TTR Mobility Extensions for Core-Edge Separation Solutions to the Internet's Routing Scaling Problem

Robin Whittle
First Principles
Rosanna Vic 3084 Australia
rw@firstpr.com.au

Steven Russert
Boeing Phantom Works
russerts@hotmail.com

RW 2011-10-30: Steve left Boeing in late 2009. This version of the paper uses the term "DITR" rather than "OITRD" for "Open ITR in the DFZ". In 2010 the IRTF Routing Research Group recommendation a core-edge elimination (locator-identifier separation) protocol ILNP. APT, TRRP and Six/One Router are no longer being developed. Sections which concern these have been rendered in light grey. I removed mention of full-database ITRs in Ivip, since this is no longer part of the design. I added notes about NERD scaling and the LISP approach to mobility. I had mistakenly considered both LISP and Ivip to be "locator-identifier separation" protocols. Using terminology developed in the RRG in 2009/2010, they are both "Core-Edge Separation" architectures.

ABSTRACT

Several router-based "~~locator/identifier separation~~" solutions have been proposed for the Internet's routing scaling problem, including the "map and encapsulate" systems LISP, APT, Ivip and TRRP. These are part of a class of scalable routing solutions known as "core-edge separation" systems – along with similar proposals involving address translation and novel forwarding techniques rather than encapsulation. These "core-edge separation" systems use a global system of Ingress Tunnel Routers (ITRs) near sending hosts to tunnel traffic packets to an Egress Tunnel Router (ETR) close to the destination network. Existing mobility techniques will not take advantage of such an architecture. Here we describe a new "Translating Tunnel Router" (TTR) wide-area mobility architecture which builds on the ITR, ETR and mapping system infrastructure of the core-edge separation system. This TTR approach to mobility promises to provide generally optimal paths for all traffic whilst supporting all existing IPv4 and IPv6 hosts as correspondent hosts, without need for upgrades. The mobile node (MN) retains a stable public IP address or prefix at all times, no matter what its current care of address(es) is or are. Furthermore, MNs will be able to use any access network, including those which provide care-of addresses behind NAT, since no mobility capabilities are required in the access network. This TTR global mobility architecture will work equally well with MNs and correspondent nodes using any local Mobile IP architecture. TTRs behave like ETRs to the core-edge separation system and somewhat resemble MIP home agents - however the MN chooses TTRs which are close to its access network, so there is no fixed home agent.

Keywords

Routing, addressing, mobility, multihoming, routing scalability, core-edge separation, locator-identifier separation, map-encap.

1. INTRODUCTION

Here we discuss a novel mobility scheme that can serve as an extension to the "core-edge separation" class of proposed enhancements to the Internet's routing and addressing architecture. [1] We describe the Internet's routing scaling and IPv4 address exhaustion problems together with several core-edge separation schemes which are being developed in response. It seems likely that one such scheme will be developed and widely deployed to enable the Internet to efficiently provide hundreds of millions or billions of end-user networks with multihoming, portability of address space between provider networks, and traffic engineering. Such a scheme would form a unique enabling system for a new "TTR" (Translating Tunnel Router) approach to IP mobility, which has little in common with current Mobile IP techniques, but which promises to surpass current techniques in several important respects:

The new system will work with the correspondent host being any existing IPv4 or IPv6 host. No changes are required to correspondent hosts or their networks, although in practice many networks will be upgraded to support the new core-edge separation architecture.

The TTR mobility system promises to provide generally optimal paths between mobile nodes (MNs) and correspondent nodes (CNs), including those which are mobile.

The MN needs only a Care-of Address (CoA) in any access network. This address can be behind one or more layers of NAT. Indeed the CoA can itself be an address provided by the same TTR-mobility scheme. There is no need for special mobility features in the access network, or for any business relationship between that network and the other elements of the mobility system.

The TTR mobility architecture builds on a core-edge separation scheme, using a mapping change in that system to switch traffic to TTR "near" the MN. These mapping changes are not needed frequently, since the one TTR is typically optimal or acceptable even when the MN is moving within its access network, using another access network in the same general area etc. As we discuss below, a mapping change is typically only likely to be needed if the MN's point of connection to the Net moves by some distance such as 1000km.

Since core-edge separation schemes do not inherently ease any of the technical or business-case difficulties which have so far prevented widespread deployment of traditional mobile IP, the future success of mobility arguably depends on making the best use of the core-edge separation enhancements to develop a new, global, mobility architecture.

While we discuss the proposed TTR mobility system as if it were an extension of an established, global, IETF-standardized map-encap scheme, the principles described here could be implemented by a single operator to create a novel and profitable mobility service, using an existing ITR and ETR system, or creating their own – without waiting for IETF standards.

We provide an overview of the proposed mobility architecture, how map-encap and other approaches to core-edge separation work with TTRs to form Level 3 of the new three level mobility system, how the TTRs and MNs form Level 2 and how any mobility systems inherent in the access network form Level 1. These techniques apply equally to IPv4 and IPv6. Next, the challenges to further growth in the Internet are presented, followed by a description of the proposed core-edge separation solutions. Finally, we show how the TTR mobility system applies to these core-edge separation approaches and end with a detailed example of MN mobility across various distances.

2. THE TTR MOBILITY SYSTEM

Our proposed mobility solution is applicable to today's globally routed Internet, as it would be enhanced by one of the core-edge separation scalable routing systems we discuss below. Since the TTR system requires no special software in correspondent nodes – nor extra features the networks used by correspondent nodes – all TTR mobility users will be able to use their mobile IP address(es) for 100% of their communications, so providing business incentives for early implementers.

The Translating Tunnel Router (TTR) is the foundation of the mobility system. A TTR need not be a hardware-based router. In the early years of deployment, it is more likely to be implemented as software on a COTS (Commercial Off The Shelf) server. TTRs are most likely to be located at Internet peering points, but they may also be located within access networks, particularly those of 3G and other major wireless networks.

A TTR mobility provider company would likely maintain TTRs at hundreds of sites, as close as possible to access networks all over the world. However, mobility could still be achieved with a single TTR, in which case it would appear much like a Mobile IP Home Agent (HA). By deploying a greater number of widely dispersed TTRs, the company would enable generally shorter paths to the destination node, through the TTR which is closest to the MN. Either the TTR itself or a server at the same site is responsible for managing the TTR, authenticating the MN's attempts to create 2-way tunnels to the TTR, and assisting the management system in determining which of the company's TTRs is topologically closest to the MN, for each access network through which the MN currently connects.

Each MN runs specialized tunneling software provided by the TTR company. This software may be globally standardized, but could be proprietary since it operates only between the company's TTRs and the MNs of that company's customers. Each MN obtains a Care-of Address (CoA) from one or more of its current access networks. It then establishes a 2-way encrypted tunnel from each such CoA to one of the company's TTRs. This TTR may be the optimal TTR in terms of network location, load, etc. or it may be an initial TTR to begin

providing service whilst the TTR management software selects a better TTR and signals the MN to open a new tunnel to that TTR.

2.1 3 LEVEL MOBILITY MODEL
While some Mobile IP techniques provide mobility only within certain networks, the TTR model provides mobility, with one or more stable IP addresses, on a global scale. The complete mobility model discussed here is composed of three distinct levels, with minimal interaction between each level.

2.1.1 Level 1: access network & MN
While some access networks such as wired Ethernet have no inbuilt mobility functions, for this discussion we will assume the use of access networks which do: the terrestrial wireless systems 3G/4G cellular, WiFi and mobile WiMax. All such access networks have their own internal mobility mechanisms concerning the MN connecting via different base-stations, access points etc. while maintaining a relatively stable CoA. Proxy Mobile IP (PMIP) also constitutes a level 1 mobility mechanism, since (within the PMIP domain) the MN sees no change in its CoA.

These are the Level 1 mechanisms of the complete system. Each access network is assumed to provide the MN with a single fixed or dynamically assigned CoA, perhaps behind one or more layers of NAT. No technical or administrative aspects of the access network are required to interact with elements of the other two levels of the TTR mobility model.

In practice, a 3G network in a large city would not necessarily provide a stable CoA as the MN roams from one area to another. Each region of the city may have its own IP gateway, so the MN may lose one CoA and gain another in the course of day-to-day movement, or even as base-station loads change so the MN is switched from one region's base-stations to those of an adjacent region. The new region's IP gateway will probably connect to the rest of the Internet at a different topological location to that of the previous CoA, potentially causing sub-optimal path lengths with the currently chosen TTR.

We class all movements of base-station, or change of L1 access technology (e.g. Wired to WiFi Ethernet) which result in the MN retaining its current CoA as instances of Layer 1: the local access network's inbuilt mobility mechanisms.

2.1.2 Level 2: MN & TTR
Level 2 of the TTR mobility model includes the MN being told (by a management system we discuss below) the address of one or more nearby TTRs, and establishing a 2-way tunnel from each of its one or more CoAs to the one or more such TTRs. Level 2 also concerns the one or more TTRs determining the reachability of the MN through each tunnel, and each TTR being aware of the costs, bandwidth limitations and packet loss characteristics of each tunnel.

These Level 2 mechanisms are all new and yet to be developed. They are not directly related to the core-edge separation scheme. Considerable sophistication would be required to achieve optimal outcomes in a wide range of circumstances. However, since the MN to TTR protocols need not be globally standardized, and would be chosen by negotiation between the

TTR and the MN, there is great scope for a variety of IETF standardized and proprietary techniques to be used to optimize Level 2's performance.

2.1.3 Level 3: TTR & global ITR network

Level 3 is the global core-edge separation system of ITRs tunneling packets to ETRs – or for mobility, to TTRs. Core-edge separation systems are not capable of tunneling packets directly from ITRs to MNs, since MNs' CoAs may be behind NAT and may change very frequently.

2.1.4 Relationships between the three levels

The Level 2 (MN-TTR) mechanisms function irrespective of the geographic or topological distance between the MN and any one TTR. In order to ensure optimal path lengths for packets to and from correspondent nodes all over the Net, the MN should tunnel to a TTR close to its current location. Once this is done the Level 2 management system changes the Level 3 (core-edge separation) system's mapping for this MN's address space so all ITRs will tunnel packets addressed to that MN to the new TTR.

The Level 2 management system controls which TTRs the MN tunnels to and the mapping of the MN's IP address or subnet in the core-edge separation scheme. The Level 2 management system does not need any particular knowledge of the topology of access networks, or of any Level 1 mobility features they provide.

Below we discuss the routing scalability problem and the core-edge separation schemes which are proposed to solve it. From that basis, we give examples of the core-edge separation scheme and TTRs working with MNs to create the complete, three level global mobility architecture.

3. INTERNET GROWTH CHALLENGES

Two problems stand in the way of future growth and manageability of the Internet: IPv4 address depletion, and inter-domain routing scalability. We discuss these because they are the impetus for the development of a new architectural enhancement to ensure routing scalability. The best developed of the scalable routing proposals are all "core-edge separation" (CES) schemes. Any one such CES scheme can be the basis for the TTR approach to global mobility.

3.1 IPv4 ADDRESS DEPLETION

Existing BGP techniques are administratively constrained to manage IPv4 space in large chunks of at least 256 addresses, each with a global cost to the BGP routing system. This leads to inefficient utilization [2] which, together with the growing demand for PI (Provider Independent) space, has lead to the imminent exhaustion of fresh space. [3][4] Since core-edge separation schemes generally enable address space to be managed much less expensively and in smaller segments than is practical with BGP, they are likely to encourage improved IPv4 address utilization and so help alleviate the IPv4 address depletion problem.

While opinions vary on how much scope there is for better utilization of IPv4 address space, it is inevitable that pressure to use the limited space more intensively will lead to a greater number of divisions, so fueling the growth in the number of advertised prefixes.

3.2 Routing Scalability

The core of the Internet uses BGP (Border Gateway Protocol) routers to forward packets between all its connected networks – those of providers and of the larger end-user networks. A large subset of these routers – probably well over 123k in number [5] – have two or more upstream links. These routers are regarded as being in the Default-Free Zone (DFZ) due to their need to develop a best path route for each BGP advertised prefix, rather than use a single default route for all packets not matching the local network's prefixes, as can a router with a single upstream link.

There are currently about 250k advertised prefixes (also known as "DFZ routes") in the global BGP routing table [6]. This number is growing unsustainably with a doubling time of approximately four years.

Each DFZ router conducts a separate BGP "conversation" with each of its neighbors for each of these 250k+ prefixes. For each prefix, the router chooses the "shortest" path advertised by each of its neighbors, subject to local policy (which may exclude or prefer certain neighbors for this prefix), and then advertises that path (perhaps made artificially longer to some neighbors, according to local policy) to its other neighbors. The metric by which alternative paths are evaluated for "shortness" is intentionally crude: the number of Autonomous Systems the path traverses before the destination network is reached. This simplification of some elements of the BGP control plane is crucial to its ability to scale to large numbers of routers and prefixes. (The interdomain routing system is far too large for routers to determine the best path based on complete knowledge about the current state of the network. BGP enables each router – and the entire network – to do a good job of choosing paths, while each router's "field of view" extends only as far as the paths offered by its immediate neighbor routers.)

While the whole BGP network today will converge on (adapt its best path decisions until a stable condition is reached) a good set of paths for all routers for each of the 250k+ prefixes, there are significant scaling problems which lead to concern about the ability of the BGP system to continue operating reliably in the future. Firstly, each router's RAM and CPU requirements depends largely on the number of DFZ routes, multiplied by the number of neighbors – and according to how often its neighbors change their best path advertisements.

There are scaling problems in the FIB (Forwarding Information Base) section of routers which handle the traffic packets, but the most urgent part of the routing scaling problem is due to the growth in the number of DFZ routes, and the rate at which each router sees them change. Projected rates of growth in the number of advertised prefixes exceed expected gains from faster CPUs and memory, and raise concern about the ability of the whole network to adapt rapidly to major outages, in which tens or perhaps hundreds of thousands of prefixes are affected.

A considerable proportion of the 250k+ currently advertised prefixes are those of providers - and this number is expected to grow. While there is no formal consensus on the matter in IRTF Routing Research Group (RRG), there is a widely held view that the biggest contributor to unsustainable growth in the number of BGP advertised prefixes is not the provider networks, but end-user networks.

A potentially vast number of individuals and organizations want and arguably need address space which can be multihomed via two or more providers and/or which is portable between providers (Provider Independent (PI), as opposed to Provider Assigned (PA) address space). At present, the only means of attaining such space is for each organization to obtain a prefix and advertise it in the BGP interdomain routing system. Stephen Sprunk, writing on the RRG mailing list [7] in late 2007, summarizes this viewpoint:

"As of last week, 87% of all ASes visible in the DFZ are origin-only. There are tens of thousands of medium and large leaf ASes not_ visible in the DFZ because they don't need ASNs, either because they have a upstream (transit) AS(es) announce for them or they're stuck on PA space. There are hundreds of millions of small leaf ASes, like my house, that can't get BGP from their upstreams, period, but might want EIDs so they can multihome over their DSL/cable/wireless lines.

"While the total number of visible ASes is going up, the number of origin-only ASes is growing faster than the number of transit ASes (i.e. the percentage of the former is growing). This is due to the increasing number of leaf ASes (e.g. large corporations) that are starting to visibly multihome, which is happening significantly faster than new transit ASes (e.g. ISPs) are being created.

"We've heard, in RIR meetings, over and over again that operators in the DFZ are scared of widespread multihoming and PI because each leaf AS requires a slot in the DFZ [a separately advertised prefix AKA a "DFZ route"], and there aren't (and won't be) enough slots available to handle the demand. This has resulted in high artificial bars to entry, denying a huge fraction of the Internet reliable service.

"If we are able to constrain BGP tables to only transit ASes, the DFZ becomes a lot smaller and we can afford to let everyone, even home users, multihome with PI space. (Of course, there's a hidden assumption there that the number of transit ASes will remain under control, but I haven't seen anyone dispute that.)"

4. Core-Edge Separation

Core-edge separation proposals to solving the routing scaling problem do not aim to reduce the number of provider prefixes which are advertised in the global BGP system (sometimes referred to as the "global routing table" or simply "the DFZ"). These proposals aim to create a new type of address space, which we will refer to as "Scalable PI" (SPI) space. SPI space is intended to suit the needs of end-user networks, not providers. Each core-edge separation proposal has its own way of providing this space, and ensuring that there is a much lower number of additional prefixes advertised in the BGP system than there are end-user networks using the new scheme.

Broadly speaking, this is achieved by making all new "SPI" space either not appear in any BGP advertised prefix, or by allowing for a relatively small number of advertised prefixes, each typically containing the SPI space of dozens to millions of separate end-user networks.

Initially the only proposals in what we now call the core-edge separation (CES) class of scalable routing solutions were known

(incorrectly) as "Locator/Identity Separation" protocols and/or (less formally, but quite reasonably as) "map-and-encaps" ("map-encap"). There are now two other classes of core-edge separation proposals, which we discuss briefly below. All these approaches are, in principle, capable of taking the role of Level 3 in the TTR mobility architecture.

4.1 Map-Encap Schemes
The map-encap proposals of 2006 onwards have their roots in a 1992 proposal by Robert M. Hinden, published in 1996 as RFC 1955. [8]

The Internet Research Task Force (IRTF) Routing Research Group [9] is currently discussing a number of broadly comparable router-based "map and encapsulate" proposals, each of which is intended to solve the "Routing Scalability Problem", as defined by the Internet Architecture Board's October 2006 Routing and Addressing Workshop [10][11]. The five most prominent proposals are LISP-ALT (Locator Identity Separation Protocol Alternative Topology) [12], LISP-NERD (A Not-so-novel EID to RLOC Database) [13], APT (APT: A Practical Transit Mapping Service) [14], Ivip (Internet Vastly Improved Plumbing) [15][16] and TRRP (Tunneling Route Reduction Protocol) [17].

Each of these "map-encap" proposals is applicable in principle to IPv4 or IPv6 and is intended to manage a subset of each address space to provide Scalable PI (SPI) space which is suitable for end-user networks which need multihoming, portability and traffic engineering.   (There is no accepted term for this new type of space, but we use SPI in this paper.)

While these map-encap schemes differ considerably, they share a common basic structure of a global system of Ingress Tunnel Routers (ITRs) which intercept traffic packets addressed to the end-user networks handled by the scheme and Egress Tunnel Routers (ETRs) that forward the packets on to their destination.

The SPI destination address of the packet is known as the "identifier" and is used by the ITR to look up some mapping information for the micronet (Ivip terminology) or EID prefix (Endpoint Identifier, in LISP and APT terminology) within which the destination address is located.   The "mapping" information determines the ETR to which the ITR tunnels the traffic packet.   The ETR, which is close to the destination network, decapsulates the packet and forwards it to the destination.   The authority to control the mapping for each micronet of mapped address space belongs to the end-user who rents or has been assigned this space.

Whether this ITR system is a single global IETF-standardized system or an independent special network using proprietary protocols and installed by a single company, the aim is for ITRs to be as close as possible to all sending hosts, so that the total path between the sending host and the ETR/TTR is no longer, or typically not much longer, than necessary.

These router-based CES schemes enable small or large blocks of address space, including individual IPv4 IP addresses and IPv6 /64 prefixes, to be used by end-user networks (or hosts) via any ETR in the access network of their chosen provider(s), with generally optimal paths for packets travelling from all sending hosts to the destination hosts in the end-user network.

These CES schemes differ considerably in their ITR and ETR functionality and in where these devices are located. One set of important differences between these schemes lie in the methods by which the ITR gains access to the mapping information it needs to correctly choose which ETR to tunnel each traffic packet to. Another set of differences concerns whether the scheme integrates the processes of detecting and responding to multihoming failures into ITR and ETR functionality, or makes it a separate task to be performed by some outside system, such as one run by the end-users. We discuss these differences in sections below.

4.2 Translation schemes

The first alternative to map-encap in the core-edge separation class of scalable routing solutions is Christian Vogt's Six/One Router proposal [18] (not to be confused with an early SHIM6-like host-based proposal "Six/One"). Six/One Router does not use encapsulation, but has Translation Routers, at the core-facing borders of provider networks, which translate the source and destination addresses of packets entering and leaving the network.

Broadly speaking a Translation scheme (of which Six/One Router is currently the only instance) resembles a map-encap scheme, with its mapping system, and separation of edge end-user networks, using what we refer to as SPI address space. However Translation Routers replace ITRs and ETRs and packets are not encapsulated, or made any longer at all.

Each SPI end-user edge network has its own prefix of address space which is not globally advertised. This achieves the central aim of core-edge separation. Each such network connects to the Net via one or more provider networks, and at each such provider, a similar-sized portion of a provider prefix is matched to the SPI prefix of the end-user network. Thus an end-user network with a /48 of SPI space would be accessible from the core by two /48 prefixes within larger blocks (shorter prefixes) of space advertised by each provider.

In principle, Six/One Router is applicable to both IPv4 and IPv6, but due to the shortage of IPv4 address space, this "prefix mirroring" approach is only practical for IPv6.

Translation schemes have a profound advantage over map-encap: the packets are no longer. This makes the "tunneling" part of the core-edge separation system 100% efficient in terms of bandwidth, and does not create extra Path Maximum Transmission Unit (PMTU) problems due to traffic packets being made longer. In principle, a translation scheme might be capable of supporting standard RFC 1191[19] Path MTU Discovery (PMTUD) – which is something which map-encap schemes cannot do without a great deal of extra complexity in ITRs and ETRs. [20]

We have not yet discussed with Christian Vogt whether the TTR system would work with Six/One Router. In this paper, we assume that the two systems could be adapted to work together. In this paper we treat Six/One Router as being functionally similar to a map-encap scheme, or to one of the versions of Ivip with a forwarding approach to transporting data from ITR to ETR.

Six/One Router is a core-edge separation scheme, with a mapping system and a method of directing traffic packets to any desired end-user network in a scalable fashion. However, it does not use ITRs or ETRs. The packets which are addressed to a given end-user network are not tunneled to a single device such as an ETR, so there is no obvious point in Six/One Router for the Translating Tunnel Router.

To use the TTR approach with Six/One Router, for any given edge prefix (the minimum span of address space which can be mapped to a transit prefix) the TTR function would be performed by a router which advertised that transit prefix. A TTR inside a provider network which was able to handle N MNs would need to have N prefixes of a size suitable for however large the edge prefix is for each MN. For instance, if all MNs used a /64, and a TTR could handle 1024 simultaneous sessions with MNs, it would need a /54. This could be within a shorter prefix of a given provider, so multiple TTRs could each use a part of a shorter prefix which is advertised by the provider border routers.

4.3 Forwarding schemes

Two new approaches to transporting packets from ITRs to ETRs have recently been proposed by one of the authors. Both involve using a modified format of the existing IP header to carry enough bits to control the forwarding behavior of core routers, in order that the packet will be forwarded to the ETR – while the packet retains its original source and destination addresses. The major advantages of both schemes are absence of encapsulation overhead and direct support for RFC 1191 PMTUD without ITR involvement. The major disadvantage is the need to upgrade essentially all core routers, and some or all internal routers, to support the relatively simple alterations to processing packets with the modified headers

Both systems are applicable to Ivip and could be used in the long term, as a more efficient and elegant approach than encapsulation. In the future, if the requisite routers could be upgraded in a sufficiently short time, it would be possible to introduce Ivip with the forwarding technique alone, without encapsulation and the complex ITR functions this requires in order that PMTUD is properly supported. The first scheme is for IPv4: ETR Address Forwarding (EAF) [21]. The second is for IPv6: Prefix Label Forwarding (PLF) [22].

4.4 Mapping Distribution systems

Each core-edge separation scheme requires that information relating to current mappings be conveyed to distant parts of the network

The mapping data is of a different nature, or has different terminology for the different schemes. For LISP and APT, the mapping is "EID to Locator": for a given endpoint identifier prefix, which one (or more, for multihoming) locator (ETR) address the ITR should tunnel packets to.

For Ivip map-encap, EAF and PLF, it is "micronet to ETR": for a given micronet of SPI end-user address space, the ETR address which ITRs should tunnel the packets to.

For Six/One Router, the mapping is "edge prefix to transit prefix": for this edge (end-user) prefix, the one or more provider (transit) prefixes the destination address should be translated to.

The mapping information distribution system must push the mapping information to the ITRs, have ITRs pull the

information on demand from local or remote query servers, or use some hybrid of push and pull.

4.4.1 Pure push

Pure push provides the full global database of mapping information at every ITR, so each ITR already has the mapping information it needs whenever it receives a packet whose destination is to an SPI address (a mapped address, within a micronet or EID prefix). Pure push (LISP-NERD), however, cannot provide the complete global set of mapping information in an up-to-date manner without incurring excessive costs, both in transmitting the mapping data across the network and in storing the entire database at each ITR. [The Jan 2010 update draft-lear-lisp-nerd-07 argues that NERD scales to 10^8 EIDs.]

Mass-market hard disk drives and DRAM are capable of storing the multiple gigabytes of data which would constitute a mapping system used by billions of individual cell-phone users. Ignoring the storage cost objection, the cost of maintaining the full feed of mapping updates to each ITR is still a scaling concern. For a given financial cost, the data carriage costs of full push reduces the number of ITRs and so limits the flexibility with which they can be placed in the network, while requiring each one to handle more traffic. Pure push precludes the nearly zero cost option of having caching ITR functions in sending hosts or DSL modems.

4.4.2 Pure pull

Pure pull systems (LISP-ALT and TRRP) avoid this problem but must trade-off timeliness of the mapping information, caching times and query-response volumes. When a packet arrives addressed to an EID prefix for which the ITR has no mapping information is cached, the ITR must drop or delay the traffic packets whilst the mapping information is fetched. Both these schemes have alternative delivery schemes for these initial packets, but these too involve significant delays and reliability problems.

4.4.3 Hybrid push-pull

Hybrid push-pull systems (APT and Ivip) chart a path between the extremes of pure push and pure pull to create a responsive system that does not excessively burden the global mapping distribution system with control plane overhead in the form of having to push all mapping updates to all ITRs. The full mapping information is pushed to local full database. All ITRs cache the mapping they receive after sending a map request message to a nearby (such as in the same ISP network) full database query server.

4.4.4 Ivip's fast hybrid push-pull mapping system

Ivip differs from the other proposals in several respects. Ivip uses an ambitious "fast push" system to transmit the end-user's command for a new mapping for their micronet to all full database query servers in the Net, within 5 seconds or so. This provides each end-user with essentially real-time control of to which ETR all the world's ITRs will tunnel packets which are addressed to the end-user's micronets.

One benefit of this a simplification of the mapping data which must be provided for each micronet – to just a single ETR address. In all other core-edge separation schemes, the mapping for a multihomed network consists of two or more ETR addresses, with weights and priorities by which each ITR can choose which to tunnel packets to, depending on whether each

ETR is reachable and according to the traffic engineering (load sharing) desires of the end-user.

The primary purpose of the fast hybrid push-pull mapping distribution system is to give end-users complete control of the decision making process which determine the mapping of their micronets, including complete control of all reachability testing which needs to be carried out in order that these decisions can be made.

This gives rise to the most important difference between Ivip and the other CES schemes developed so far: Ivip is a modular subsystem which contains no mechanisms for reachability detection of multihoming service restoration decision making. In contrast, all other current CES system monolithically integrate these functions into ITRs and ETRs, leading to greater complexity and costs and more detailed and voluminous mapping information. This integration prevents end-users from implementing any approaches which are more sophisticated or suitable to their needs than whatever is provided by the necessarily limited functionality built into every ITR and ETR.

LISP-ALT/NERD, APT and TRRP all require each ITR to test ETR reachability and make decisions, in isolation from other ITRs, about which alternative ETR to tunnel traffic to in the event the preferred ETR becomes unreachable. Ivip requires end-users (or some system operating on the end-user's behalf) to perform multihoming failure detection and to make their own decisions about mapping changes, such as to direct traffic to a different ETR.

While all these map-encap schemes are in principle suitable for supporting the TTR approach to mobility, Ivip would support it best because it enables each end-user to change their mapping effectively in real-time, (~5 seconds). In many TTR mobility scenarios, such short response times are not required. Nonetheless, it is desirable to control ITR tunneling as rapidly as possible.

APT and LISP-NERD aim for mapping update times much longer than this, in the range of tens of minutes to hours. LISP-ALT and TRRP, being "pure-pull" systems can in principle provide fresh mapping information in map reply messages within a second, or so. However this would not allow rapid control of ITR tunneling, except to the extent that ITRs repeatedly queried mapping information for all EID prefixes (micronets) for which they are currently handling traffic. It would be impractical to achieve, for instance, 30 second response times in this manner due to the heavy load on the query servers and the high volumes of query traffic traversing the distributed global query server network.

4.5 Support for hosts in networks without ITRs

It is vital that any CES scheme support packets sent from hosts in networks which have not been upgraded with ITRs. If end-users adopting SPI space were to find that the new portability, multihoming and TE arrangements only applied to packets sent from other networks which had adopted the CES scheme, then there would be very little incentive for early adopters to use the system. Even if adoption rose to 90% or so, there would still be serious difficulties with multihoming etc. not working for

packets sent from the 10% of networks which have not yet installed ITRs.

APT and TRRP are in principle capable of supporting packets from non-upgraded networks. Six/One Router supports only basic connectivity from non-upgraded networks: multihoming etc. only works for packets sent from upgraded networks. Below we describe the two best developed techniques by which CES systems provide "backwards compatibility": portability, multihoming and TE for all incoming packets, including those from hosts in networks without ITRs.

4.5.1 Ivip DITRs

While many provider and end-user networks will have ITRs to tunnel outgoing packets which are addressed to SPI address space, Ivip will involve numerous widely dispersed "Default ITRs in the DFZ" (DITRs). [Until 2010 these were called "Open ITRs in the DFZ" (OITRDs).] which will tunnel such packets sent from networks without ITRs.

Every prefix of address space which is removed from conventional BGP management, and instead handled by Ivip's mapping systems, ITRs and ETRs, is known as a "Mapped Address Block" (MAB). Each MAB is operated by a single organization, who leases space in smaller chunks to end-user networks, who themselves decide on how their space is split into micronets, and to which ETR each micronet is mapped.

DITRs will be operated by the organizations who will lease SPI space to end-users. The cost of running the DITRs will be recovered by charging end-user networks for the traffic handled by DITRs for their.

While only one DITR is required to ensure connectivity – to attract packets sent by hosts in all networks without ITRs – generally the best outcomes will result from numerous DITRs being placed around the Net, so there are generally shorter paths between each sending host, the nearest DITR and the ETR which handles the micronet to which the packet is addressed.

All DITRs for a given MAB advertise this MAB in BGP, so causing packets from any ITR-less network to be forwarded to the closest DITR which advertises the MAB which matches the packet's destination address. In principle it would be sufficient to have a single global system of several thousand of DITRs, each advertising every MAB. A more likely scenario is a mix of DITRs run by specific organizations who lease out MAB space – and DITRs run by companies for those organizations, and so which advertise the MABs of multiple organizations.

Generally, DITRs need to be widely distributed, due to sending hosts and ETRs being located potentially anywhere. If, however, it was known that all ETRs for all a MAB's micronets were located in a given country or region, then generally optimal paths from sending hosts all over the world could be achieved by locating DITRs only in that country or region. This might be the case if one or more MABs were run by an organization such as a university or government, purely to provide SPI space for its own departments, which were all located within the one country or region.

4.5.2 LISP Proxy Tunnel Routers

LISP Proxy Tunnel Routers (PTRs) are in principle capable of perform much the same functions as Ivip's DITRs. However the usage models and business cases for PTRs are less developed than for Ivip's DITRs.

5. Layer 2: MN to TTR

In principle, the TTR approach to mobility is equally applicable to any of the core-edge separation schemes: encapsulation, translation or one of the new modified header forwarding schemes. All these systems have a similar overall structure of a mapping system which controls the tunneling behavior of ITRs for packets addressed to each of potentially billions of SPI destination prefixes. Packets are tunneled from ITRs to ETRs across the core, with all ITRs tunneling packets addressed to a given micronet of SPI space to any given ETR at a particular time. (For simplicity of discussion, we ignore how the ITRs of LISP, APT and other non-Ivip schemes can be told by the mapping information to load share traffic between multiple ETRs, and to tunnel to a second ETR if the first one appears unreachable.)

Each TTR behaves to the core-edge separation scheme exactly like an ETR.

TTRs always use two-way tunnels, established by the MN, for communicating with the MN, irrespective of whether the core-edge separation scheme uses encapsulation, translation of forwarding to tunnel packets from ITRs to the ETR function of the TTR. So a TTR never initiates contact with a MN. The MN must establish contact with one or more TTRs, and through that tunnel may be directed by the TTR company's management system to establish tunnels to one or more other TTRs.

5.1 INFREQUENT MAPPING CHANGES

Frequent mapping changes are not required in the TTR mobility approach. Each mapping change selects a new home-agent-like TTR – which typically only needs to occur when the mobile host moves a significant distance, likely more than about 1000km.

This 1000km figure is a very rough estimate, based on the assumptions such as the extra latency involved in distances up to 1000km or so being acceptable for VoIP packets. If very high volumes of packets with physically nearby correspondent hosts is part of the usage pattern, and/or if the move to a new area will result in a lasting new location for the MN, then it makes more sense to choose the closest possible TTR rather than whichever one has been used previously.

5.1.1 Research into individual movement patterns

Some estimates of the total frequency of mapping changes for a system serving a large population could be gained from studying existing research of individuals' physical movements. Airline flight and other transport statistics are a good source of raw data for such enquiries.

For instance, the US Bureau of Transport Statistics [23] lists 677 million domestic airline passengers per annum, with 10.2 million aircraft departures. Assuming a worst case scenario of each such flight involving three mapping changes for every passenger, this is an average of 21 mapping changes a second. An Ivip IPv4 mapping change involves about 12 bytes of data, so while this represents only fraction of global airline traffic,

and while there would be peaks and troughs in the update rate, the data rate required to carry these updates averages only 2k bits per second. A more realistic estimate would involve fewer mapping changes, due to many flights being only a few hundred km, and would account for only a subset of passengers wanting continual Internet access on their own portable device during their flight.

Albert-Laszlo Barabasi and colleagues followed the movement of 100,000 individuals, measured by cellphone basestation data. [24]. While this survey would not detect airline travel, it shows patterns of movement where in a week long survey, "most individuals travel only over short distances, but a few regularly move over hundreds of kilometres".

5.1.2 Mapping changes are not crucial to connectivity

The mapping change is part of Level 3 of the TTR mobility model. It is generated by the TTR company's management system. As long as a micronet is mapped to a given TTR, Level 2 involves the MN establishing multiple tunnels from various CoAs to that TTR. Level 1 is any intrinsic mobility features of the access network(s) currently used by the MN. These enable the MN to switch between multiple base-stations while retaining the same CoA.

While 5 second response times for ITRs changing their tunneling from one TTR to another may seem excessive in a mobile IP setting, the selection of a new TTR is not needed due to any problem with connectivity, but solely to maintain generally optimal total path lengths. As such, while it is desirable if mapping changes can be made at any time with minimal delay, the mapping change is not urgent or required to maintain connectivity, but simply to choose one TTR over another, for reasons such as one being closer to the MN, or being less congested, more reliable etc.

If the MN has a tunnel to its old TTR (which is close to its initial access network but distant from a second and now preferred access network) then the management system will detect the new location and instruct the MN to establish a tunnel to one or more closer TTRs. Once the new tunnel is established, even if the mapping change to tunnel packets from ITRs to the new TTR is delayed by seconds or minutes, no disruption will occur, since the MN will receive incoming packets from the old TTR, and sends outgoing packets via either the old or new TTR.

Some access networks, such as 3G networks in large cities, use multiple IP gateways, giving the MN a different CoA when it moves only a short distance. Some loss of connectivity may be inevitable in any radio mobile network. In the event of the loss of one CoA and the gaining of a new one, the MN will establish a 2-way tunnel to the same TTR and resume its communication sessions, without requiring any mapping change.

In addition to its basic ETR function of decapsulating traffic packets tunneled from ITRs, the TTR is the end-point of 2-way tunnels from the MN, and so may be simultaneously handling such tunnels from hundreds or thousands of MNs simultaneously.

5.2 TTR company's real-time control of mapping

The MN itself does not control the mapping of the micronet(s) to one or another TTR. The one or more micronets of SPI space "belong" to the owner of the MN – via a lease arrangement with the company who runs the MAB the micronet is within. In order that the one or more micronets can be used with the TTR mobility system, the owner gives the TTR company the permission and requisite username, password etc. the company needs to control the mapping for these one or more micronets. The MN owner may withdraw this permission at any time, and select another TTR company to control the mapping of the micronet(s). The TTR company physically controls the mapping of the micronet by an authenticated session which directly or indirectly interfaces with the Root Update Authorization Server (RUAS) company which controls the mapping for the MAB which each micronet is a part of. (RUAS is an Ivip term.) The RUAS organisation may be the same company who the MN owner leases the micronets from, or the micronets may be leased from a separate MAB company who contracts this RUAS company to handle the mapping for this MAB.

With Ivip's fast hybrid push-pull mapping update distribution system, commands from the TTR's management system will be fanned out to all the world's full database query servers within a few seconds.

Those query servers will immediately convey the changed mapping to any local ITRs which recently requested the mapping of this micronet. This is achieved by sending a cache update message directly to these ITRs, secured by a nonce which the ITR sent to the query server in its initial map request.

5.3 Mapping changes incur a small fee

Since the RUAS engages the considerable global resources of the distributed Ivip fast hybrid push-pull mapping distribution system, it charges end users per mapping change. So the MN owner ultimately pays for each mapping change, and will probably pay for their share of traffic flowing through the DITRs which the MAB company runs in order to make the micronets in the MAB reliably reachable from hosts in networks without their own ITRs.

It is a vital part of the Ivip approach to scalable routing that the end-user pays for most or all of the burden their traffic and mapping changes place on the shared infrastructure of the global fast hybrid push-pull mapping system, and of the MAB company's DITRs.
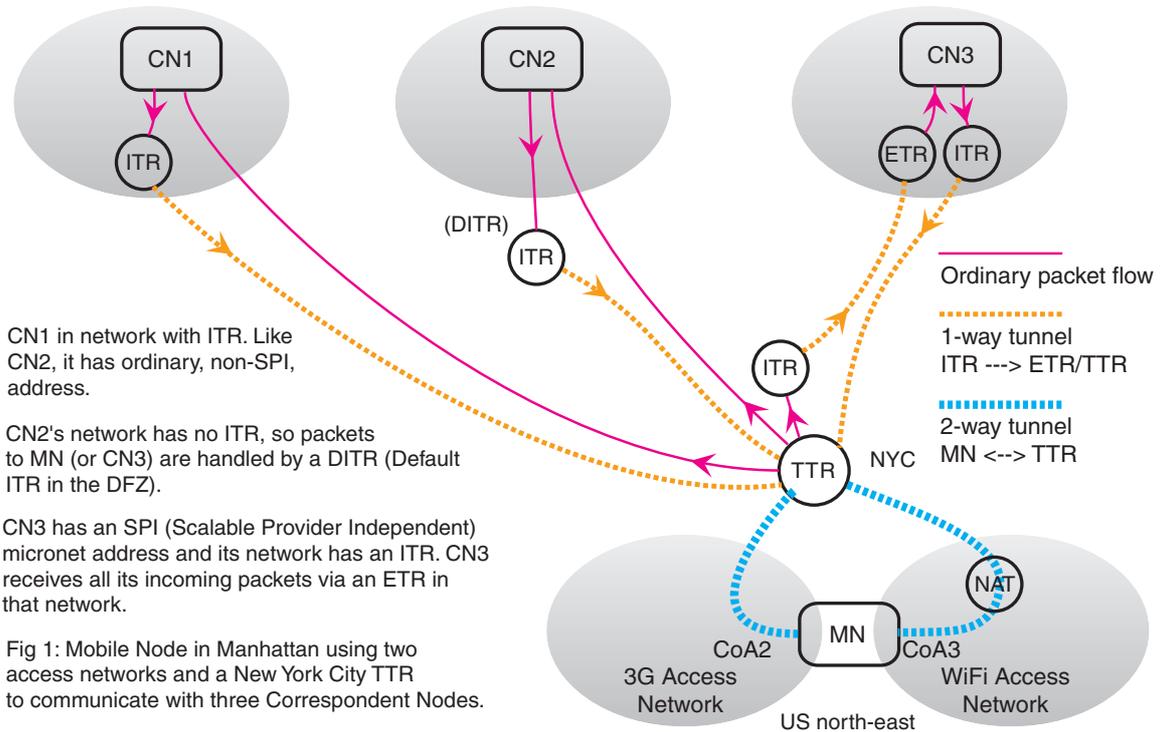
These fees need not be so high as to discourage widespread adoption, since mass adoption leads to great economies of scale.

While the MN owner authorizes the TTR company to control the mapping of their one or more micronets, it is the owner who pays for those changes. Consequently, depending on customer preferences which prioritize low costs or rapid selection of the closest and best TTR, the TTR company would employ a variety of strategies in determining how frequently to change to a closer TTR.

Since the Internet works quite well on a global basis, most mobile end users would not need frequent mapping changes to select a better TTR simply because they moved a few hundred km. As long as the TTR is within about 1000km of the border

router of their current access network, there should be little or no perceivable problem with latency or packet loss.

CN1 in network with ITR. Like CN2, it has ordinary, non-SPI, address.

CN2's network has no ITR, so packets to MN (or CN3) are handled by a DITR (Default ITR in the DFZ).

CN3 has an SPI (Scalable Provider Independent) micronet address and its network has an ITR. CN3 receives all its incoming packets via an ETR in that network.

Fig 1: Mobile Node in Manhattan using two access networks and a New York City TTR to communicate with three Correspondent Nodes.

6. OPERATIONAL EXAMPLE

We now discuss TTRs, the management system of a TTR network, and the tunneling software which is installed in the MN. In the following examples depicted in Figures 1 and 2, the MN is a mobile laptop computer with an SPI "micronet" (range of address space covered by a single mapping) – of a single IPv4 address. The same principles apply for one or more micronets of any size.

In the following examples, the end-user leases their micronet space from an organization which is part of the basic Ivip system and does not have any direct mobility role. The end-user is also a customer of a TTR mobility provider company. In practice the TTR company may also provide the micronet space as part of the service. There may be many such TTR companies, with the end-user being a customer of several, but in the following we assume the end-user's MN uses a single global network of TTRs. The end-user will pay for traffic passing through these TTRs, as well as for traffic packets handled for their micronet by the DITRs operated for or by the company they lease their address space from.

The end-user may be a customer of the one or more access networks. However, no mobility arrangements are needed in any such business or technical relationship. The TTR mobility system works equally well with an ad-hoc connection such as an office Ethernet cable, or a free WiFi system in a public space.

In our example, a laptop MN can connect simultaneously to 3G and WiFi networks, as well as via cabled Ethernet. Its operating system automatically gains a CoA on each such access network, and its TTR-company-supplied tunneling software makes a 2-

way encrypted tunnel from each such CoA to whichever TTRs the TTR company's management system suggests. At all times, the MN's tunneling software maintains a link to the TTR management system via one or more tunnels from one or more of its CoAs to one or more TTRs and/or to a centralized server.

Our example begins with the laptop plugged into a home DSL service in Manhattan, which gives it an address, behind NAT: CoA1 – which is not shown in the diagrams. The MN has established a 2-way encrypted tunnel to the NYC TTR. The MN could use its care of address CoA1 conventionally (for communication with any host in the Net), but here we assume the tunneling software uses each CoA to tunnel to one or more TTRs, thus maintaining the MN's public, stable, globally mobile, SPI micronet address so applications can use that address for initiating and accepting communications.

6.1 TTR discovery

Initially, the MN would connect to the TTR company via a centrally located TTR, for which it would obtain the address via a conventional DNS lookup.

Once the MN has established a tunnel to one or more such TTRs – which may be located in a country distant from the MN – the TTR company's management system attempts to determine where the MN is located, and to find one or more closer TTRs for it to tunnel to.

The MN plays in important role in this process, since tunnels can only be established from the MN to the TTR, not from the TTR to the MN – due to the MN's CoA being potentially behind NAT. However the process of suggesting TTRs and deciding

which ones to connect to, and which to use for traffic is made by the TTR company's management system, rather than by the MN.

6.2 Tunnels to multiple TTRs

In our example, several application programs each open a SSH session from the MN's stable address. The MN can also run servers, since its micronet of SPI address space is public and remains the same no matter where its current CoA is. Outgoing packets for each SSH tunnel are encapsulated by the MN's tunneling software and pass through the DSL modem's NAT function. They arrive at the NYC TTR, where they are decapsulated and forwarded normally to the rest of the Net. The TTR may integrate an ITR function so outgoing packets addressed to SPI addresses (such as the addresses of CN3) are encapsulated and tunneled immediately, without relying on any external ITR. However, in Figure 1, we show a raw packet emerging from the NYC TTR and being forwarded to a nearby ITR, which encapsulates it and tunnels it to the ETR which handles the micronet of space which CN3 is within.

Correspondent hosts all over the world send packets addressed to the MN's stable public (SPI) address. In principle the MN could have multiple stable, public, SPI addresses, if one micronet spans multiple IP addresses and/or if the MN has multiple micronets. These one or more micronets must be mapped to the one or more TTRs with which the MN currently has 2-way tunnels.

6.3 Establishing a second CoA

In our Figure 1 example, the MN finds a 3G signal and establishes a CoA2 address in that network. CoA2, like all other CoAs, may be used by the MN to communicate with other hosts in that access network, or (perhaps via NAT) with hosts anywhere in the world. Here we discuss how the MN uses CoA2 to exchanges packets with the TTR company's management system, via new 2-way tunnels from this CoA2 address to one or more TTRs.

The new tunnels may be to the first TTR in NYC, or to a central TTR whose address is obtained from DNS. The TTR company's management system uses traceroute and/or other techniques to determine that the 3G access network's IP network has border router in NYC too. So the MN is instructed by the management system to establish a second 2-way tunnel to the NYC TTR, if its tunnel from CoA2 is not already to that TTR. (It is also possible to modify the TTR selection algorithm to emphasize robustness over path length, by ensuring that when multiple tunnels are established, they go from the MN to more than one TTR).

When the Ethernet cable is unplugged, the MN and TTR detect this and use the 3G tunnel instead. Ideally this would involve a fractional-second delay and no lost packets. There is no requirement to change the mapping of the MN's micronet(s) in the global Ivip mapping system, since both active tunnels are with the same TTR, to which the micronet(s) are currently mapped.

As the MN is carried out of the house and into a subway station, it acquires a WiFi connection from the subway, and similarly establishes a CoA3 there, and a third tunnel to the NYC TTR. Sophisticated TTR management software would ideally direct traffic to the faster, cheaper, WiFi tunnel, while maintaining the 3G tunnel for management purposes and in readiness to carry traffic in the event the WiFi link failed.

Note that all layer 1 mobility arrangements, which give the MN the same CoA as it moves from cell to cell, are within the 3G and WiFi networks. These are Level 1 of the complete mobility system and require no coordination with the MN's software, the TTR system or the Ivip core-edge separation system.

The 3G link fails when the laptop enters the subway carriage, and (ideally) traffic continues on the WiFi link. At the end of the trip, a new 3G connection and CoA4 is acquired and a 2-way tunnel built from there to the NYC TTR.

The 3G link carries the traffic after the WiFi connection ends, and when the laptop acquires a WiFi or cabled Ethernet connection in the office, the tunneling software establishes another 2-way tunnel to the NYC TTR from this new CoA5 (not shown). Ideally, the TTR management system recognizes this link's lower cost and higher capacity – and perhaps with configuration information from the user, the fact that this connection is likely to persist for many hours.

Optimal decision making by the TTR management system is likely to involve some degree of end-user customization, such as to nominate particular access networks which are preferred in terms of low cost, high speed etc. For instance the end-user would configure their account with the TTR company to prefer the wired or WiFi Ethernet link at home, and those links at work, over other forms of connection.

The TTR company's management system would generally not be aware of the nature of the final physical link, but would be able to detect which network the MN had CoAs on, by the access network prefixes within which each CoA falls. If the CoA is behind NAT, the TTR ascertains the public address of the NAT box from the source address of the packets it receives from the MN, and so determines which access network this particular CoA is within.

In our example, the TTR company's management system instructs the MN's tunneling software to end the 3G connection and continue using the office CoA5 tunnel for all traffic.

In all this time, all applications including servers and clients continue to function within the limits of at least one access network being connected, and the MN's stable public IP address is maintained, with generally optimal paths to and from correspondent hosts in all locations.
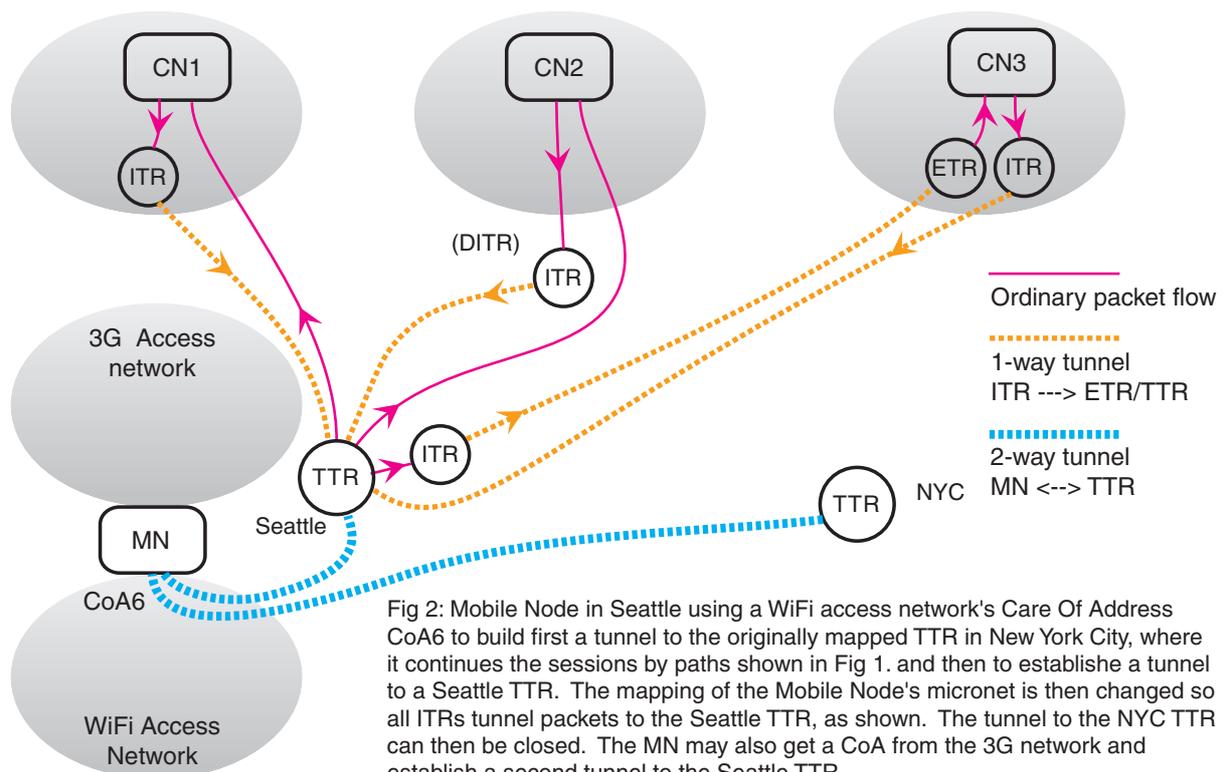
Fig 2: Mobile Node in Seattle using a WiFi access network's Care Of Address CoA6 to build first a tunnel to the originally mapped TTR in New York City, where it continues the sessions by paths shown in Fig 1. and then to establishe a tunnel to a Seattle TTR. The mapping of the Mobile Node's micronet is then changed so all ITRs tunnel packets to the Seattle TTR, as shown. The tunnel to the NYC TTR can then be closed. The MN may also get a CoA from the 3G network and establish a second tunnel to the Seattle TTR.

6.4 Moving across country

We now consider (Figure 2) the MN being turned off in NYC and then on again, in Seattle, where it acquires a WiFi signal in the airport. Its micronet is still mapped to the NYC TTR, and the MN establishes a tunnel initially from this new CoA6 address to that TTR, restoring connectivity. However, the management system is able to determine (for instance by tracerouting from one or more of its TTRs to CoA6) that the MN is now closer to its Seattle TTR and far from its NYC TTR.

The TTR company's management system instructs the MN to establish a 2-way tunnel with the Seattle TTR, and when this is operational, the management system changes the mapping of the MN's micronet(s), so ITRs all over the world tunnel packets to the Seattle TTR instead. No connectivity need be lost during this time and the same TTR can be used for 3G and Ethernet connections in the Northwest. The MN may now establish other tunnels to the Seattle TTR from CoAs in other access networks.

There is no absolute need to change the mapping. Each mapping change will probably have a low, but non-zero, financial cost to the end-user So it will be worth changing the mapping whenever the new closest TTR is a significant distance, such as 1000km or more, from the currently used one.

7. ADVANCED TECHNIQUES

This TTR approach to mobility could be used for a micronet of a passenger airliner using satellite or terrestrial links to various ground stations.   Here we discuss a variety of advanced techniques which demonstrate the flexibility of the TTR

approach, and its ability to work without any prior arrangement in whatever access network the MN connects to.

The following examples involve a passenger aircraft using the TTR mobility approach to maintain ideally continual Internet connectivity, using different satellite ground-stations and therefore different TTRs as it travels. It is tempting to assume that airliners of the future could provide continual connectivity at all times. This may be possible over land, using radio links direct to numerous ground stations, but for flights over oceans (in the absence of fiber-optically connected buoy ocean-stations catering for the trans-Atlantic route), the airliner must rely on geostationary or other satellites for its connectivity.   In Appendix 1 we briefly discuss some barriers to achieving continual connectivity via satellites.

7.1 Nested mobility systems

A MN with a CoA behind NAT with a public address which is part of the airliner's SPI address space (the airliner's micronet) could itself use the TTR mobility system for global mobility with its own micronet(s), tunneling via the aircraft's MN to TTR link to its own TTR.

This illustrates the ability of the TTR approach to work for MNs on any kind of address, including those behind NAT and/or those on the micronet address space of another system. This includes micronet space of an end-user network using purely the CES system of ITRs and ETRs, and of an end-user network using the TTR mobility extensions to the CES scheme.

In principle, any recursion of the above is true. For instance a MN1 could establish tunnels to one or more nearby TTRs and so

have generally optimal paths to and from all hosts it communicates with, even if its CoA1 was part of a micronet of another MN2, while MN2's CoA2 was behind one or more layers of NAT with a public address within the micronet of a MN3 which had a CoA3 was part of a micronet of passenger airliner which switched its ground-station, and therefore its TTR, as it travelled around the globe. This rather contrived example might involve MN2 being a laptop in an aircraft cabin, which itself had its own links to other devices, such as MN1, providing each with an address from MN1's micronet.

None of the three MNs need to have any knowledge of each other, and they all may be using different TTRs, including from different companies.

Furthermore, each MN may be using an access network which involves considerable local mobility functionality (Level 1 in our TTR model). None of the various levels of access network or the various MN's mobility arrangements, need be known to any other MN. For instance, in an even more contrived instance of the above 3 MN example, MN2 might be the gateway for a MANET in which multiple other laptops in the cabin communicate via WiFi. MN2 provides CoAs from its micronet for all nodes such as MN1 which access the MANET. MN1 may be a Bluetooth device and the MANET may involve extensive Layer 1 mobility functionality, such as enabling MN1 to retain the same CoA1, no matter which of the MANET laptops in the cabin it is currently communicating with via Bluetooth.

Our example below is less elaborate. It concerns a passenger aircraft using TTR mobility for its own micronet, of which one address is the public address of a NAT box. Behind the NAT box, multiple laptops gain Internet access, and in our example, one of them is also using TTR mobility to use its own micronet of one or more IP addresses, which it retains no matter what its current CoA(s) and no matter what access network(s) it is using.

Whether using a single IETF-based global system, or proprietary protocols and a private ITR and TTR network just for this purpose, the TTR mobility architecture would have significant scalability and performance advantages over the BGP-based approach of moving the advertisement of each aircraft's /24 prefix from ground-station to ground-station, as with Boeing's Connexion system [25] or the MIP-based nested NEMO solution. This BGP approach – the only one available with current techniques – involves an excessive number of BGP advertised prefixes being used, with such frequent changes of which router advertises the prefix as to unreasonably burden the global BGP system.

Furthermore, such frequent changes may be deemed by some routers to be symptomatic of router instability and/or unreasonable use of the BGP system, leading to such routers not recognizing and so failing to propagate the changes for these prefixes. This would result in some parts of the Net being unreachable from hosts in the aircraft.

7.2 Care-of-Addresses within mobile micronets

If the MN used WiFi in the aircraft cabin, to gain a CoA7 behind a NAT box in the aircraft, the NAT box's public address would be unchanged for the whole flight, but its point of connection to the rest of the Net would change. For instance on a flight from Seattle to London, a connection via one geostationary satellite and then another would move from a ground station in Colorado to one in Switzerland. The following discussion applies whether the plane's micronet is part of the main Ivip SPI system, or is implemented with similar principles for a proprietary Ivip-like ITR, ETR and tunneling system just for aircraft and their ground stations.

The TTR management system would need to detect the change of ground station by periodic traceroutes or by some other mechanism, such as by monitoring the mapping of the plane's micronet. (This last technique would be straightforward if the aircraft used the global ITR, ETR and mapping system. If the aircraft used a proprietary system, the mapping information would only be available to the TTR company's management system by special arrangement.)

Continuing from our Figure 2 example, when the MN is carried into the aircraft in Seattle and established its CoA7 address, the MN's mobility software automatically establishes a tunnel from CoA7 to the Seattle TTR, since this is the TTR it currently has one or more tunnels to, or last had a tunnel to. The TTR company's management system would have detected that the new CoA7 (or rather the NAT box's public address by which CoA7 appears to the outside world) was distant from the Seattle TTR and much closer to a TTR the company runs in Colorado.

The MN itself would not necessarily detect this, but the Seattle TTR could easily do so, by tracerouting a few hops towards the CoA7 public address and by finding that the chain of responding routers led away from Seattle and to a distant state. In order to do this, the TTR company's management system would require considerable sophistication and to be configured with relevant topological information.

The TTR company's management system would then instruct the MN to establish a 2-way tunnel to a Colorado TTR run by the same TTR company. In fact, TTRs might be run by some intermediate service company, and their capacities rented out to multiple TTR companies. It is even possible that the TTR company's TTR in Colorado is the same TTR as used by the aircraft's mobility system, but in our example, we assume these are two separate TTRs.

Once this tunnel to the Colorado TTR was established and proven to be robust, the TTR company's management system would change the mapping of the MN's micronet to that Colorado TTR.

So far, we have discussed mapping changes being prompted by a new CoA address, leading the TTR company's management system to determine that the new CoA's point of connection to the Net is far enough away from that of the current CoA to warrant a change in mapping. Now we discuss the need for the TTR company's management system to detect a changed point of connection while there is no change in CoA.

In flight, the Colorado TTR can still be used for the MN no matter which satellite ground-station and TTR the plane uses. When the uses the second GEO satellite with its Swiss ground station, the mapping of the plane's micronet is switched to a new TTR in Switzerland. This will need to be detected by the management system of the TTR company which this MN is using. That TTR company's system will then instruct the MN to

establish a link to a TTR the company runs in Switzerland. Once that tunnel is established and tested, the TTR company's management system will change the MN's micronet's mapping to that Swiss TTR.

One method of detecting the change of the aircraft's point of connection (from the TTR in Colorado to one in Switzerland) would be the MN somehow gaining link-level information from the aircraft's mobility system. However, this involves information flows which the aircraft operator doesn't necessarily have a reason to support, and would require considerable coordination of protocols, software etc.

A more robust approach would be for the TTR company's mobility system to periodically traceroute towards the MN's CoA7 address, and note any changes. Alternatively, various TTRs (or servers at TTR sites) all over the globe could send ping packets to the MN's CoA7 address, and note any changes in the timing of the responses. A European node would notice a shorter response time while a US-based node would notice a longer time when the aircraft's micronet's mapping switched from the Colorado to the Swiss TTR.

This level of probing would be onerous except when global movements of a CoA's point of connection was expected. Probing MNs needs to be done judiciously so as not to waste expensive bandwidth. It would be reasonable not to probe continually if the MN's CoA7 was an ordinary non-SPI address – on the assumption that non-SPI addresses are unlikely to involve changes in connection to the rest of the Net involving thousands of kilometres. If NAT was involved, as it is in this example, the public address of the NAT box behind which CoA7 was located would be checked instead of the CoA7 address, which is an RFC 1918 private address. However, in this example, the public address of the aircraft's NAT box is in SPI address space. Therefore, the MN's TTR company's management system should recognize the potentially mobile nature of this address, and probe the MN regularly to see if its point of interconnection to the Net has changed.

This assumes the aircraft mobility system used the one global CES system to tunnel packets to its TTRs. If it used a private mobility system and so did not use the global mapping system – that is, if the aircraft used address space which was not part of the main Ivip etc. SPI space – the TTR company's management system would need to be configured to recognize the CoA7 address (or its NAT box public address) as being part of this airline mobility system, and therefore to probe the MN's point of connection regularly.

The flow of packets in and out of NAT, tunnels and TTRs is quite complex in this example, but it can be seen how a well designed automated management system would ensure generally optimal paths, irrespective of the nature of the access network, and ideally even if that access network involved changed points of connection to the Net while the CoA remains stable. If the airplane's satellite link provided connectivity to one ground station while also connecting to new ground station, then continued connectivity could be maintained, on the end-user's stable, public, IP address, from Seattle to London and beyond.

While it may be considered overkill to maintain a single IP address for a laptop travelling internationally, and while various new protocols and application capabilities might traditionally be suggested as a better approach to mobility than maintaining a constant IP address from one month to the next, once a global ITR and TTR network is established, this TTR approach may prove to be more efficient and cost-effective than any other approach to global mobility.

7.3 Optimizing choice of TTR

While there is no absolute requirement that the MN software or the TTR management system be aware of any details of the access network, standardized protocols which enable the MN to detect conditions and changes in a mobile access network (Level 1 in the TTR model) could be used by MN software to communicate this information to the TTR management system.

The IETF DNA (Detecting Network Attachment) WG [26] is developing protocols which enable the MN to be notified of link-level events.

The TTR management system controls both Level 2 and Level 3 of the TTR architecture. Any awareness the management system could gain of the moment-to-moment vagaries of the Level one physical access network is likely to be useful in optimizing the Level 2 arrangement of which CoA and TTRs each MN should use. For instance, information on signal-strength and lower level bit error rates and congestion from the one or more access networks a MN is connected to would enable the TTR management system to choose the best of potentially several CoAs and/or TTRs to use.

7.3.1 Alternatives to traceroute

We have mentioned Traceroute – from MNs to TTRs and from TTRs to MNs – as a method by which the TTR management system can automatically discover the location of the MN's CoA(s) in its current access network(s). Traceroute may suffer from robustness problems or be prevented by ICMP filtering. Alternative methods to traceroute would be highly desirable, since determining the best TTR to use is a crucial element of a successful system.

Physical or topological proximity to a TTR – as traceroute might easily detect – is not necessarily the best criteria for deciding which TTR to tunnel to or use for traffic. Ideally, the MN would tunnel to several candidate TTRs and continually monitor round-trip packet times and packet loss rates in order that the TTR management system could choose which access network would best support the support current traffic.

Despite the general principle that most mobile users would not need to change TTRs as long as the current TTR is within about 1000km, some users who are happy to pay for more frequent mapping changes, would prefer their TTR company's management system to expend considerable resources choosing the optimal TTR, especially in unfavorable local network conditions

A MN with two or more micronets could use multiple TTRs simultaneously, using one micronet for one kind of traffic an another for traffic with different latency or reliability requirements.

7.4 Optimized TTR Tunneling Protocols

A TCP-based encrypted tunnel between MN and TTR has disadvantages when dealing with lost packets: the tunnel is blocked until retries are successful. A more sophisticated UDP-

based protocol could use QoS attributes to queue short non-delay-sensitive packets to piggyback with a VoIP packet in a single tunnel packet, and to avoid retries for VoIP packets.

Sophisticated tunnel protocols could duplicate packets over wireless links to improve robustness, or spread loads over multiple links to improve throughput. Each tunnel could handle traffic for multiple micronets, enabling great flexibility in spreading the load over multiple access networks and potentially multiple TTRs. Furthermore, these "mobility" techniques could be used with multiple DSL, cable modem and WiMax links as an inexpensive approach to multihoming a small non-mobile corporate fiber access link.

While the core-edge separation system is singular and global, and so requires TTRs to comply with its tunneling protocols, there are no such restrictions on how TTRs and MNs communicate. While IETF standards in this field would no-doubt be helpful, Level 2 of the model can be engineered in whichever way the TTR company chooses, as long as they provide appropriate software for their customer's MNs.

This flexibility, combined with the great scope for innovation in designing a good TTR management system, should enable a great deal of service innovation and competition, even if IETF standardized tunneling techniques are used between the MN and TTR.

8. CONCLUSION

We have described a promising new mobility architecture which applies equally to IPv4 and IPv6, which maintains a stable IP address or prefix for each MN, which works with all existing hosts as correspondent nodes, and which can use existing hosts as MNs, with suitable additional software. The additional software required for the MN could be added at runtime to most operating systems, and would enable all existing protocols, existing applications and the rest of the operating system to communicate with all hosts.

With a reasonably well deployed system of TTRs, the system should be capable of providing generally optimal path lengths – without using a fixed home agent or traditional Mobile IP techniques. The approach grew from a core-edge separation solution to the routing scalability problem, and it could be implemented independently of any IETF standards as the basis for a global mobility business.

Initial consideration of a core-edge separation architecture (using encapsulation, translation or forwarding) being used for mobility might lead to the impression that mapping changes would be as frequent as the MN's changes of CoA, or that connectivity depends on the rapidity with which the mapping change could be executed. Both notions are based on the erroneous assumption that the mapping system directs ITRs to tunnel packets directly to individual CoA addresses. [RW 2010-01-12: draft-meyer-lisp-mn-00 does this - the MN is its own ETR, which can't work behind NAT.]

CoA addresses are not suitable destinations for tunneling packets from ITRs - nor are they usually suitable for sending out packets with SPI source addresses.

The TTR forms a stable, typically nearby, bridge between the global ITR-ETR system, and the potentially unstable addresses and local attachment points of the MN as it connects to various access networks. The TTR is somewhat like a nearby home agent of choice, except that the MN can use multiple TTRs at once, and is directed by the TTR management system to use the closest or at least the most appropriate TTRs of the potentially thousands which are located all over the Net.

By providing a global network of strategically located TTRs, with a sophisticated management system, the TTR company can adapt to the MN gaining any kind of CoA, in any access network, and maintain generally optimal paths from all correspondent hosts, by judicious choice of TTRs for the MN to tunnel to, and then by judicious choice of which TTR at any point in time the global ITR system will tunnel packets to.

The result is that while a rapid response mapping system is highly desirable, it is not absolutely essential. Similarly, while in some extreme cases rapid changes of mapping may be required to produce the best results, for the great majority of MNs, there need be no mapping change for one month to the next, since most people do not travel distances in such times which would put them so far from their current TTR as to adversely affect performance.

9. ACKNOWLEDGEMENTS

10. REFERENCES

[1] On 2008-07-24 Jari Arkko wrote on the IRTF Routing Research Group list two messages regarding taxonomies of scalable routing proposals and a diagram. One of the branches was core-edge separation schemes. http://psg.com/lists/rrg/2008/msg01942.html , http://www.arkko.com/ietf/rrg/designspace_dataplane.jpg & http://psg.com/lists/rrg/2008/msg01943.html (retrieved 2008-08-20).

[2] Heidemann. J. et al. 2008. ISI-TR-2008-649 Census and Survey of the Visible Internet (extended). http://www.isi.edu/~johnh/PAPERS/Heidemann08a.pdf (retrieved 2008-08-20).

[3] Hain, T., "IPv4 Address Pool Projection", 2008-07-15 http://www.tndh.net/~tony/ietf/ipv4-pool-combined-view.pdf (retrieved 2008-08-20).

[4] Huston, G., "The End of the (IPv4) World is Nigher", http://www.potaroo.net/ispcol/2007-07/v4end.html (retrieved 2008-08-20).

[5] On 2008-08-07, Ricardo V. Oliveira of the iPlane project http://iplane.cs.washington.edu wrote on the RRG list that 123k was a lower bound for the number of DFZ routers. http://www.ops.ietf.org/lists/rrg/2007/msg00255.html (retrieved 2008-08-20).

[6] Huston. G. BGP Routing Table Analysis Reports. 2008. http://bgp.potaroo.net (retrieved 2008-08-20).

[7] RRG mailing list 2008 archives: http://psg.com/lists/rrg/2008/maillist.html.

[8] Hinden, Robert M., IP Encaps, June 1996, RFC 1955 http://tools.ietf.org/html/rfc1955 .

[9] IRTF Routing Research Group Charter (early 2007) http://www.irtf.org/charter?gtype=rg&group=rrg (retrieved 2008-08-20).

[10] Internet Architecture Board 2006. Routing and Addressing Workshop reading list and meeting materials. http://www.iab.org/about/workshops/routingandaddressing/ (retrieved 2008-08-20).

[11] Meyer, D., Zhang, L., Fall, K. 2007. Report from the IAB Workshop on Routing and Addressing. RFC 4984. http://tools.ietf.org/html/rfc4984 Workshop materials: http://www.iab.org/about/workshops/routingandaddressing/ .

[12] Farinacci, D., Fuller, V., Meyer, D. 2008-04. LISP Alternative Topology (LISP-ALT). http://tools.ietf.org/html/draft-fuller-lisp-alt-02 (work in progress).

[13] Lear, E. 2008-04. NERD: A Not-so-novel EID to RLOC Database. http://tools.ietf.org/html/draft-lear-lisp-nerd-04 (work in progress).

[14] Jen, D., Meisel. M., Massey. D., Wang. L., Zhang. B., Zhang. L. 2007. APT: A Practical Transit Mapping Service. http://tools.ietf.org/html/draft-jen-apt-01 (work in progress).

[15] Whittle, R. 2008. Ivip (Internet Vastly Improved Plumbing) Architecture. http://www.firstpr.com.au/ip/ivip/Ivip-summary.pdf (retrieved 2008-08-20).

[16] Whittle, R. 2008. Ivip Mapping Database Fast Push. http://tools.ietf.org/html/_draft-whittle-ivip-db-fast-push (work in progress).

[17] Herrin, W. 2007. Tunneling Route Reduction Protocol (TRRP). http://bill.herrin.us/network/trrp.html (retrieved 2008-08-20).

[18] Vogt, C. 2008-07. Six/One Router A Scalable and Backwards-Compatible Solution for Provider-Independent Addressing and IPv4/IPv6 Interworking. http://users.piuha.net/chvogt/pub/2008/vogt-2008-six-one-router.pdf (retrieved 2008-08-20).

[19] Mogul, J. Deering, S. 1990 RFC 1191. http://tools.ietf.org/html/rfc1191 .

[20] Whittle, R. 2008-04. MTU, fragmentation and Path MTU Discovery. http://www.firstpr.com.au/ip/ivip/pmtud-frag/ (retrieved 2008-08-20).

[21] Whittle, R. 2008-08. ETR Address Forwarding (EAF). http://tools.ietf.org/html/draft-whittle-ivip4-etr-addr-forw (work in progress).

[22] Whittle, R. 2008-08 Prefix Label Forwarding (PLF). http://www.firstpr.com.au/ip/ivip/ivip6/ (retrieved 2008-08-20).

[23] US Bureau of Transport Statistics. http://www.transtats.bts.gov .

[24] Barabasi, A-L., Hidalgo C, A. and Gonzalez, M. C. Understanding individual human mobility patterns Nature 453, 779-782 (5 June 2008) http://www.nd.edu/~chidalgo/Papers/nature2008/nature08images.htm .

[25] Arbarnel, B. 2004. Connexion Global Network Mobility. http://www.nanog.org/mtg-0405/pdf/abarbanel.pdf (retrieved 2008-08-20).

[26] IETF DNA WG. http://tools.ietf.org/wg/dna/ .

11. Appendix 1: Continuous Connectivity for Aircraft

Our Advanced Techniques example above illustrates how the TTR mobility approach would be capable of providing continuous connectivity for a laptop, from one country to another, via a variety of local access networks – if the aircraft's own access network was capable of providing continuous connectivity to the Net as it travelled between countries, over the Pacific Ocean etc. Here we discuss some L1 aspects of Internet connectivity for passenger airliners, which illustrates some of the challenges to achieving continual connectivity for passenger aircraft via any mobility architecture.

There are three basic approaches to providing 2-way data communications for aircraft, for Internet access or other purposes. These could be combined – for instance using ground stations where available over land to reducing reliance on satellites.

One approach is to use geostationary (GEO) satellites, 35,800km above the equator, to provide a link to a particular ground station for each such satellite. GEO satellites are physically distant and are limited in number. The total bandwidth available via this approach is limited and expensive. The full round trip for all communications involves an additional 477ms latency. A GEO satellite covering a large expanse of the Earth's surface, such as the Atlantic or Pacific ocean, also faces challenges with sufficient transmit energy and receive sensitivity, considering the high data rates a single passenger airliner might require, and the small size of the antenna which can be fitted safely to any passenger aircraft. One solution to this is a large phased array antenna in the satellite, with multiple synthetic beams tracking each aircraft it is currently communicating with. However the complexity, cost and weight of such antenna systems is challenging for any satellite application.

Another approach is to use MEO (Medium Earth Orbiting) satellites, which can be more numerous, closer, and therefore have greater total bandwidths and lower latencies. LEO (Low Earth Orbiting) satellites might also be used, but these travel even faster across the sky and remain in view for only a few minutes at a time. LEOs would poses still greater challenges for rapid steering of the aircraft's transponder beam(s).

A third approach is to use a series of ground stations. This can provide low latency and high bandwidth, and can cope with high aircraft densities better than a satellite-only approach.

Retaining continuous connectivity for long flights which require switching to a new satellite would only be possible if the aircraft's transponder can communicate with two or more satellites simultaneously. This is not possible where a single steerable parabolic dish antenna is used, for instance in a radome installed in the top of the fuselage.

Unfortunately the only method by which multiple satellites can be reached simultaneously has considerable weight and cost problems: phased array antennae on the top and sides of the fuselage, or on the bottom and sides for ground stations. Radome-based steerable dishes can more easily reach satellites close to the horizon, which is more difficult for phased-array antennae of any reasonable size, particularly in the forward and aft directions.