Transport for Carrier Grade Internet

Raimo Kantola, Marko Luoma, and Olli-Pekka Lamminen, *Member IEEE,*
Department of Communications and Networking
Helsinki University of Technology
P.O. Box 3000, 02015 TKK,
Finland

***Abstract*— This paper discusses the problems of the modern Internet and prospects for its future development. The paper proposes to gradually phase off IP as the core packet networking protocol, move from fundamentally flat addressing and routing with numerous add-on improvements to recursive addressing and routing by developing Ethernet technology. The goal is to meet user network connectivity needs on top of a Carrier Grade transport network. The paper shows how the idea is being implemented in an Ethernet based Internet architecture and points out several deployment scenarios.**

***Index Terms*— carrier grade transport, Ethernet, IP, trust-to-trust.**

I. INTRODUCTION

There were about 1.6 Billion Internet users in March 2009. The number has grown by about 340% over the past 8 years [1]. The number of wireless subscriptions was around 3.8 Billion in September 2008 showing an even higher growth rate of tripling in 5 years [2]. The growth of wireless has been mainly based on circuit switched voice but packet data networking has been penetrating wireless networks over this decade and lately all new mobile network technologies are optimized for packet networking. The number of fixed broadband subscriptions was around 430M showing a growth rate of doubling in about 5 years [3]. The Internet traffic has grown by 50 to 100% per annum over the past few years and is predicted by Cisco to continue doubling in two years [4].

The nature of Internet usage is also changing. For businesses of any size high quality Internet access and visibility are a must. The Internet has changed our everyday lives irrevocably. All information businesses and information goods delivery, be that business, consumer or entertainment related, is moving to the Internet. We have only seen the beginning of the impact of the Internet on media industries that have long been fighting a loosing battle to defend their industrial age copyright earning models.

This development has created and continues to create a serious pressure on the Internet architecture that originates from the 1960's. So far, the architecture has been surprisingly robust. It has been possible to incrementally stretch the original architecture by numerous add-on solutions to meet new scalability and functional requirements. These add-on solutions include classless inter domain routing, private addressing and network address translators, security protocols and firewalls, fitting IP to run over all new transport technologies, adding label switching to operator networks as a broad range connectivity services platform and means for traffic engineering etc.

A symptom of the pressure is the expected exhaustion of IPv4 addresses in 2012 [5]. Another symptom of the pressure is that actually in modern ISP networks there is very little if anything left of the original IP network principles of end-to-end and IP-over-everything by Dave Clark and Vinston Cerf. In 2007, Dave Clark himself recognized the existence of all kinds of middle boxes (NATs, Firewalls and Application level gateways) that break the end-to-end principle by introducing the Trust-to-trust principle. He formulated the new principle as follows: *"The function in question can completely and correctly be implemented only with the knowledge and help of the application standing at points where it can be trusted to perform its job properly"* [6].

IPv6 has long been designed and projected as the solution for the future Internet. However, very little of the users, traffic or Internet services have started using IPv6 so far. In this paper, we argue that IPv4 or IPv6 do not meet the current networking requirements nor does IPv6 solve the problems inherent in IPv4 networks today. Disbelief in IPv6 is visible in the research community. People are coming up with new ideas but so far the main focus seems to be on approaches that can be characterized as application oriented. Examples are the Data Oriented Network Architecture by ICSI in USA and Distributed Web by HIIT in Finland. We argue that, instead, a new networking paradigm based on recursive addressing and routing is needed. This means that we propose a way of implementing trust-to-trust as a principle rather than an afterthought. We show how this idea can be implemented by leveraging the Ethernet technology. The idea is to solve the problem bottom-up in the protocol stack. Actually, the IPv4 research community is grouping in the same direction with the idea of Address Indirection, a proposed solution to the BGP scalability problems. Thus, the idea of recursive routing can be engineered to work with different forwarding protocols.

Manuscript received July 9, 2009. This work was partly supported by the EU funded, FP7 strep project ETNA Ethernet Transport Networks, Architectures of Networking (project nr 215462) [13].

All authors are with the Helsinki University of Technology, Finland (corresponding author phone: +358-40-750 1636; fax: +358-9-451 2474; e-mail: Raimo.Kantola@ tkk.fi).

978-1-4244-5626-0/09/$26.00 ©2009 IEEE

The rest of the paper is organized as follows. Section II depicts the mismatch between current network requirements and IP. Section III summarizes key networking challenges over the next 10 to 15 years. Section IV briefly discusses what is happening under IP in the protocol stack. Section V suggests the new networking principles that meet the current network requirements and future challenges. Section VI depicts an architecture following the new principles that we are implementing in ETNA [13] covering the foundation of the overall vision and discusses how the architecture can be widely and smoothly deployed. Section VII analyzes the benefits from the new principles. Finally, Section VIII gives some conclusions.

II. What is Wrong with IP

Let us recall that the original IP principle called for a network that would not keep connection state or flow state in network nodes. The routing system was designed to work on the background of user traffic and create the forwarding entries in the routers and those were supposed to be all that was needed for end to end packet delivery.

A. Middle boxes

Internet is a network of networks made of the consumer Internet and corporate IP networks including Internet Service Provider's own IP networks. Instead of direct visibility of any user and node by any other user and node based on global IP addressing as assumed by the principles of end-to-end and IP-over-everything, this network of networks is segmented into islands isolated by NATs, Firewalls, Application level gateways and Session Border Controllers. A packet sent by a user from a source address to another user indicated by a destination address in the IP packet header traverses several legs of a connection. On network boundaries the packet moves up in the protocol stack, is re-addressed and re-encapsulated when it is pushed to the next leg of the connection.

Middle boxes introduce connection state into the network. This state is mostly managed by complex configuration information and by implicit signaling that is embedded into the client server request-response message pattern. Complex processing of each packet adds cost and delay to packet transport. Modification of packets in the middle boxes creates numerous problems to protocols that were designed with the assumption of end-to-end. The recommended solutions are application specific and scale poorly for mobile use [15].

B. Scalability of routing

IP routing is divided into interior and exterior routing. Exterior routing uses the BGP protocol that needs to create routes for all address prefixes from the global address space for the packets that need to follow different paths across core Internet service provider networks. Address blocks to ISP and corporate networks are allocated from the same global address space. In case of selective multi-homing routing changes in corporate networks are propagated to core ISP networks. Due to the fact that more and more companies see the Internet as part of their critical business infrastructure, demand for multi-homing of corporate networks to several ISP networks is growing. With the current architecture, this leads to non-default networks having to make the selection between alternative routes to multi-homed destinations and also to the growth of the core routing tables when more entries are needed for long prefixes that can not be aggregated. Large tables are hard to maintain in a global network and take a lot of expensive and power hungry fast memory. This adds cost and makes further scaling of the Internet challenging. Also, convergence of the global routing system has become ever more difficult. Achieving a stable routing state after a major failure can take days rather than seconds or minutes.

An example of operator's own corporate IP networks is the 3G access network architecture. The solution has two IP networks one over the other. The lower network is used for packet transport by the mobile operator. The resulting complex protocol stacks are one of the reasons of high cost of packet traffic to and from mobile devices.

C. Lack of mobility support

IPv4 does not natively support mobility. This is related to the dual semantics of IP addresses that are both identifiers and network locators. Considering that the role of wireless communication is increasing and soon most Internet users will be wireless, a core networking protocol that does not support mobility will have a problem.

D. Virtualization of infrastructure

Best effort IP connectivity is a commodity service. Most ISPs have hard time making any money from the consumer Internet. Instead, they survive by providing corporate connectivity services based on the same infrastructure that is used to provide services to consumers. The physical network infrastructure made of fibers, radio links and sites housing different kinds of network nodes is virtualized and allocated to the consumer Internet and corporate networks using different kinds of multiplexing and virtual private network technologies.

Routing vendors and operators introduced MPLS as an add-on to ISP networks in order to facilitate traffic engineering and virtualization of the network infrastructure. MPLS is now used as a general purpose platform for connectivity services provisioning in ISP networks. Often this comes down to emulating leased line communication, Ethernet virtual LANs or private IP routing over the IP/MPLS core network. Scaling of MPLS services is improved by using several labels in a single packet.

MPLS uses locally significant labels and label swapping for forwarding. This means that MPLS nodes have label path specific state that is managed either by a management system or a combination of routing and signaling protocols. Generalized MPLS applied to Ethernet Label Switching (GELS) uses MPLS for establishing Ethernet label switched paths. GELS replaces MAC learning by a management system for populating forwarding entries. This approach is limited to point-to-

point intra-domain services.

Managing MPLS networks is costly and locating network failures is difficult. A fully fledged OAM is still missing. On this point MPLS as a transport system lags behind SDH and SDHng. Tracing faults and correlation of faults to individual connectivity services in MPLS domain is hard. One reason is that core nodes only see labels but routing information behind the label generation may be hidden from the control logic of the core router. Even with the correlation of internal routing information and label allocation actual impact to the connectivity services is hidden. This information is only visible at the edges of the provider network where the connectivity service is realized as a virtualization of the routing platform. However, at these points hardly any other OAM functionality than LSP ping and traceroute exists. Therefore, MPLS can be seen as a yet another add-on next to IP. The purpose of MPLS has changed several times during its design and use. This is no surprise because IP principles give very little guidance on how add-ons should be designed.

E. Unwanted traffic

The Internet users suffer from unwanted traffic. The assumption the Internet was built on was that a sender would not send if the receiver would not want to receive. This assumption has long been abandoned by connectivity and information providers. Manifestations of this are the constant increase of advertisements and pop-ups on web pages. Also the operation of current email system suffers from this design principle, where the network is using all of its resources in support for the sender – no matter whether the communication is based on legitimate goals or not. As a result, the cost of reception is high while the marginal cost of sending approaches zero. This seems to be mainly an economic problem rather than a technical one. Lack of trust has an impact on behavior: according to OECD [10], 20 to 50% of users do not buy on-line because of security or trust concerns in different countries.

F. Summary

To summarize, IP assumes visibility of all-to-all but ISPs and corporations spend a lot of money for hiding their networks from the outside world. Hiding of networks takes place due to legitimate business reasons. There is no turning back to the idealized past when we could assume benevolent users and a common good that every user would be willing to put above his or her own interests. IP provides one low level of trust while differentiation in terms of trust between users is implemented by costly add-on techniques. IP itself is not the money maker to ISPs, rather the add-ons that largely can be seen as misuse of IP technology are critical for earning money.

III. Networking Challenges

We argue that the challenges relate to two factors. One is providing and managing trust and the second is scalability. The latter can be further broken down into scalability in terms of (a) number of users, applications and network nodes, (b) scaling the core network by a factor of at least 100 and (c) providing cost efficient mobile access for packet switched services.

Trust requires efficient techniques in terms of both capital and operational costs for network hiding and virtualization. Brokering trust is also a business opportunity for operators.

We expect that the next couple of Billion Internet users will mainly be wireless and will come on-line in countries that have poor fixed network coverage. The wireless access technologies will use variants of 802.xx, the current mobile radio and new types of radio technologies such as LTE that will be connected to the core packet network. Low ARPU and the need to carry all kinds of traffic over wireless packet access will require simplifying the access architecture.

Traffic growth will come from new users, higher access speeds and new services especially all kinds of entertainment and media that will be carried over the network.

IV. Development of Transport Technology

Originally IP traffic was carried over 2Mbps TDM circuits allocated from the PSTN infrastructure. The first packet based transport protocol that was widely deployed under IP was Frame Relay. Frame Relay scaled up to some tens of Mbps. Telephony operators and vendors created ATM to become the uniform and ubiquitous broadband network technology. ATM was indeed used to scale up the speeds in packet data networks but IP remained as the core unifying network protocol while ATM provided transport pipes for IP from the SDH based transport infrastructure typically at the speed of 622Mbps (STM-4). The next architecture used in ISP networks was IP over SDH (Packets over SONET – PoS). PoS scaled the speeds to the next level by leveraging STM-16 (aka 2.4Gbps) circuits for IP traffic. Lately these links are upgraded to STM-64 (aka 9.9gbps). SDH is still playing a major role in packet transport as a means to frame information on the links of modern DWDM /CWDM networks providing speed of tens or even hundreds of Gbps over the same fiber infrastructure. However, this role is diminishing with new framing types like GFP and OTN. The role of MPLS in these networks is twofold. One is to provide separation of connectivity and routing and the other is to make efficient aggregation and multiplexing of low speed connectivity services to high speed core transport technology. We illustrate in Figure 1, how major steps in scaling link speeds up have been followed by changes in the transport architecture.

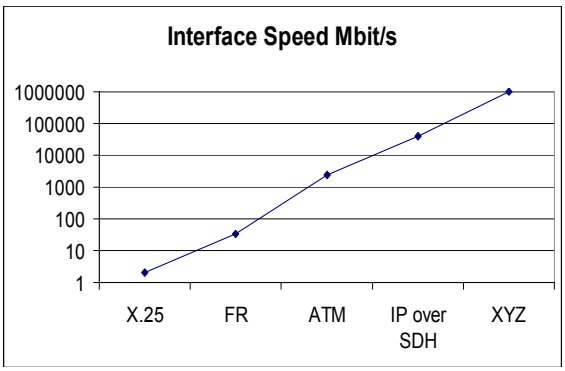

Figure 1. Generations of Internet transport

Figure 1 poses the question which transport architecture will be used when we next scale the speeds up by a factor of 10 to 100. Currently, the best that is commercially available in core routers electronically is 40G link speeds. Also, this is the upper limit for the current SDH technology. Technology for 100G Ethernet [7] is on the drawing boards and is expected to become standardized in 2010 or 2011. Soon after, we expect that link speeds with electronic processing approaching 1Te will be within reach by bundling several 100G links into one.

A. Carrier grade

Carrier grade, the current buzzword, means that operators expect their transport network to be manageable, exhibit predictable behavior and provide highly resilient connectivity.

Several years ago, packet data traffic took over circuit traffic in operator networks. To optimize an operator network for data, a native carrier grade packet transport technology that could be used to carry all types of traffic is needed. *Carrier grade* means that the operator can with fine granularity control which traffic is carried from where to where. Unless allowed explicitly, all traffic is destroyed. In addition, carrier grade includes requirements for resiliency, protected connections, network virtualization and overall predictable behavior. Verification and traceability of what is being provided to a customer are also important. This calls for a fully fledged OAM. It is obvious that IP is not a carrier grade technology while SDH is. IP over MPLS with traffic engineering is only aspiring to become carrier grade. For this purpose for example the MPLS OAM capabilities would need to be improved. Nevertheless, IP/MPLS footprint is expanding to access networks. The typical service provided is Ethernet over MPLS [11]. This particular use of IP/MPLS leads to a protocol stack that seems to be far from optimal.

A recent development in the area of transport for operator networks is the introduction of 10GbE and Metro Ethernet in metropolitan and core networks. In such networks the SDH layer as a transport network technology is not needed. One should also note that most if not all new radio access technologies are 802.x based or at least friendly and that the 100G Ethernet will further increase the Ethernet footprint in operator networks. Some vendors are pushing Ethernet as a carrier grade native packet transport technology to operator networks. An early sign of this development is Provider Backbone Bridging and its extensions for traffic engineering. Also link, domain and end-to-end OAM for operator Ethernet networks has been specified and is available from some vendors.

Looking at the trends in packet transport, it seems that there is very little interest in scaling SDH over 40G while it seems that scaling electronic link speeds up cost efficiently will be best done using a packet technology. The natural choice is to develop 802.x technologies further. Scaling transport networks calls for more manageable transport network services. In this respect, a major limitation of SDH is that the frame structure does not contain network addresses.

V. New Principles

IPv6 does not address the needs for network hiding and virtualization. It does not have a widely accepted and scalable solution for multi-homing. We argue that this means that IPv6 does not address the legitimate needs of operators and corporations in trust and resiliency. Moreover, with IPv6 we may end up in the same scalability problems in routing that are present in IPv4 networks.

A. Addressing and network location

A premise in the design of IPv6 was that a huge address space that makes a huge number of devices and nodes visible to each other in the network is a blessing and that it will be the responsibility of the routing system to implement that visibility. We argue that this is a wrong premise.

The alternative for addressing is a format that we can present as $u@e$ where u is a locally unique address of a user or a server in domain U and e is a globally unique routing address of an edge node owned by an operator. So, u is an address in one trust domain and it is chained to another address e that is applicable in another trust domain. This addressing logic is quite similar to IPX addressing. The operators are responsible for address allocation for their nodes. With this premise a user can be reachable in $u@e_1$ ... $u@e_n$. It will be the responsibility of the routing system to carry packets from any e_1 to any other e_m. Location of routing destinations has three steps: (1) directory search, (2) connection state lookup on trust boundary and (3) routing. On a given unique name, directory search routed through an edge node will return u_{dest} to the requester u_{source} and cache connection state $u_{dest} \rightarrow e_{dest}$ at e_{source}. At the reception of a packet from a remote source e_{source}, e_{dest} will cache connection state $u_{source} \rightarrow e_{source}$ using information in the packet.

In this mechanism, user devices, servers or even corporate network nodes never see any of the routing addresses used by operator network nodes and as a result, operator networks are effectively isolated from corporate networks.

Multi-homing is implemented and mobility is facilitated by a directory search that caches at e_{source}: $u_{dest} \rightarrow e_{dest1} \mid \ldots \mid e_{destN}$. In addition, the destination edge node will have to make a directory search at some point to expand the connection state: u_{source}

$\rightarrow e_{source}$ to $u_{source} \rightarrow e_{source1} \mid \ldots \mid e_{sourceN}$. Any of the edge nodes can use their own algorithms to make the selection between the different edge nodes through which the remote party is reachable. Even balancing load among the multiple paths is easy. It is possible to have an OAM protocol for monitoring the reachability of all remote edge nodes. To fully support mobility, it must be possible to update user location into the directory in real-time.

A difference with IP is that instead of DNS and routing, location of destinations uses three mechanisms, i.e. directory search, connection state on trust boundaries and routing. Instead of breaking a hard problem into two, we break it into three. The result is a significant increase in scalability. Multi-homing has no impact on the routing system. It is taken care of by the directory search and caching of connection state. Like in case of NATs, connection state on boundary nodes is managed by implicit signaling (because it scales best for short flows). Contrary to NATs that are an after-thought in the current Internet, Provider Edge nodes in our solution are assumed to be a legitimate and necessary part of the architecture from the start.

B. Forwarding

An ingress edge node will encapsulate a packet for transport over the operator domain. An egress edge node will decapsulate the packet and strip off the operator network headers including addresses that are used in the operator domain.

Forwarding in the network can be based on one or more VLAN tags, source MAC or destination MAC or any combination thereof. VLAN tags can be assigned global, domain wide or even link local significance. Scaling of MAC based forwarding could be significantly improved by using operator allocated node addresses to find a filtering (forwarding) entry.

A natural model of forwarding uses one or more VLAN tags to identify a forwarding tree that either spans the whole or most of the physical network for one or more public services or a subset of the physical network for a corporate network. Having identified the tree, a node uses the destination MAC to forward the packet onto a branch of the tree. Either the service VLAN tag or source MAC address can be used to avoid downstream merging of traffic from different sources onto the same outgoing link. This may be used for traffic engineering purposes.

Alternative models are also possible. For example, link local VLAN tags lead to a label swapping style forwarding.

Avoiding forwarding loops is necessary. One way to solve the problem is to introduce a TTL or hop count ether type and field into all frames traversing routed Ethernet networks. Other options are explored under the heading of Provider Link State Bridging that makes use of the well known IS-IS routing protocol.

C. Network virtualization

The common network infrastructure of an operator needs to be split to many logical networks such as the consumer Internet, different corporate networks and the operators' own corporate networks. Among the last ones are different service networks.

Virtualization of an operator network is achieved by marking users as $u \in v$ at edge nodes and by modifying the search and caching of connection state mechanisms. First, a directory search will cache an entry at e_{source}: $u_{dest} \rightarrow e_{dest1} : v_a \mid \ldots \mid e_{destN} : v_z$. Second, assuming multi-homing of the source, the destination edge node will make a directory search at some point to expand the connection state entry to $u_{source} \rightarrow e_{source1} : v_a \mid \ldots \mid e_{sourceN} : v_y$.

The virtualization of corporate networks is transparent to the operator networks. Forwarding is modified accordingly. For traceability, each packet will have to carry the identity of the virtual network it belongs to.

The fact that Ethernet frames can carry one or many stacked VLAN tags in the header is a great strength of the technology. Many developments, such as Multiple Registration Protocol, are under way to further leverage this capability. The technology for the future infrastructure should include inherent support for VLANs be they for the needs of the operators or corporations. This can not be achieved by IP/MPLS because IP is essentially a point-to-point technology.

D. Control and Data Plane separation

The IP network routing system is vulnerable to attacks and overload because IP does not make a separation between the control and the data planes. The situation is easy to remedy by running the operators control plane protocols in one or several virtual networks created by the operator for its own needs. This will enhance robustness.

E. Control Plane

The protocol for forwarding is independent and secondary in importance to the structure of the control plane.

Populating of forwarding entries in the data plane can be based on autonomous learning in network devices like in PBB, a management system like in PBB-TE, implicit signaling like in NATs or a distributed control plane like in IP networks or a combination thereof. Packets in an end-to-end Ethernet network carry a string of addresses that point to a unique location. A combination of two MAC addresses and a VLAN tag can even point to a unique path from a device to another. Like addressing, in our solution, routing is recursive. Each domain runs intra-domain routing. Edge nodes of a domain map incoming flows to tunnels based on inter-domain routing, implicit signaling or configuration maintained by a management system.

We argue that a control plane is needed that would make full use of the rich capabilities already specified for Ethernet such as VLANs and MAC-in-MAC. Support for multicast will also be important because of TV distribution and as a service enabler. Label switching at the moment has only limited support for multicast in form of point-to-multipoint LSPs. It seems that the control plane can be quite generic and can be adapted

to work with different forwarding structures and network policies.

F. Network hiding and interconnection

Isolation of the corporate networks from operator networks is achieved by the previous mechanisms. Similarly, operator networks are also invisible (like ether that magically moves information from one place to another) to users.

For routing among their networks operators use globally unique addresses (in the union of all operator domains) for their nodes. Like in IP networks, the routing, broken down into interior and exterior routing will take care of routing packets from ingress to egress through as many operator networks as it takes. This approach seems most suitable for the consumer Internet.

Some virtual network identifiers *v* may have global significance. It would, for example, seem reasonable to allocate a globally unique code point for the consumer Internet. Virtual network identifiers allocated to corporations may need to be mapped to new values on operator interconnection points. This is a simple renaming or label swapping operation. The renaming state needs to be maintained by a management system or by inter-domain routing.

G. Mobility

User mobility requires elaborate trust services such as operator assured identities and registration of users at edge nodes. Existing mechanisms such as SIM/USIM –cards, Home Subscriber Servers and e.g. the Multiple Registration Protocol can be fitted into our solution to provide operator assured identities.

User registry at edge nodes is a fourth mechanism in addition to the ones that were introduced in Section V (A). The registry serves a similar function to Visitor Location Registry in GSM/3G or the Foreign Agent in Mobile IP.

The registry will be helpful also for operator controls and assuring reachability of any users.

H. Carrier grade control

The solution will need to provide means for the allocation of network capacity to different virtual networks. This requires a resource management and traffic engineering solution at the transport level. The solution can leverage the connection state that is needed in Edge nodes.

The registry and the ingress node will have algorithms for operator control of traffic generated by users.

A combination of a registration protocol such as MRP and an Authentication, Authorization and Accounting (AAA) system will make possible auto-configuration of VLANs in a domain. This may be extended to a global network with an appropriate inter-connection method.

Provisioning of corporate VLANs across multiple operator networks becomes possible if problem resolution can be supported by an OAM solution. One has been already specified for Ethernet in Y.1731 by ITU-T. OAM makes it possible to monitor each link and each multi-hop connection, check availability of a connection and measure parameters such as delay and capacity over connections. With these tools tracing a problem to a link or node and linking network problems to services is straightforward.

The combination of AAA and the proposed location methods make it possible to trace spam back to its source given a set of appropriate reactive protocols. Having this technology in the networks, fighting SPAM becomes a matter of effective inter-operator agreements.

VI. IMPLEMENTATION

We have implemented a subset of the principles presented in Section V in a prototype [8]. The first prototype was mainly used for the purpose of verifying the forwarding mechanism directly over Ethernet without using IP at all. We are now implementing a significantly more comprehensive prototype in the ETNA project funded by the EU in FP7. The next section will outline the design we have adopted in ETNA. It covers the foundation layers of the overall vision proposed in Section V. More details are provided in [9, 14].

A. Layered Architecture

The network functionality is split into three layers as depicted in Figure 2. The bottom layer is the *transport layer*. The middle layer builds *transport services* such as leased lines using resources of the underlying layer. More elaborate services such as interconnection and mobility are implemented on the *network services layer*.

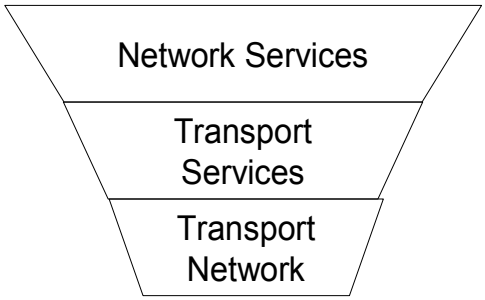

Figure 2. Layered network architecture

The Transport layer takes care of transmission of packets between network end points. A network endpoint is the termination of a core interface at an Edge Node (EN). Transport layer supports traffic flows from Edge Nodes through Core Nodes (CN) to other Edge nodes. Transport layer functionality can be broken down to forwarding plane and control plane and it also has its own management functions. Note that no customers are connected to this layer. Core nodes and core interfacing functions in Edge nodes reside on this layer.

In our implementation we use IS-IS modified for 802.1 for network discovery and a centralized routing server on the Transport layer. This routing system is responsible for setting up and maintaining tunnels from all edge nodes to all other

edge nodes in a domain. A centralized routing system was chosen to facilitate protection and traffic engineering rather than just shortest path routing. Also, one of the key aspects of Carrier Grade is predictability. By not allowing network to make its dynamic decisions on routing, we can predict the performance and quality of network connections.

Transport Services layer – Contains a set of services required to support customers of the transport network. These are the conventional transport services, such as point-to-point connections (e.g. leased lines) and more advanced packet transport services like virtual forwarding (e.g. private LAN services). The Transport Service layer builds its services on top of primitives provided by the transport layer. These primitives are different types of forwarding path bindings (e.g. tunnels). Services make use of additional logic which is generic for future transport networks, like customer dependent forwarding tables and encapsulations. This layer contains generic edge functions in Edge Nodes.

Network Services layer (or Value Added Services) provides services that are dependent on the network functionalities (like routing) or primitives (like identities) or are used to enhance trust. It also makes it possible to build service logics that are not generic in nature and are therefore not efficiently implemented in the Transport Service layer. Network Services are realized by Service Nodes (SN). An SN is a modular processing unit connected to an edge node. Modularity provides means to build variable service logics based on demand. A Service Node realises its offerings either through manipulation of forwarding logic on the Transport Service layer or by processing and manipulating the traffic itself. These functionalities are necessary when support for mobility and/or inter-connection with other network are pursued.

There may also be other kinds of value added services which are native for future broadband networks. These services may include for example network assisted media distribution and support for session control.

The reason for dividing functionalities of the network into layers is that these layers are distinctly different and form a clear separation between the services of a provider and its customers. The Transport layer is hidden from the user of the transport network. It is the infrastructure that the provider has for delivering packets efficiently between access points of the transport service. In this respect, the same service may be delivered with a range of different implementations of transport layers and thus allows migration and evolution without influence on the customers. We argue that evolution of network technologies will follow certain rules in market economy. One of these rules is that existing technology is pushed to its limits before total replacement is done. This means that current SDH networks are likely to be upgraded to SDHng networks before leap to pure Ethernet networks. Similarly current MPLS networks will evolve towards MPLS-TP which inherently removes the complex IP control plane and replaces it with provider dependent OSS and control. Ultimately this will evolve towards ubiquitous packet network – namely Ethernet.

B. Deployment

Our three layer network can carry IP packets [15] or packets formatted according to some other protocol the users will choose to use. Synchronous stream oriented services can also be emulated on the architecture similarly as it is done over other packet transport networks. Since most users are connected to their own Ethernet networks, actually IP will not always be a necessary layer in the stack for end to end communication. It will be possible to support end to end Ethernet services in parallel to supporting the other likely modes of communication.

From the point of view of the end-to-end Ethernet, IP is an identity protocol. No IP routing functionality is required for end to end delivery across the proposed network. The proposed network is still an Ethernet and it smoothly carries IP packets as before. This means that corporations and operators can make independent decisions on deploying the technology. If we turn IP from a routable protocol into an identity protocol, the problem of IPv4 address exhaustion goes away. Effectively, all IP addresses become private. We are discussing this issue more thoroughly in another paper.

It is feasible to build Ethernet-to-IP gateways for supporting hosts that do not themselves run the IP protocol. The allocation of IP addresses is then moved from such hosts to the gateways. In terms of impact on connectivity, an Ethernet to IP gateway is similar to a Network Address Translator on the border of two IP networks.

In parallel to IP, other identity protocols can be deployed over the end to end Ethernet. Some services (e.g. voice) do not require an identity protocol at all and can be transported over the proposed network directly provided the media flows are managed properly with signaling. If we allow end to end transport without IP, a research area for new transport protocols competing with TCP will open up.

Legacy networks such as IP/MPLS can be used as transport layer networks in the proposed architecture. Also, a transport profile for MPLS has been proposed. The goal of our architecture is to provide compatible functionality with MPLS-TP but in a manner void of IP legacy.

The proposed solution allows implementing a simpler and more cost efficient mobile access network than the one used today in 3G and GPRS networks.

VII. BENEFITS

802.x networks today use MAC -addresses that are 48, 60 or 64 bits in length. These addresses are not globally unique. In our solution, unique identification of users and servers is achieved by a string of addresses. Each address refers to a domain with clear engineering rules for address allocation. Thus, an address space of practically any size can be built.

Conventional manufacturer dependent MAC-addresses can not be aggregated and are therefore difficult to use in the global network. Therefore, we propose to use topology aligned

NSAP-addresses as routing identities on the provider networks. NSAP-addresses map to nodal addresses of provider infrastructure devices. These topological MAC-addresses are the system identifier parts of the NSAP address regime. With this method forwarding table entries are easy to aggregate forming a hierarchical network view on a global scale. Allocation of NSAP prefixes and MAC trees need to be controlled globally in order to avoid collisions and depletion.

A premise of the solution is that MAC and NSAP addresses are for network location. This immediately solves the problem of dual semantics of addresses in IP networks. By using four mechanisms (directory services, caching of connection state on trust boundaries, routing and the registry) instead of the two in IP networks (DNS and routing), we increase network scalability significantly.

All well understood connectivity services such as point-to-point connections, virtual networks, multi-homing and mobility are implemented in a uniform manner. Protection mechanisms (1+1 and 1:1) are implemented by the routing and forwarding systems.

By breaking down the functionality into several layers, we achieve modularity leading to simple implementation and low cost.

VIII. Conclusions

We have proposed clear networking principles for future packet networks based on recursive addressing and routing. We are implementing the principles using Ethernet as the forwarding protocol. The control plane is rather generic and could be adapted to different forwarding protocols.

An alternative to end-to-end Ethernet is IP over IP. This is pursued by some people. An example is the proposed Address Indirection approach to solving the BGP scalability problem. We chose to simplify the target architecture and the resulting protocol stacks for the sake of low cost. In our opinion the experience of two IP stacks one on top of the other in the 3G access network shows that complex stacks lead to long packet delay and high cost. By having the IP layer in the network, implementing inherent infrastructure support for virtual LANs would become impossible. By assuming Ethernet as the forwarding plane, we are free of the IP legacy in creating new networking principles and solutions. There is a natural fit of recursive routing with Ethernet. Compared to MPLS shim and IP, Ethernet natively supports traceable network virtualization which we see as a great advantage.

Another alternative approach of leveraging Ethernet networking capabilities is the TRILL architecture and the RBridges protocol [12]. We are pursuing this avenue of development in the Finnish subproject of the 100GET Celtic effort [7]. The solution we have suggested here is more ambitious. RBridges carries a lot of Ethernet legacy and assumes that IP is used for wide area. Our solution shows a smooth deployment path without sacrificing generality and pursues the final goal of solving most of the problems that are inherent in today's Internet in addition to being a carrier grade transport solution. The minimum one can conclude is that conducting research into enhancing Ethernet technology is very timely and that we should look at all reasonable alternative ways forward.

We have partially verified the forwarding solution over Ethernet by a prototype. We are building a more comprehensive prototype at the moment covering mainly the transport layer and the transport services layer. It is presented in more detail than was possible here in another paper [9].

We believe that the proposed architecture opens a way for Ethernet to become the network of networks and to phase off IP from packet networks as the core networking protocol.

References

[1] Internet World Stats, http://www.internetworldstats.com/stats.htm, checked June 26, 2009.

[2] Gsmworld, http://www.gsmworld.com/newsroom/market-data/market_data_summary.htm, checked June 26, 2009.

[3] World Broadband Statistics, Q1 2009, http://point-topic.com/.

[4] http://gigaom.com/2008/06/

[5] http://www.potaroo.net/tools/ipv4/index.html

[6] David Clark, MIT Communications Futures Program, Bi-annual meeting, May 30-31, 2007, Philadelphia, PA.

[7] http://www.celtic-initiative.org/Projects/100GET/.

[8] J. Ryynänen, Routed End-to-End Ethernet Network - Proof of Concept, Department of Communications and Networking, TKK, Diploma thesis, 2008.

[9] O. Lamminen, M. Luoma, J. Nousiainen, T. Taira Control Plane for Carrier Grade Ethernet, 2009, submitted to BIPN 2009.

[10] OECD, The Future of the Internet Economy – A statistical profile, June, 2008, available at http://www.oecd.org/dataoecd/44/56/40827598.pdf.

[11] M. Seery, Packet Transport Trends: IP/MPLS Success Challenged as Deployment Footprint Expands, IEEE Commun. Mag., July 2008.

[12] http://www.ietf.org/html.charters/trill-charter.html.

[13] http://www.ict-etna.eu/index.html.

[14] Network and Service Architecture, Deliverable D2.1, http://www.ict-etna.eu/documents.html#