Internet Evolution: Critical Issues

Mehdi Bezahaf, David Hutchison, Daniel King, Nicholas Race

Abstract—The Internet has been gradually evolving since its inception. In this paper, we highlight the crucial factors that have driven this evolution, and describe how the Internet is nevertheless still struggling with several critical issues that need to be solved to meet predicted requirements of future applications. We discuss possible approaches, solutions, and open issues, bearing in mind the considerable inertia of the Internet's key architectural features.

Index Terms—Internet Evolution, Architecture Design, Fragmentation, Management, QoS, Resilience.

1 Introduction

The Internet architecture continues to evolve, gradually. It has been a great success beyond even the most optimistic expectations. From the initial remote log-in, file transfer and electronic mail applications, then the rapid transition to the continuous media of audio and video services. Since then, changes have been more gradual. Now, voice over IP and content distribution have become ubiquitous, and we see the onward growth of video streaming platforms, social network services, cloud computing, and Internet commerce. This success has created a universal and near-global dependence on networking; any outage, due to a technical glitch, a hacking attempt, or a hardware malfunction, potentially has an enormous and adverse effect on governmental, political, economic, and societal activities.

The Internet, however, suffers from some deep-rooted problems related to its original design. These are now more significant issues because of society's dependence on Internet technologies. Keshav [1] points out that even after 50 years of Internet evolution, the Internet architecture still suffers from some fundamental problems such as email spamming, lack of privacy and security, and deficient Quality of Service. He also discusses the positive and negative impacts of the original Internet design philosophy on the Internet success.

In this paper, we first investigate the requirements or elements that we believe are crucial in the Internet evolution (Section 2). In Section 3, we discuss the difficulties of deploying these approaches, and in Section 4 we highlight and explain the range of critical open issues that need to be addressed for the future Internet, to ensure its continued success as an enabler and basis of the Information Society.

2 Internet Evolution

In highlighting eight requirements or elements we believe are vital in the evolution of the Internet, it is worth noting that almost all of them have been through the same Internet evolution cycle (Fig. 1). In other words, when a new event occurs this has triggered a development in the architecture, which then promotes new actors that go on to generate new behaviors. Finally, these behaviors set in motion new requirements. This section interprets how these points evolved and affect the Internet architecture from the early days of the Internet to the next generation of network.

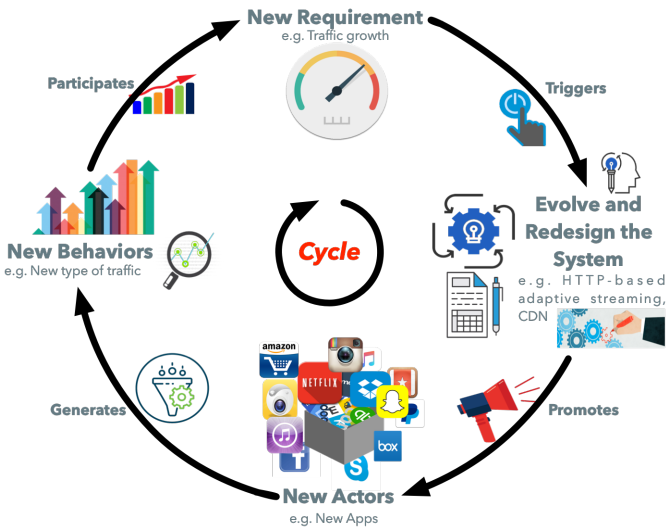


Fig. 1: Internet evolution cycle.

2.1 Traffic is changing

If we travel back to the summer of 1969, the volume of traffic on the original ARPANET network links was a few thousand bytes per month exchanged between four nodes. The number of nodes and volume of data transferred have kept increasing over the entire history of the Internet and will continue to grow in the future.

The Internet Protocol (IPv4) address exhaustion is a good example showing the impact of this rapid growth on the Internet architecture. The Internet Engineering Task Force (IETF) introduced Network Address Translation (NAT), as a short-term solution (since the mid-nineties), and Internet Protocol version 6 (IPv6) as a long-term solution, but the latter is still yet to be widely deployed.

The Internet architecture evolution promotes the emergence of new types of application and therefore new types

• *M. Bezahaf, D. Hutchison, D. King and N. Race are with Lancaster University.*
E-mail: (mehdi.bezahaf), (d.hutchison), (d.king) and (n.race)@lancaster.ac.uk

of traffic. Evolving challenges and requirements will reshape network traffic in the future, for example in holographic applications and with autonomous vehicles, which will cause further introspection about the Internet architecture. Many examples in the history of the Internet exist, and we can illustrate this point by highlighting how content sources have moved toward the edge of the network. In the last decade, video streaming has become the dominant traffic type on the Internet, which encouraged the use of different techniques to reduce server/network loads, notably Content Delivery Networks (CDNs) and the deployment of caches that bring the content closer to the end-users.

2.2 Users are mobile

Following the evolution of telecommunications and smartphones, almost every device in daily life is equipped with at least one wireless communication technology that has the ability to exchange data while moving. Because of this mobility feature, the entire architectural design of the Internet and its principal IP protocol were questioned. TCP/IP is a set of communications protocols in which an IP address is used to identify a unique end-point in the network. Address shortages mean that multiple devices behind a network address translator (NAT) can share a single public IP address. Additionally, however, an IP address indicates the device's location. This dual functionality (ID and locator) affects the transport layer during mobility. Changing the physical location (i.e. possibly changing the IP address) can break an open transport layer session.

The Internet of the future will be high-bandwidth and low-latency, and applications - such as remote surgery, augmented reality, holographic and autonomous vehicles - will require support for deterministic behavior. Furthermore, with the upcoming era of mobile communications, notably 5G+ and Low-Earth-Orbit satellite clusters (space-based infrastructure), the Internet will be expected to accommodate extremely high volumes of traffic from high-speed autonomous vehicles.

Different solutions may be found in the literature about how to deal with mobility, and splitting ID and location. However, mobility is still an open issue for future applications: see the survey by Akyildiz et al. about mobility management in IP-based wireless systems [2].

2.3 Inadequate end-point Internet bandwidth

With application traffic growth and the huge expansion of the user base, the Internet is encountering massive (over-) demand in terms of bandwidth. In most developed countries it is commonly reported that advertised broadband speeds are frequently not met, leading to many users being dissatisfied with the Quality of Service (QoS).

The advances in telecommunication systems and the economics of scale as the increased bandwidth helped the proliferation of cloud-computing networks. Even if we have witnessed a huge improvement in term of network bandwidth, cloud-computing providers adopt different techniques to cut down the bandwidth resources use as incremental backup technologies. CDN providers use caching techniques to bring the content closer to the edge and then reduce bandwidth consumption. New applications such as Internet of Things (IoT), holographic communications, and remote critical operations, will generate a significant amount of data and traffic control such that the bandwidth demand will be even bigger.

Note that the main and visible change/improvement was made at the physical layer (DSL, ADSL) and on physical equipment (copper, fiber optic).

2.4 High Internet delay

Høiland-Jørgensen et al. [3] show that variation in latency is both common and of significant magnitude. They show that latency variation has not been improved over time and that there are significant regional differences.

Bufferbloat, which represents the large queuing delays on the Internet, often lead to network performance degradation and packet loss. To avoid this bufferbloat, different approaches have been studied such as scheduling and controlling the bandwidth using multiple disjoint paths [4], or leveraging software-defined networking (SDN) to address issues of data transport service control and resource provisioning [5].

The deployment location of the network components, such as servers, caches; and the strategies for accessing data can have a significant impact on the application latency. For instance, reducing Domain Name System (DNS) lookup latency by redirecting data connections of a client to a geographical closer server or with lower estimated latency; placing the content in proximity to the end-user, such as network proxies, network/client caches, or prediction and latency-hiding techniques used for gaming [6].

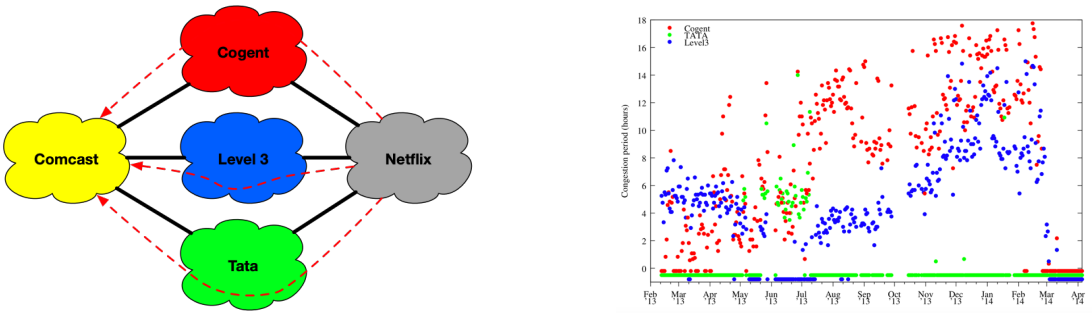
Historically, the Internet has been designed around three parameters: the capacity of the network, the number of bits in a data packet and the reach of the network. However, future applications such as holographic-type and haptic communications will be more about responsivity and interactivity and today's Internet latency will be too large for future high-precision network applications.

2.5 Interconnections impact

End-to-End packet delivery is achieved through multiple interconnections between heterogeneous entities called Autonomous Systems (ASes). The initial AS interconnection was relatively simple, involving mainly Internet Service Providers (ISPs) with a balanced mixture of inbound and outbound traffic. However, this interconnection becomes significantly more diverse and complex with time due to financial and sometimes political reasons.

Content providers are playing a role in changing the hierarchy between ASes. Prior to March 2014, all Netflix traffic toward Comcast users used to transit through several paths via multiple providers, including Cogent, Level 3, Tata, and others (Fig. 2a). Clack et al. [7] evaluate the impact of these interconnections in term of congestion. They show that the congestion duration on Cogent, Level 3 and Tata's links drop from almost 18 hours/day to relatively no congestion after March 2014, due to the agreement between Netflix and Comcast to peer directly (Fig. 2b).

The fragmentation of the public Internet and the growth of content providers will probably result in a complex and unbalanced ASes interconnection, causing Internet architecture flattening.



(a) (b)

Fig. 2: The rise and fall of congestion on Comcast links. (a) A (simplified) illustration of how Comcast users actually stream Netflix traffic (2013 - early 2014). (b) Estimated congestion duration for links connecting three major networks to Comcast.

2.6 Poor availability

In the Oxford dictionary, *Availability* is defined as the quality of being able to be used or obtained at any time and any condition. Applied to networking, availability means if working paths exist, the Internet protocols are designed to find them, even in the face of failures. This is true in theory, but in practice, routers can fail to detect or reroute around a failed link, causing *silence failures*.

A comprehensive survey on Internet outages [8] pointed out that numerous short-lived outages are due to the Border Gateway Protocol (BGP) route changes and route convergence delays. Additionally, outages are also due to misconfigurations, mobility and criminal attacks (e.g., prefix hijacking, and DDoS).

It is true to say that the Internet has been reliable for most of its history. Nevertheless, there are shreds of evidence that BGP is increasingly complex and too fragile to be patched in the future.

2.7 Internet insecurity

As highlighted above, the Internet is based on multiple interconnections between heterogeneous ASes to deliver packets End-to-End. In 1989, BGP was created to decide how packets are routed between these ASes. Since then, four versions have been released. Other versions have been proposed to improve the protocol, but BGP is still suffering from a variety of problems such as link failure, scaling issues, and security concerns. Traffic interception and route hijacking (using Man-In-the-Middle) are dangerous threats, where traffic follows a route that it is not supposed to traverse before reaching the final destination. In 2013, Renesys reports that more than 150 diverse victims (financial institutions, VoIP providers, and world governments) have been targeted [9]. They show for example that traffic, supposedly going from New York to Los Angeles, has been hijacked and rerouted to Moscow and Belarus before reaching the final destination.

Due to the likely criticality and confidentiality of future applications, and considering that BGP, for example, has not been updated or replaced since 1994, security is one of the major topics to address. Moreover, autonomic network management and orchestration can open a security breach.

2.8 Transparency

In the Catenet concept [10], which can be described as an early version of the Internet, a clear assumption was that a single logical address space would cover the whole network. Two important results emerge from this address transparency conceptualization: 1- packets could cross the network without any alteration, and 2- the source/destination pair addresses could be used as unique labels for the end-to-end communication.

With the growth of the network, various reasons have led to the disappearance of end-to-end transparency. We can cite the emergence of the Intranet concept or private network; dynamic address allocation; firewalls; private addresses; NATs; application-level gateways, relays, proxies, and caches; middleboxes; and so on. For more information, we invite the reader to study RFC 2775 about Internet transparency.

Today, a sender cannot obtain any guarantee that its packet will travel along a certain path. This is because routers with forwarding tables can update their tables at any arbitrary point of time. Even if the sender would know the entire forwarding state of all the routers in the path, which is an extreme case, it cannot guarantee that one or more routers will update their table after the packet has been sent.

In the future, critical applications can benefit from this property. However, transparency has to be implemented appropriately to avoid blacklisting some ASes.

3 Discussion

3.1 Internet inertia

Many alternative clean-slate approaches sound promising on paper, but they experience huge inertia because of the adherence of current applications' designs to the TCP/IP model and the current transport protocols. Another significant point is the difficulty of field-testing and validating these approaches. Some testbeds have been created to experiment with these architectures at a large scale, for example the MBone for Multicast, QBone for QoS, 6Bone for IPv6 solutions, GENI, and PlanetLab.

In addition to the difficulties cited above, we know that even the TCP/IP improvement approaches can be difficult to deploy, notably and not least with IPv6.

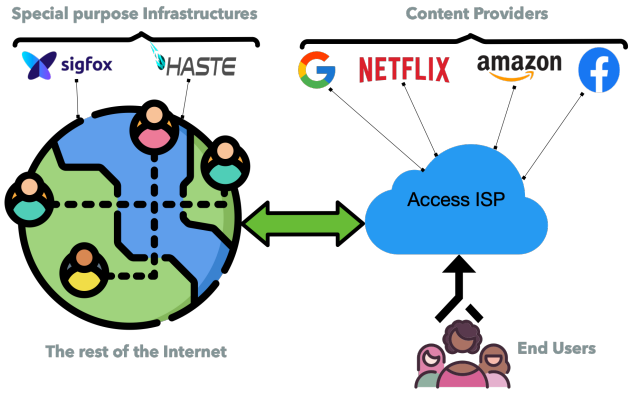


Fig. 3: Internet fragmentation.

Of course, the public Internet is still working and providing services, probably better than we could have imagined if we look back at where it started. However, as discussed above, the Internet may not be able to provide for all future applications' requirements. Also, there are now large corporate networks (based on Internet protocols) but operating and evolving independently.

3.2 The fragmentation of the Internet

One reason for Internet success is the service-infrastructure cycle, where we establish and/or upgrade the current service, when this one need to be scaled or new service (requirement) emerges. This cycle has been working for almost 40 years to produce the current Internet but since 2005, some researchers started denouncing the Internet as becoming ossified and complaining about the inability of the Internet architecture to adapt to new pressures and requirements.

Before becoming the one-network that we all know, the Internet used to be a set of separate networks, where no interconnections existed between these networks. Ammar [11] believes that this ossification is pushing the Internet to a new paradigm, where everything old is new again. In fact, he believes that the Internet architecture is fragmenting again and transforming from one-network to multiple-networks. If we look closer to the current Internet, we can see that in parallel to the public network, already a number of private networks exist to meet certain requirements as Sigfox for IoT, and HASTE for gaming latency (Fig. 3). Additionally to the private networks that bypass the Internet, the role that the content providers are playing by peering directly with ISPs and by pushing their content to caches closer to the end-users results in fragmenting the Internet.

3.3 New initiatives

On top of the clean-slate approaches, the Internet Research Task Force (IRTF) focuses on longer-term research issues related to the Internet, while the IETF focuses on the shorter-term issues of engineering and standards making. Fourteen Research Groups tackle different topics like cryptography, privacy, network management, global access to the Internet and so on. In recent months, the Internet Advisory Board (IAB) has shown some interest in starting an activity on resilience - CHallenges for Internet Resilience (CHIRP).

The International Telecommunication Union (ITU) established in July 2018 the Focus Group on Technologies for Network 2030. Its objectives are to study, and survey existing technologies, provide guidelines for standardization roadmap and formulate all aspects of Network 2030, including vision, requirements, architecture, novel use cases, evaluation methodology, and so forth. This activity is running in parallel to the IRTF and IETF work.

3.4 Programmable networking

Programmable networking has a long history - for example, application level programmability [12]. Over the last decade, sustainable development and innovation have evolved from the networking world. In 2008, the first API for OpenFlow was released, which allows the control plane to communicate with the data plane. OpenFlow is considered as an enabler of software-defined networking (SDN), which itself is defined as a technology that promotes network management and configuration in order to improve network performance and monitoring. Additionally, ForCES and POF represent approaches for designing and deploying programmable data plane devices. Thanks to OpenFlow, new concepts as network-operating system were reborn with the introduction of OpenFlow-based network operating systems, such as NOX, Onix and ONOS.

Later on, and with the efforts made at the European Telecommunications Standards Institute (ETSI), Network function virtualization (NFV) [13] became a hot topic for both industry and academic world. By decoupling network functions (NFs) from the physical devices on which they run, NFV has the potential to provide efficient scalability and ease of migration, which adds certain flexibility to the service.

Networking is becoming more automated and malleable with the introduction of intent-based networking, and new programming languages and orchestration techniques, and moving away from CLI-based configurations and the need for highly specific technology experts to deploy services. For example, in conjunction with OpenFlow, P4 is a high-level language for programming protocol-independent packet processors. T-NOVA is a full software-based management and orchestration (MANO) stack that operates with OpenStack (cloud orchestrator) and OpenDaylight network controller, both capable of being driven by high-level intent interfaces.

4 Open Issues

The future Internet will need to ubiquitously connect massive numbers of physical entities, such as smart terminals, sensors, wearables, vehicles, and industrial control devices. The popularization of in-network computing and machine learning technology will greatly increase the service resources, such as micro-services, processes and functions. The content in the network will also be moved away from highly centralized model that we see now, to a massively distributed model, where end-users and applications have low-latency access. We will also see content and functions no longer bound to specific locations or specific hosts. We outline critical open issues in the following sub-sections.

4.1 Viewpoints

The Internet architecture is a complex system with extensive specifications, where no single individual can fully comprehend all aspects of its requirements. In order to deal with the current issues, we should be able to see the Internet from different angles. The International Organization for Standardization (ISO) Reference Model for Open Distributed Processing (RM-ODP) [14] specifies a set of viewpoints for partitioning the design of a distributed software/hardware system. They define five different viewpoints (Enterprise, Information, Computational, Engineering, and Technology). So for example, we should not just think about technological or engineering aspects but also the business or enterprise point of view, which is becoming extremely influential.

4.2 Management

Conceptually, in the networking context, the logical separation of a network's operation into planes was first introduced in the ITU on ISDN in the 1980s. Three planes were adopted, namely the forwarding, the control, and the management planes.

As the control and forwarding planes, the management plane, which concerns methods of configuring the control plane (CLI, SNMP, etc.), is also vital to the system. In literature, most of the contributions omitted the management plane or considered it as a subset of the control plane. In fact, as an example, SDN concept, which is the new emerging architecture in recent years for network management and control, goes a bit further in term of separation and claims the physical separation between the forwarding and the control planes. The Open Networking Foundation (ONF) described it as *"In the SDN architecture, the control and data planes are decoupled, network intelligence and state are logically centralized, and the underlying network infrastructure is abstracted from the applications"*. Note that the management plane is not mentioned at all in this definition.

The constant need to autonomically deploy and supervise services in a cost-effective way has driven the evolution of the control, orchestration and management planes. Velasco et al. [15] propose a high-level point of view of a control, orchestration, and management architecture that enables dynamic provisioning of services based on SDN and NFV principles.

4.3 Quality of Service

One of the most challenging requirements for future Internet is managing application traffic throughput and latency, described in section 2.1. As new types of network traffic and applications have emerged, best-effort communication is no longer enough. Current Internet transport methods can differentiate class of services and using complex protocol stacks they can also guarantee bandwidth. However, these techniques are unable to provide guaranteed throughput and latency, or even to predict deterministic latency.

Emerging applications will need different processing and requirements for their own characteristic flows to make the delivery successful over the Internet. We believe that to address these kinds of requirement, there need to be well-defined QoS mechanisms in the network, complementing the original de facto best-effort delivery model. Various types of QoS architecture to support QoS provisioning were proposed, notably the Integrated Services (IntServ) and Differentiated Services (DiffServ) models.

With its centralized global view, the SDN controller has a visibility of the whole network unlike conventional networking, which gives it the ability to optimize flow management and reallocates resources actively and dynamically. Because of this dexterity, it becomes feasible for the network to perform per-flow or application-level QoS provisioning.

Karakus et al. [16] classified in their survey different studies into seven categories that are the most prominent ways in which QoS can benefit from the concept of SDN. Each category reflects a problem/challenge for QoS in SDN.

4.4 Resilience

Hutchison and Sterbenz [17] introduced the scientific disciplines that serve as the basis of network resilience, and provided the well-accepted definition of resilience as the ability of a network to continue to offer a satisfactory level of service despite the challenges that it faces. Due to the critical aspect of future Internet services and applications, we believe that structural and operational resilience should be a native feature of the Internet architecture, in order to enable the network to continue to offer its service in the face of, e.g., cyber-attacks, hardware failures, misconfigurations, accidents or natural disasters, and human error.

Fast convergence has typically been used by routing protocols for ensuring Internet resilience. The reason for a routing protocol to reconverge is mainly due to network topology changes caused by router or link failure or removal, or the addition or repair of routers or links. However, in future the Internet will need to consider predictive routing capabilities, which would allow a change in the state of a router or host to be predicted; hence the routing algorithm can make route changes before or as an event occurs. There is a new category of applications that may benefit from predictive routing: these are applications where packet loss or delay is potentially very harmful such as with cars driving on a highway, or robots moving in a factory. Predictive routing will likely require machine learning to help profile and classify the nature of failure, and the most efficiently solution to mitigate or negate impact on applications and end-users.

Moreover, these new predictive routing techniques will also need to be combined with QoS (which of course resilience underpins) is arguably the most important feature of a service in the eyes of network end-users and perhaps increasingly for operators and service providers.

4.5 Resource identification and addressing

New methods for resource identification and addressing will also be required for the future Internet. Although the IPv6 fixed 128bit addressing helped with the constrained Internet address space, its internal structure is ambiguous and too complex for IoT and Internet of Everything (IoE) applications and the lossy and low power devices used in Smart City and Industry 4.0 applications.

Furthermore, as content, resources, and Mobile Edge Computing (MEC) functions are growing very fast and are

highly distributed, it is crucial to have suitable mechanisms to find, and then efficiently connect to these endpoints using variable addressing and a semantic definition of the identifier for physical and virtual objects [18].

4.6 Instrumentation

The Internet architecture has been evolving since decades, by deploying new services, applications or protocols, monitoring them, measuring their performances and re-adjusting them if needed. In the early days of the Internet, collecting and reporting data was straightforward. However, nowadays, instrumentation becomes problematic due to the network complexity, systems distribution, and systems becoming increasingly more mobile and heterogeneous, which involve multiple measurement entities.

Instrumentation will be an essential feature for the future Internet, with a proper framework for data collection and reporting. Mehani et al. [19] suggest a list of requirements as good practice guideline, which we believe is a good start.

After more than a decade since David Clark proposed the *Knowledge Plane (KP)* for the Internet, where key information about the network design and operation would be maintained, Mestres et al. [20] present a KP-based architecture and the potential benefits of its deployment. The authors believe that progresses in SDN, data analytic, and machine learning makes a KP solution achievable.

4.7 Orchestration

The term orchestration is heavily used in different parts of the field. As a SDN term, orchestration refers to a universal function that manages and automates network behavior. The term was also adopted by ETSI in the scope of NFV, where its definition leads to a vague distinction between orchestration and management. In fact, very often, the distinction between the terms orchestration, automation, and management is blurred. Automation describes the mechanism of making a task running without human intervention. Management is all about maintaining the infrastructure in good condition. Orchestration, in turn, is concerned with the execution of each task or process in the correct order.

For future network generations, orchestration will be an essential element. It needs to be able to handle the complexity of services and to support rapidly-changing intents in the network coming from both applications and network operations: a key challenge is to develop a scalable orchestration process. Resilience (in orchestrators, controllers, and managers) and state synchronisation between functional components is a critical and open issue.

4.8 Intent-based networking

We highlighted how rapid growth in the variety and volume of traffic will apply pressure on the operation and management of existing network technologies, and only increase the management manpower and technical knowledge required to operate new networks. Instead, an efficient and faster deployment of infrastructure and services will be required using a high level of automation; this issue is being explored under the title “intent-based networking” (or IBN as it is becoming known).

In intent-based networking, the application or end-user will interact indirectly with the network layer through the intermediate intent (sometimes called the knowledge layer) based on application and technical agreements. This is a high-level abstract concept known as the declarative request. Then the declarative request can be mapped to imperative policy, with the help of policy database, the translation module and decision module. Eventually the policy will be applied as action such as configuration device model, or end-to-end service setup.

Intent-based networking will allow the Internet to evolve from a human-driven static resource management system, to an automated dynamic system that continuously and consistently adapts to end-user and application demands.

4.9 New network varieties

New network varieties such as IoT, IoE, MEC, and Edge Computing aim to reduce network latency and improve overall application performance. However, they raise yet further issues that need to be resolved, including: device and resource bootstrapping and location dissemination; end-to-end addressing across non-traditional IP domains (”ManyNets”) and network path decision-making; privacy and security (e.g. DNS over HTTP); and system resilience.

Internet via space-based communication infrastructure also presents several challenges when using existing packet transport and control technologies. These include minimisation of traffic delay and ensuring highly responsive routing, in an environment where frequent changes will occur due to the spatial network connection within satellite clusters and dynamic space-to-earth downlink selection.

5 Conclusion

The Internet has been further evolving in recent years, not least in terms of new practice, with traffic bypassing and Internet fragmentation. Going from one network to multiple networks (viz. ManyNets) has both advantages and disadvantages. On one side, this will promote innovation and ease the deployment of clean-slate approaches. On the downside, it could be an economic drawback for new actors to build new services. Nowadays, a new company can potentially deploy a new service on the Internet (i.e. one shared network) accessible to all of us without too much cost. The challenge is how to keep these economic features low while embracing the evident advantages of fragmentation.

The Internet is still suffering from critical, ongoing issues such as the lack of security and privacy, and the evident need to provide resilience and QoS. It is also becoming clear that replacing the current Internet with a new clean-slate architecture is impossible to realize without fragmentation, though the development of parallel networks (by large corporations) may offer the most likely way ahead for radical change. We need to keep in mind that the Internet evolution seems to indicate the practical impossibility of developing solutions to resolve all of the issues. However, we believe that virtualization of networks forms a very promising basis for the future. One specific area of investigation that needs to be re-visited is the autonomic management of