced as a possible solution that addresses each of the requirements. This is followed by a discussion of several specific wireless access use-case scenarios such as device mobility, multihoming, vehicular access and mobile content delivery along with sample protocol evaluation results.

II. Cellular-Internet Convergence

As Internet-connected mobile devices will soon outnumber fixed PCs, a convergence of business models and technical standards associated with cellular networks and the Internet may be expected over the next decade. This process has already started, with cellular standards embracing the concept of "flat" IP-based networks without centralized gateways. In 4G/LTE, the cellular access network architecture has been significantly flattened with only a single specialized MME (mobility management entity) in the control path and SGW (service gateway) in the data path, and with commodity routers everywhere else in the network. We predict that this trend will continue with the evolution of 5G radio access [5]–[7]. In our view, the next logical step in this direction is a completely flat mobile network architecture with native support for basic services such as authentication, dynamic association and handover, inter-network roaming, and disconnection tolerance. As shown in Fig. 1, in the integrated "mobile Internet" architecture, it will be possible to "plug in" multiple wireless access technologies such as 4G, 5G or Wi-Fi without requiring gateways. Such a uniform protocol solution across wired and wireless network technologies will eventually lead to convergence of cellular and Internet standards, in view of the fact that both industries are serving the same mobile

*Research supported by NSF Future Internet Architecture (FIA) grants CNS-1040735 and CNS-1345295

†currently with Tarana Wireless, Berkeley, CA

978-1-4244-8953-4/11/$26.00 © 2015 IEEE

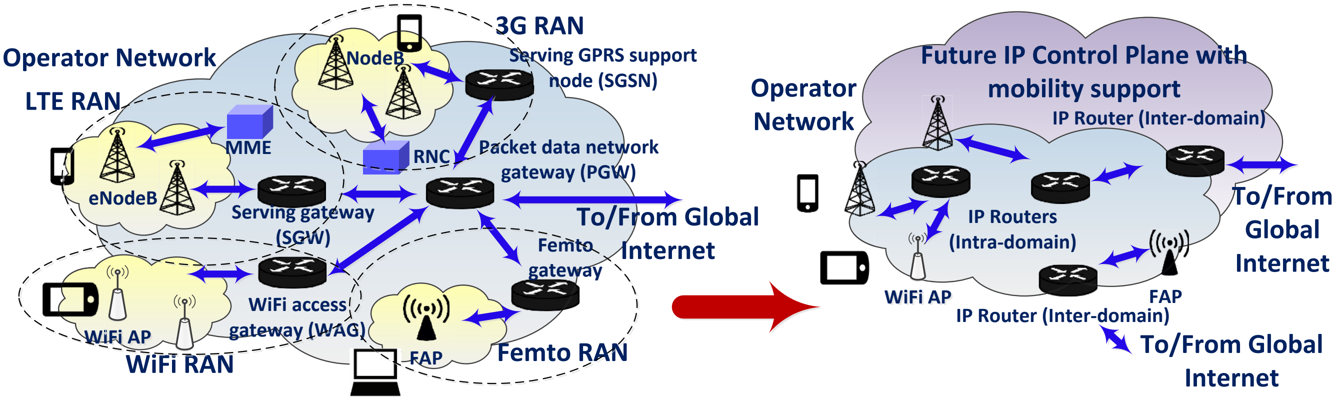

Fig. 1: Cellular-Internet convergence for the future Internet

end-users. Beyond mobile data, any new protocol architecture should also support the requirements of emerging machine-to-machine (M2M) communications between embedded sensors, vehicular networks, and Internet-of-Things devices, which are expected to grow significantly over the next decade to an estimated 1.5 billion devices by 2017 [1].

We note that a unified mobile Internet architecture is useful to both cellular network operators seeking to improve performance, as well as to more general Internet service providers (ISPs) aiming to introduce mobility services across heterogeneous access networks. For example, an ISP that currently offers standard Internet access service could expand the offering to include seamless mobility across multiple wireless networks such as Wi-Fi hot-spots using standard network elements (router, basestation, access point) without the need for a specialized control framework. This type of heterogeneous wireless access service is sometimes referred to as "open wireless networks" [9] in which loosely coupled access networks use a common protocol to support basic mobility needs such as authentication, handover and inter-network roaming. Cellular providers incorporating Wi-Fi hot-spots and 3G/4G small cells to supplement their existing macro-cellular deployments could also use the same flat future IP protocol to provide mobility services across these heterogeneous networks without the need of any specialized network equipment, as shown in Fig. 1.

III. Wireless Access Requirements

In this section, we analyze specific wireless access and mobility service requirements and identify the corresponding protocol implications for their support.

A. Host and Network Mobility

The primary characteristic of mobile nodes is that their points of attachment to the Internet can change easily and rapidly. The need for supporting mobility arises when an individual node or a group of nodes, for example a bus/train/plane network, moves and reconnects to the Internet. Previous studies on opportunistic Wi-Fi through vehicular nodes have shown that mobile nodes suffer frequent disconnections (a mean disconnection period of 75 seconds). In addition, nodes change their IP addresses every time they associate with a new access point (median connectivity period is only 13 seconds for vehicular mobility in an urban scenario) [10]. A cellular network provider performs handover between its basestations transparent to the user, enabling them to hold on to their static IP address assigned by the network provider. However data is routed through a gateway which reroutes it to the current basestation the client is connected to. In this regard, Mobile IP tries to achieve the same with the use of fixed mobility anchors [2]. However, the concept of having a fixed "home network" with infrequent network transitions, is changing. Given host names and their actual locations are increasingly becoming uncorrelated, a fundamental requirement for mobility support is to separate the two and identify hosts only via a permanent name. This functional requirement can be translated to the following protocol design requirements:

1) Disambiguation of the dual-roles of an IP address as both an identifier and a locator into two different primitives - a permanent name and a network-specific temporary locator.
2) Dynamic binding of names to network addresses/locators.
3) Support for weak connectivity and disconnection in wireless environments.

B. Varying Wireless Link Quality and Disconnection

Achievable bit rates in both Wi-Fi and 4G systems, can show large variations within a fraction of a second. Temporary disconnections due to mobility and/or insufficient signal strength is also common. While these variations are usually handled at the PHY and MAC layers, they invalidate some implicit assumptions in the control algorithms used in the Internet. For example, it has been long known that TCP congestion control treats wireless link errors as congestion losses and performs poorly in high variation and multi-hop wireless channels [11]. Given the last mile connectivity is increasingly becoming wireless, such link quality variations need to be natively supported at different layers of the Internet architecture. This leads to the following requirements:

1) Link quality awareness at both the intra-domain and inter-domain routing layers to enable robust packet delivery strategies.
2) Disconnection-tolerant routing with support of forwarding in-transit packets to new points of attachments.
3) Reliable transport protocol capable of temporary storage and asynchronous delivery of data in the presence of poor link quality and/or disconnection.

C. Accessing Multiple Networks

A typical wireless device in an urban area today might see 3-5 cellular networks and 10-20 Wi-Fi access points, but accesses only one of these due to both technical and business model constraints. Current techniques supporting simultaneous use of multiple interfaces rely on enhancements to the underlying end-to-end transport layer (see [3] and references therein). Specifically, these mechanisms require a multihomed end-point to inform the sender about its multiple interfaces prior to the commencement of data-flow, and a data-striping algorithm on the sender stack that adapts the packet rate of each interface. This results in rigidity in two key aspects: (i) There is no mechanism by which users can specify under what conditions, and in what manner the interfaces are to be used; (ii) Since all decision logic is implemented only at the end-nodes, in-network routers cannot adapt or buffer the flows in accordance with wireless channel quality variations. Thus efficient support for host multihoming induces the following key requirements:

1) Support for binding a single name to multiple addresses and interfaces.
2) A routing plane capable of modifying the data-striping and storing decisions in accordance with the link quality at each interface.
3) Service semantics to support interface selection and utilization (e.g. "send to all interfaces", "send to higher-throughput interface", "send only to Wi-Fi", etc.).

D. Ad hoc Networks

Wireless ad hoc networks are important for infrastructure-less vehicle-to-vehicle (V2V) and sensor network scenarios, last-mile connectivity and applications such as photo/video sharing, local social networking, and multi-player gaming. One view of Internet design is that ad hoc networks are just a type of edge network; as long as they are connected to the Internet via a boundary IP router, the protocols used within the ad hoc network can be ignored. However, the ubiquity of non-specialized devices requiring support for ad hoc networking (e.g. phones, tablets, laptops, vehicular infotainment systems, etc.) forms a strong argument for an integrated design that avoids boundary translation solution. Integration of such networks within the framework of a future Internet design results in the following distinct requirements:

1) Critical network services such as authentication and dynamic binding of names to addresses should be capable of disconnected-mode of operation.
2) Routing and transport protocols should be robust to opportunistic association and changing network topologies.

E. Content and Context Addressability

Along with the shift from fixed to mobile nodes, the Internet is increasingly becoming content and context-driven. In contrast to communicating with a fixed destination, information-centric networking refers to the retrieval of named content, which could potentially be cached at multiple end-hosts. According to the Sandvine global Internet phenomena report 2013, Netflix and YouTube account for more than 50% of downstream Internet traffic in North America [12] and the demand is only predicted to increase. Current Internet architecture deploys content delivery networks (CDNs) or peer-to-peer (P2P) systems to support content-delivery, but such application layer over-lays suffer from efficiency and cost issues. Context-services on the other hand use external conditions, including time, location, and network attachment, to deliver information to/from end-hosts [13]. With the advent of Internet of Things (IoT), providing context-aware computing on large volumes of sensor data becomes crucial [14]. In these use-cases, it is necessary to use the content or context as a first-class primitive in packet transmission, i.e. it should be as easy to use content/context semantics like "fetch content X from nearest source" or "send to all nodes at location Y", as the traditional end-to-end semantic of "send to address Z". Supporting these use-cases in mobile scenarios lead to the following requirements:

1) The architecture should enable dynamic identification of endpoints based on content/context attributes.
2) Since the context attributes of mobile nodes can change rapidly, there is a requirement for fast mechanisms that capture the context and make it available as a packet delivery primitive.

F. Spectrum Access Coordination

Finally, a critical challenge that differentiates wireless networks from wired networks, but which is common across all forms of wireless networks - cellular LTE, Wi-Fi, white-space networks, etc., is the need for devices to coordinate their use of spectrum. These coordination schemes, whether centralized, distributed, or a hybrid, are typically implemented through overlay channels. For example, the IETF PAWS protocol for accessing white space database uses an HTTPS overlay [15], and the X2 interface between LTE base stations uses SCTP over IP [16]. However supporting these wireless control plane functions at the scale of thousands of devices/km requires an integrated approach satisfying the following requirements:

1) Support for a low-latency control plane that is unaffected by data plane congestion.
2) Dynamic multicast of control messages, based on geographic location and radio-range of the sender, to enable efficient distributed coordination schemes.

G. Other Requirements

Although we do not focus on security aspects in this paper, the requirements of location privacy, strong authentication of ownership, mechanisms against mobility spoofing attacks and fast authentication mechanisms must also be taken into account for a mobile-centric future Internet architecture.

In the following section, we introduce MobilityFirst as a clean slate solution, that is built from ground up considering the above mentioned requirements.

IV. MobilityFirst Architecture

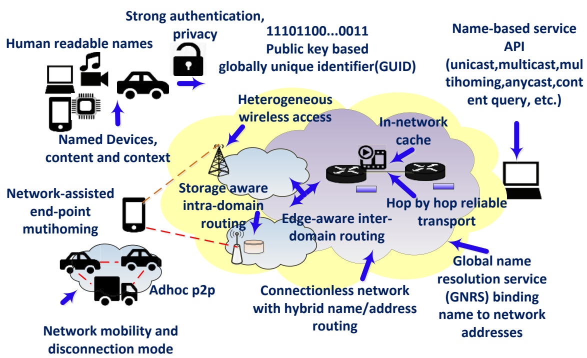

Fig. 2: MobilityFirst architecture overview

The MobilityFirst architecture [8] is based on the idea of separating "names" of end-users or other network-connected objects, and their routable addresses or locators. As shown in Fig. 2, the name-based service layer uses flat globally unique identifiers (GUIDs). Every network attached object (a device, a piece of content, a vehicle, a collection of devices, etc.) is assigned a 160 bit GUID by name certification services (NCS). This GUID could be derived from a cryptographic hash of an associated public key for self-authentication purposes. The GUID being directly derived from the public key gives it a self-certifying property; authenticating a node does not require an external authority [4]. The dynamic mapping of GUIDs to NAs is made possible through a logically centralized, but physically distributed infrastructure called the global name resolution service (GNRS). GNRS is implemented in MobilityFirst as a distributed DHT-based direct mapping (DMap) infrastructure, that can achieve round-trip update and query latencies under 100 ms [17]. Optimization based on location and popularity have been shown to further reduce GNRS median latency to the range of 10 ms [18].

For intra-domain traffic forwarding, MobilityFirst employs a storage aware and delay-tolerant link-state routing. Generalized storage aware routing (GSTAR) [19] provides integrated storage at the routers and link-state flooding throughout the network. The decision to store or forward at every hop is based on both short-term and long-term path quality metrics. GSTAR is also disruption tolerant. In particular, each router maintains two types of topology information: (i) An intra-partition graph is formed by collecting flooded link state advertisements which carry fine-grained, time-sensitive information about the intra-network links; (ii) A DTN graph is maintained via epidemically disseminated link-state advertisements which carry connection probabilities between all nodes in the network.

For inter-domain traffic, MobilityFirst exposes a global network graph with some visibility of edge network properties, as a general solution for mobility-related requirements. Edge-aware inter-domain routing (EIR) combines three key techniques to enhance network visibility. Autonomous systems (ASes) have the flexibility to abstract network entities and their associated properties to aggregation nodes, called aNodes, and connectivities to its neighbors to virtual links, called vLinks. Networks announce their internal state using a network state packet (nSP) which contains the internal network graph in terms of aNodes and vLinks that the network operator wishes to expose. nSPs are flooded across the network using telescopic updates designed to keep the total routing overhead within limits. As a side effect of telescopic route dissemination, nSP updates that a network receives from a distant network could be obsolete and hence result in routing failure. To address this, EIR uses the concept of "late binding" by which routers can temporarily store packets and query the GNRS to rebind names (GUIDs) to addresses (NAs). Further details on inter-domain routing design can be found in [20].

V. Wireless Access Use Cases

In this section, we further discuss design of the MobilityFirst (MF) protocol stack, through a set of wireless use-case scenarios. The MF protocol key components such as GNRS, storage-aware and delay tolerant routing and edge-aware inter-domain routing are highlighted along with sample evaluation results. Table I summarizes how these components achieve the set of requirements identified in each of the use-cases.

A. Host Mobility

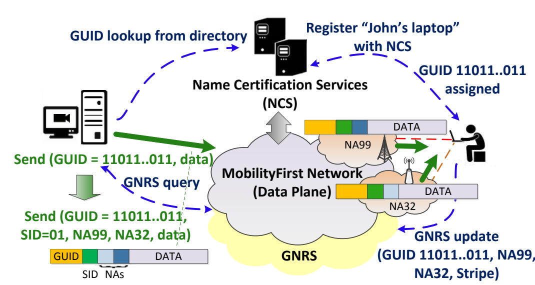

Fig. 3: Example showing message delivery to "John's laptop" that is dual-homed using MobilityFirst

MobilityFirst is based on the idea of separating "human-readable names" of end-users and their routable addresses, with the mapping of flat GUIDs to its corresponding network attachment points (NA) maintained in a distributed fashion at the global name resolution service (GNRS). Any router within the network can update the GNRS with new mappings and query for up-to-date GUID to NA translation.

Consider the example scenario shown in Fig. 3: When "John's laptop" connects to the Internet, it is assigned a GUID by the name certification services (NCS). When another host wishes to send data to "John's laptop", it obtains the corresponding GUID from the NCS. The GUID is then resolved through a GNRS lookup at the edge router to the set of current NAs. The GUID assigned to the host remains constant for the lifetime of the device. As the device moves, its up-to-date network location's mapping changes in the GNRS. In addition, generalized storage aware routing (GSTAR) [19] provides integrated storage at the routers. The decision to store or forward at every hop is based on both short-term and long-term path quality metrics. In the disruption tolerant (DTN) mode of GSTAR, each router calculates the connection probability of all other nodes in its network, and epidemically forwards data towards the destination. For example, if John's laptop, temporary disconnects from the Wi-Fi, data packets would be temporarily stored in NA_{32}, and would be forwarded once the Wi-Fi comes back up. Results [21] indicate that by intelligently utilizing in-network storage, GSTAR outperforms traditional and storage-augmented link-state protocols in both wired and wireless network environments.

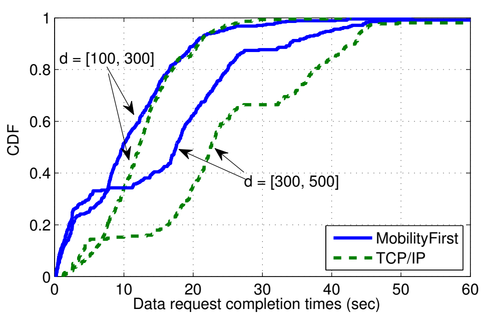

Fig. 4: Cumulative distribution of the data request completion times for TCP/IP and MobilityFirst in a web-browsing emulation scenario

1) Evaluation scenario: The simulation topology consists of a single mobile client with a single 802.11 radio moving along a straight roadway, with access points deployed along the road at random inter-AP distances, *d*. A remote data server is assumed to be connected to the access points through the back-end wired network and the client requests different sized content packets (uniformly distributed between 10 KB to 5 MB, roughly representing HTTP packet lengths in use today [22]) from the server. Fig. 4 compares the performance of the MobilityFirst architecture with the current TCP/IP based Internet access, through a detailed NS-3 simulation. For the TCP/IP comparison, we do not assume a managed Mobile IP implementation since the setup being evaluated here is that of opportunistic connections through open 802.11 APs which are deployed and managed by different entities. As in practice, the moving client in our simulation gets a new IP address via DHCP upon connecting to each AP in the TCP/IP case. Further, in order to make a fair comparison, we assume a "smart" application layer over TCP/IP that resumes transmissions from its point of last connection instead of re-requesting entire transfers. Correspondingly, for the MobilityFirst case, every time the client switches AP, GNRS update/query events are generated which take between 30 to 170 milliseconds, as per the evaluation in [17]. The speed of the car is kept at a constant 50 miles/hr, while two settings of inter-AP distance *d* is used: uniformly distributed between 100-300 meters and same with 300-500 meters. The cumulative distribution function of the request completion times are shown in Fig. 4.

The results show a significant reduction in the transfer times for both values of *d* - median gains of around 4 seconds or 30% in the [100, 300] case and 5 seconds or 22% in the [300, 500] case. Another way to interpret these results are to look at the percentage of completed requests within a given time-frame. On this scale, there is almost a 2x gain in both cases, for example, when measuring the fraction completed at 5 seconds.

B. Multihoming

Continuing with the same example from Fig. 3, after link-level association, the dual-homed device named "John's laptop" updates the global name resolution service (GNRS) with the set of network addresses corresponding to its current points of attachment . Preference policies (for e.g. best path, lowest cost path, striping over both paths, etc.) can also be expressed through this update message, as shown in the figure. When sending data to John, the GUID is resolved through a GNRS lookup to the set of current NAs, in this case NA_{99} and NA_{32} and an optional service identifier (SID) corresponding to host-specific preference policies. The packet header actually sent out into the network then consists of a destination GUID, an optional SID and both the network addresses for the network routing protocol to decide on the forwarding path. Availability of multiple paths is enabled through link-state routing utilizing GSTAR and EIR, as explained earlier in Sec. IV. If the user's policy is to stripe data across all the available interfaces, MobilityFirst utilizes a robust hop-by-hop backpressure mechanism to estimate the ratio of data to be sent across each, as explained in detail in [23].

1) Evaluation scenario: The topology remains same as the evaluation scenario of host mobility in subsection V-A, except now we assume the vehicle to be also connected to an LTE basestation, which provides it with a continuous coverage but lower achievable data rate. *d* is uniformly distributed between 300-500 m to simulate frequent disconnections through Wi-Fi. The mobile client downloads a large file from the server, while moving at a speed of 10 meters/sec (~22 mph) and we measure the raw aggregate throughput that could be achieved in such a scenario. Since baseline TCP does not support striping of data across multiple interfaces simultaneously, we focus on the advantage of using multiple interfaces in comparison to a single interface in MobilityFirst.

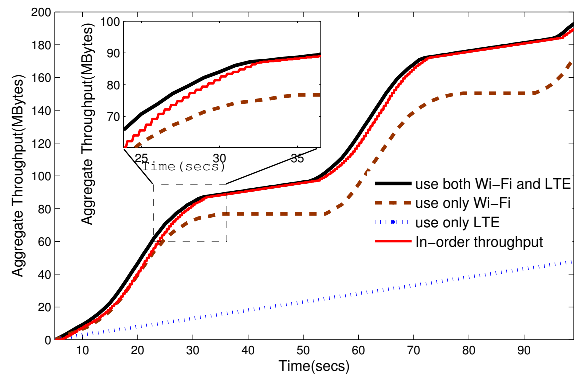

Fig. 5: Aggregate throughput for a multihomed mobile client with a Wi-Fi and an LTE interface

As shown in Fig. 5, the in-network data-striping (a detailed description of the bandwidth estimation and striping algorithm, is given in [23]) fully utilizes the Wi-Fi interface whenever it becomes available. This is indicated by the multihoming throughput being close to the sum of the raw throughput achievable through each of the individual interfaces. We also consider the case, where the application demands in-order delivery of data, indicated by the red curve. As shown in the zoomed cutout in Fig. 5, the application throughput advances in small jumps, as data arrives out-of-order across both the interfaces. However, the in-order application throughput closely follows the raw throughput trend, denoted in black.

C. Vehicular Access

Vehicular access forms an important use-case for network protocol design, from infrastructure-less ad hoc vehicle-to-vehicle communication to a more infrastructure-oriented vehicle to road-side unit (RSU) communication. Increased in-car times have given rise to new infotainment and location-aware services and targeted advertisements for vehicles, as well as crowd-sourced real-time traffic, safety and vehicular sensor data applications.

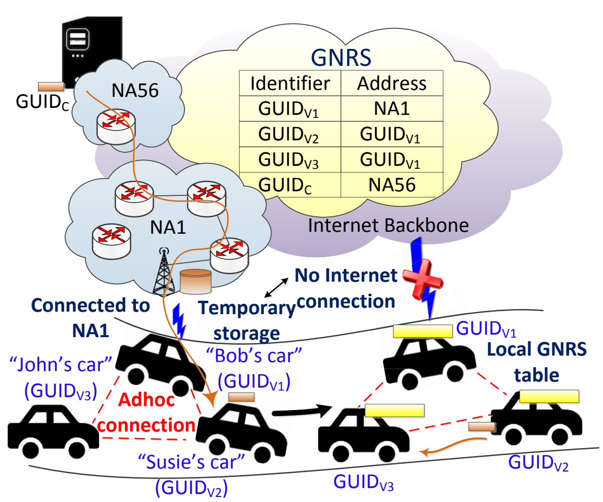

Fig. 6: Dynamic ad hoc network formation and disconnected mode of operation in a vehicular environment

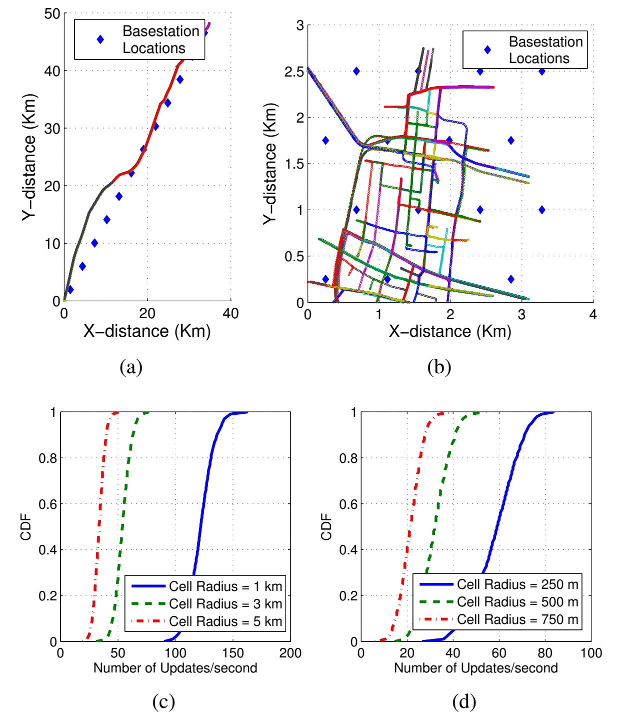

Fig. 7: PARAMICS Results:(a) and (b) Movement trace of 100 randomly selected vehicles from NJTPK and JC traces respectively, along with an instance of basestation placement (cell radius: 3 km and 500 m), (c) and (d) Cumulative Distribution Function of the number of updates/sec across all basestations from NJTPK and JC traces respectively

To understand how MobilityFirst works for V2V applications, consider the scenario depicted in Fig. 6. Bob, Susie and John are traveling along a highway in their cars which are addressable via GUIDs - $GUID_{V1}$, $GUID_{V2}$, and $GUID_{V3}$ respectively. Only Bob's car is connected to the Internet through the cellular network (NA_1). However V_2 and V_3 can form an ad hoc network with V_1 through MobilityFirst. As shown in the figure, MobilityFirst allows GUID-to-GUID delegation and iterative queries in the GNRS. In this scenario, if vehicles V_2 and V_3 can communicate with V_1 using any local area protocol, they can sustain Internet connectivity through V_1. A remote server (say $GUID_C$ wishing to send packets to $GUID_{V2}$ queries the GNRS and gets $GUID_{V1}$ in place of its current network address. It can then query the GNRS again for the mapping of $GUID_{V1}$ to find the current NA on which to forward packets for "Susie's car". Along the way, if she connects to a DSRC road-side unit directly, the GNRS gets updated again, this time directly with the NA of the new network instead of the delegated GUID. Even when there is no Internet connectivity, the vehicles can locally exchange content between themselves via a local mode of name resolution service (LNRS), as shown in the right half of Fig. 6. This is an ongoing work, where we are looking into bootstrapping and ad hoc network formation and maintenance techniques.

1) Evaluation Scenario: One of the key challenges in a vehicular scenario is to dynamically update the GNRS with up-to-date mapping of GUIDs to the routable network address of highly mobile vehicles. We analyze the scalability of GNRS updates with simulation of two real road segments - (i) a 25 mile stretch of the New Jersey Turnpike, referred as the NJTPK trace; (ii) a 3 square mile urban area of Jersey City in New Jersey, referred as JC trace. To realistically analyze vehicular mobility, we use a well-calibrated model of the New Jersey Turnpike and Jersey City that is built in a microscopic simulation tool, PARAlell Microscopic Simulation (PARAMICS) [24]. Calibrations of the model were performed using real-world road traffic measurements data and the average relative error in volumes and travel times are within 8-10% [25]. The NJTPK and JC traces model a total of $\sim$43,000 and $\sim$16,000 vehicles respectively with median speeds of 59 miles/hour and 16 miles/hour during one hour of the evening peak-time. We assume three different values for the cell radius, namely 1, 3, and 5 Km, with basestations placed along the highway in the NJTPK trace. For the urban JC trace on the other hand, we assume regular hexagon cell deployment with cell radius of 250, 500, and 750 meters. Example traces of 100 randomly selected vehicles are shown in Figs. 7(a) and 7(b).

Figs. 7(c) and 7(d) show the cumulative distribution function of the number of updates generated per second in the NJTPK and the JC vehicular trace. Even for a dense vehicular network and worst case assumptions about the frequency of updates, a maximum of around 160 and 80 updates occur per second in the two traces respectively. Since each update would typically correspond to about 40-100 bytes of transmission (the exact size depends on the nature of the network address space in use), this would lead to a traffic overhead of less than 16 KBps. At this scale, our previous analysis shows that with conservative estimates of 10 billion nodes and 100 updates/day/GUID, the worldwide combined update traffic would be $\sim$20 Gb/s, a tiny fraction of the overall Internet traffic of $\sim$50x10^6 Gb/s as of 2010 [1].

D. Mobile Content Delivery

In MobilityFirst, content is a first-class endpoint principal that is represented using GUIDs in the same manner as interfaces, devices or hosts. Content providers create GUIDs for their content and insert an entry into the GNRS denoting its network address and the content GUID. Consider the example shown in Fig. 8: A content publisher such as Netflix translates human-readable name of the content (in this case the URI, "movie_the_godfather") to a 160 bit flat GUID similar to any other network attached object. The binding of the content GUID to the network address of its location is published in the global name resolution service. Content providers also provide APIs through which any subscriber can query their name certification service and obtain the associated content GUID. "Bob's iPad" wishing to stream the movie, sends a get('GUID') primitive to the network. The access router queries the GNRS to resolve the GUID, and appends NA_{12} to the *get* request. This hybrid <GUID:NA> addressing enables forwarding the request along the path towards the content provider network. Once, the packet reaches the ingress router at Netflix, a local name resolution is performed, which maps the content GUID to the GUID of the host ("data server 0") through the local name resolution service. Allowing this two-level name-resolution allows smooth migration of content locally within data-centers of a content provider, without causing global update overhead.

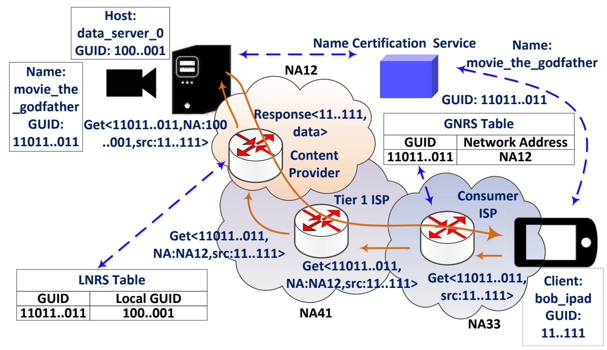

Fig. 8: Overview of content distribution and fetching of content from a client

Two-level hierarchy also reduces the query overhead. Any query for a GUID is sent to the LNRS first, before being forwarded to the GNRS. If the content is available locally, the LNRS server returns a response, else it forwards the query upstream. Optionally ISPs can leverage on this scheme to build a GNRS-assisted caching of popular content [26]. Specifically, an ISP can maintain the recent usage count (RUC) of every content request, based on which it caches "popular contents" at the local AS proxy cache. Note that revenue generating content would need a legal agreement between the two ISPs, allowing the consumer ISP to cache the content locally.

Wireless Access Requirements	MobilityFirst Elements		
	GNRS	GSTAR	EIR
1. Identity/location sep.	✓		
2. Dynamic binding	✓		
3. Link quality awareness		✓	✓
4. Topology robustness		✓	✓
5. Multiple addresses	✓		
6. In-transit decisions		✓	✓
7. Multi-homing policies	✓		
8. Disconnection-mode	✓		
9. Content addressing & delivery	✓	✓	✓
10. Secure access	✓		

TABLE I: Key wireless access and mobility requirements and the MobilityFirst protocol elements that address each requirement

VI. Conclusion

This paper presents an overview of future Internet design considerations motivated by emerging mobility services and wireless access requirements. The MobilityFirst architecture is introduced as a clean-slate solution with key protocol features including name/address separation, robustness with respect to link quality variation and disconnection, multihoming and content/context addressability. While comprehensive coverage of all design goals and protocol features is beyond the scope of this paper, we have identified specific use-case scenarios to validate individual design components. The MobilityFirst protocol has been extensively tested using a combination of simulation, emulation and experimental trials. Ongoing work aims to further validate inter-domain routing aspects, optimized GNRS design and content/context service primitives. Future work also includes trial deployments on the GENI experimental network [27], as well as early-adopter field trials with ISPs and content service providers.

References

[1] C. V. N. Index, "Global mobile data traffic forecast update, 2010-2015," *White Paper, February*, 2011.

[2] C. Perkins, "Ip mobility support for ipv4," 2002.

[3] J. R. Iyengar, P. D. Amer, and R. Stewart, "Concurrent multipath transfer using sctp multihoming over independent end-to-end paths," *Networking, IEEE/ACM Transactions on*, vol. 14, no. 5, pp. 951–964, 2006.

[4] R. Moskowitz and P. Nikander, "Host identity protocol (hip) architecture", rfc 4423," 2006.

[5] Ericsson, "5g radio access-reasearch and vision," *White Paper, June*, 2013.

[6] Huawei, "5g: A technology vision," *White Paper*, 2013.

[7] N. Solutions and Networks, "Looking ahead to 5g," *White Paper, December*, 2013.

[8] D. Raychaudhuri, K. Nagaraja, and A. Venkataramani, "Mobilityfirst: a robust and trustworthy mobility-centric architecture for the future internet," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 16, no. 3, pp. 2–13, 2012.

[9] R. Yates and W. Lehr, "Mobilityfirst, lte and the evolution of mobile networks," in *Dynamic Spectrum Access Networks (DYSPAN), 2012 IEEE International Symposium on*. IEEE, 2012, pp. 180–188.

[10] V. Bychkovsky, B. Hull, A. Miu, H. Balakrishnan, and S. Madden, "A measurement study of vehicular internet access using in situ wi-fi networks," in *Proceedings of the 12th annual international conference on Mobile computing and networking*. ACM, 2006, pp. 50–61.

[11] Z. Fu, P. Zerfos, H. Luo, S. Lu, L. Zhang, and M. Gerla, "The impact of multihop wireless channel on tcp throughput and loss," in *INFOCOM 2003. Twenty-second annual joint conference of the IEEE Computer and Communications. IEEE Societies*, vol. 3. IEEE, 2003, pp. 1744–1753.

[12] Sandvine, "Sandvine global internet phenomena report 2h 2013," 2013, [Online].

[13] G. D. Abowd, A. K. Dey, P. J. Brown, N. Davies, M. Smith, and P. Steggles, "Towards a better understanding of context and context-awareness," in *Handheld and ubiquitous computing*. Springer, 1999, pp. 304–307.

[14] C. Perera, A. Zaslavsky, P. Christen, and D. Georgakopoulos, "Context aware computing for the internet of things: A survey," *Communications Surveys & Tutorials, IEEE*, vol. 16, no. 1, pp. 414–454, 2014.

[15] V. Chen, S. Das, L. Zhu, J. Malyar, and P. McCann, "Protocol to Access White-Space (PAWS) Databases," Internet Draft, IETF Internet-Draft, September 2014. [Online]. Available: https://tools.ietf.org/html/draft-ietf-paws-protocol-19

[16] E. U. T. R. A. Network, "X2 general aspects and principles (release 8)," *TS*, vol. 36, p. V8.

[17] T. Vu, A. Baid, Y. Zhang, T. D. Nguyen, J. Fukuyama, R. P. Martin, and D. Raychaudhuri, "Dmap: A shared hosting scheme for dynamic identifier to locator mappings in the global internet," in *Distributed Computing Systems (ICDCS), 2012 IEEE 32nd International Conference on*. IEEE, 2012, pp. 698–707.

[18] A. Sharma, X. Tie, H. Uppal, A. Venkataramani, D. Westbrook, and A. Yadav, "A global name service for a highly mobile internetwork," in *Proceedings of the 2014 ACM conference on SIGCOMM*. ACM, 2014, pp. 247–258.

[19] S. C. Nelson, G. Bhanage, and D. Raychaudhuri, "Gstar: generalized storage-aware routing for mobilityfirst in the future mobile internet," in *Proceedings of the sixth international workshop on MobiArch*. ACM, 2011, pp. 19–24.

[20] T. Vu, A. Baid, H. Nguyen, and D. Raychaudhuri, "Eir: Edge-aware interdomain routing protocol for the future mobile internet," *Under submission*, 2013.

[21] N. Somani, A. Chanda, S. C. Nelson, and D. Raychaudhuri, "Storage aware routing protocol for robust and efficient services in the future mobile internet," in *Communications (ICC), 2012 IEEE International Conference on*. IEEE, 2012, pp. 5849–5853.

[22] N. Basher, A. Mahanti, A. Mahanti, C. Williamson, and M. Arlitt, "A comparative analysis of web and peer-to-peer traffic," in *Proceedings of the 17th international conference on World Wide Web*. ACM, 2008, pp. 287–296.

[23] S. Mukherjee, A. Baid, I. Seskar, and D. Raychaudhuri, "Netwrok-assisted multihoming for heterogeneous wireless access scenarios," in *Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 25th International Symposium on*. IEEE, 2014.

[24] K. Ozbay and S. Mudigonda, "Simulated Vehicle Trajectory Data," Rutgers Intelligent Transportation Systems (RITS) Laboratory, Rutgers University, Piscataway New Jersey, Dec. 2012.

[25] K. Ozbay, S. Mudigonda, and B. Bartin, "Microscopic Simulation and Calibration of an Integrated Freeway and Toll Plaza Model," Presented at the 85th TRB Annual Meeting, 2006, Washington, D.C.

[26] F. Zhang, Y. Hu, Y. Zhang, and D. Raychaudhuri, "Poster: enabling mobile content-oriented networking in the mobilityfirst future internet architecture," in *Proceedings of the 15th ACM international symposium on Mobile ad hoc networking and computing*. ACM, 2014, pp. 441–442.

[27] M. Berman, J. S. Chase, L. Landweber, A. Nakao, M. Ott, D. Raychaudhuri, R. Ricci, and I. Seskar, "Geni: A federated testbed for innovative network experiments," *Computer Networks*, vol. 61, no. 0, pp. 5 – 23, 2014, special issue on Future Internet Testbeds Part I. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1389128613004507