Beyond the Future Internet – Requirements of Autonomic Networking Architectures to Address Long Term Future Networking Challenges

M. Siekkinen, V. Goebel, T. Plagemann, K.-A. Skevik
University of Oslo, Department of Informatics
P.O. Box 1080 Blindern
NO-0316 Oslo, Norway
{siekkine, goebel, plageman, karlas}@ifi.uio.no

M. Banfield, I. Brusic
Telekom Austria AG
Lassallestraße 9
A-1020, Vienna, Austria
{mark.banfield, igor.brusic}@telekom.at

Abstract

We present in this paper an overview of current problems with the Internet from the architectural view point, and we identify and forecast trends for future developments. Our study includes perspectives from different stakeholders of the Internet as well as from the technological development point of view. We conclude that current and future requirements vary greatly between, and even within, different stakeholders, which implies that a successful future network architecture must provide maximum flexibility with a minimal set of "hard-wired" properties. In specific, a revised version of the TCP/IP layered architecture would only suffice to address some of the current issues, but would not provide a long term solution. We then show that an autonomic network architecture with a few simple architectural building blocks is enough to address a great deal of the requirements today and in the future.

1 Introduction

Our modern society is using computer networks for nearly all aspects of live, like social networks, education, health system, entertainment, industrial production, military, etc. There is no doubt that we will in the future depend even more on computer networks and especially the Internet. However, there are already today many problems with the Internet, like security threats and insufficient resilience, high costs of network management, inflexibility of the Internet, etc.

These problems have been recognized and we can see today new initiatives that will support disruptive research to develop long term solutions for the future Internet. Examples include FIND (Future Internet Design) which is a research area in the NeTS programme of the National Science Foundation [3], and the European initiative for the Future Internet in the 7th EU Framework Programme.

The Autonomic Networking Architecture (ANA) project [5] is one of the few projects that have already been launched in the 6th EU Framework Programme in Europe [29]. The central claim of the ANA project is that substituting today's Internet by another future Internet just means to repeat the error we made in the past. A particular network needs to be based on a set of standards, which in turn restricts its ability to evolve with future requirements and technical developments. In addition to the continuously increasing heterogeneity of application requirements and networking technologies, legacy systems like today's Internet need to be supported in the future. Therefore, we aim in the ANA project for an autonomic networking architecture [28] that is based only on a minimal set of standards to boot strap all kinds of inter- and intra-networks. Such a solution goes beyond what a single future Internet could provide.

In this paper, we present the outcome of our analysis of today's problems, future developments, and requirements that we performed in the scope of the ANA project [7]. We are not the first to perform such a study and we share many insights in common with previous work [14]. However, our work differs in the sense that it combines the viewpoints of academic research and commercial service providers, and it also studies the implications of the identified requirements. In particular, we identify which basic concepts an autonomic networking architecture must provide in order to address the identified (and other future) requirements. Furthermore, we study the requirements that are imposed onto these basic concepts.

This paper is based on two categories of knowledge. On one hand, we base our reasoning about requirements of the future autonomic network architecture on our own experience and expertise: Telekom Austria represents the service provider view and University of Oslo represents the academic view. On the other hand, we used the following external sources: major scientific publications, insights from

Proceedings of the 11th IEEE International Workshop
on Future Trends of Distributed Computing Systems (FTDCS'07)
0-7695-2810-4/07 $20.00 © 2007 IEEE

industry [6, 19] , and documentation and discussions from initiatives and corresponding meetings on the future Internet[1].

The remainder of this document is structured as follows: In Section 2, we present a set of problems with the Internet architecture that we experience today. In Section 3, we analyze the future developments, first, from the point of view of different stakeholders and future services and applications, and then, from the technology viewpoint. In Section 4, we present a set of core architectural concepts that we claim to suffice to meet most of the current and future communication requirements.

2 Problems Experienced Today

In this section, we identify the problems that arise already today from the restrictions that the Internet architecture imposes. First, we go through some literature to list some problems identified by others. As a second step, we focus in more detail on issues particularly important for a commercial service provider, i.e., legacy systems and Universal Service Obligations. These and other problems are illustrated with the example of Voice over IP (VoIP). In this paper, it is only possible to highlight those problems here that we believe are among the most important ones today.

2.1 Brief Survey of Some Issues with the Internet

A wide range of applications have been successfully deployed on the Internet, but there are many cases where the current design results in inefficiency. P2P overlay networking applications represent a good example of this. For optimization, these systems often make use of measurements to estimate latency, available bandwidth, or similar information about the links between participating machines. Without any support in the network for obtaining such information, the measurements need to be done by the application. As a result, each application performs these operations independently, contributing to increased traffic on the Internet. The accuracy of the measurement techniques is also limited, possibly leading to additional overhead due to inefficient optimization of the overlay network. Indirection is another area where the lack of support results in inefficiency. Use of caches, proxies, or support for mobility requires indirection, but without any specific mechanism for this, a separate infrastructure is required for each [18].

[1] FIND organized by the National Science Foundation (NSF) in December 2005 in the USA and the preparation meetings for the EU's initiatives in the 7th Framework Programme organized by Directorate D Network and Communication Technologies of the European Commission (DG-INFSO-D)

While this inefficiency is unfortunate, there are areas which are more problematic. As the wealth of literature demonstrates, security is a major issue since many years [11, 13, 22]. There are tremendous efforts from both, research community and industry, to cope with the viruses and worms [24], for instance. The problems originate from the fact that the Internet was originally designed to be an open network with no centralized control and having mutual trust among users. As a result, its use cannot be administered by a central authority, and it is difficult to ensure integrity, authenticity, non-repudiation, privacy, etc. There is no doubt about the fact that the future Internet should be secure and should protect privacy. However, this often comes with a particular disadvantage, i.e., moving from an open network to a closed or partially closed network. For example, the use of firewalls in a P2P system has a negative impact on performance of peers that are behind a firewall. Therefore, it is important for future Internet solutions to address this trade-off between advantages and disadvantages of open and closed networks. Another important issue is the need for efficient ways to know what is going on in order to identify new attacks, and trace back to study intruders and break-ins, for instance. Such properties require well designed monitoring solutions for both real-time analysis and analysis of historical data.

Resilience, which can be characterized as the ability to operate and maintain an acceptable level of service under the presence of adverse conditions, is an issue closely related to security. Hence, considerable research has been conducted on that domain as well [10, 31]. The way today's Internet has evolved lacks fundamental support for resilience both at the infrastructure level as well as at the service level. Many of the efforts to introduce resilience in the current Internet architecture have made evident that the rigidness and opacity of the layered architecture is one of the fundamentally restricting factors in addressing the problem of resilience [8].

Unfortunately, changing the Internet is not easy. Even minor changes to the address format cannot currently be performed [9]. The large number of deployed machines which support IPv4 further complicates this. ISPs have also shown themselves to be very conservative, which is a problem since any large-scale change would need the consensus of ISPs [12].

Overall, there are several problems facing the Internet, but the inertia of the currently widely deployed IPv4 means that the changes required to address the problems are unrealistic. Even IPv6, the successor of IPv4 that included already in its specification mechanisms for a smooth transition from IPv4 to IPv6, is not widely adopted.

Proceedings of the 11th IEEE International Workshop
on Future Trends of Distributed Computing Systems (FTDCS'07)
0-7695-2810-4/07 $20.00 © 2007 IEEE

2.2 Voice over IP

In this section, we focus on problems that arise from the commercial provision of VoIP services. We selected this example, because voice services have been very successful in traditional telecommunication networks. Thus, VoIP is a very important business area for commercial providers. We identify the following problems that are of particular concern for VoIP:

- *NAT traversal:* Conventional VoIP protocols separate signaling from the audio traffic of a telephone connection. The port on which the audio traffic is sent is random and NAT routers handling the signaling traffic may have no way of handling the corresponding audio traffic. As a result, the audio traffic is not translated properly between the address spaces.

- *Voice security [4]:* Additional security threats for conventional voice transmission are toll fraud and eavesdropping. For example for eavesdropping, a free program called VOMIT (Voice Over Misconfigured Internet Telephones) allows intruders with access to a local VoIP network to capture traffic, convert it to an audio file and replay the voice conversation[2].

- *Emergency calls:* The use of IP makes it difficult to geographically locate VoIP users. Thus, emergency VoIP calls cannot be easily routed to a nearby call center and if the caller is unable to give an address, emergency services may be unable to locate the caller.

- *Phone power supply:* IP phones are not supplied with power over the Internet connection, while the majority of the traditional telephones is supplied with power from the exchange. Thus, VoIP phones may not be able to make phone calls during a power outage.

- *Reliability:* Legacy telephone networks have a very high availability (estimated to 99,999%), while the reliability of an Internet access is clearly less [32]. In [32], the authors mention that many voice encoders can handle up to 1% packet loss. However, the global packet loss statistics of the Internet in [2] fluctuate between 3% and 6%.

- *QoS:* The TCP and UDP protocols inherently provide only best effort to all packets. The common approach used today for providing QoS to Internet applications is over-provisioning of bandwidth, but this does not give any guarantees. In case of congestion, real-time applications like voice transmission will hardly cope with packet loss, because it does not make any sense to retransmit lost packets. Congestion in the networks also effects the variation of delay (jitter), which is an important QoS characteristic of voice transmission.

- *Delay:* There are two sorts of delay: the first one is caused by coders and the second by routing. The speech quality is worse if more coders/decoders are used in the communication path (VoIP calling party connected over ATM backbone with GSM user). Also a long medium transmission path could exceed the maximum delay of 300 ms for real-time voice communication.

[2]Even though some VoIP clients, such as Skype, encrypt conversations, many others do not. In addition, IP PBXs are vulnerable for tools like VOMIT.

2.3 Legacy Systems

So far, we have addressed in Section 2 problems with today's Internet that are probably well-known to most of the researchers in the networking domain. However, there is another class of problems that is probably not so obvious for academic researchers, but very important for commercial service providers. The following problems have their origins in the need to support legacy systems and Universal Service Obligations (USO):

- *Heterogeneity:* A broad range of different networks exists today, with different technical, regulatory and legal requirements. Many of them also need support for adaptability of networks and applications from the future network architecture.

- *Universal Service Obligation (USO):* For telecommunication networks, an important requirement is to support USO [1]. These are obligations to provide basic telecommunication services in certain areas at fixed prices, which are imposed by the government on the network operators (both mobile and fixed).

- *Emergency Call Handling (ECH):* ECH is an essential feature of today's public switched telephone networks (PSTNs). Any new application and network which comprise PSTN replacement must provide reliable handling of emergency calls. Based on today's experience with VoIP, one of the main problems in fulfilling ECH obligations is related to the fact, that IP addresses serve as identifiers and locators at the same time, another one is insufficient reliability of the Internet.

- *Lawful Interception (LI):* is a requirement placed upon service providers to provide legally sanctioned official access to private communications. With the existing PSTN, Lawful Intercept is performed by applying a physical "tap" on the telephone line of the target in response to a warrant from a law enforcement agency.

Proceedings of the 11th IEEE International Workshop
on Future Trends of Distributed Computing Systems (FTDCS'07)
0-7695-2810-4/07 $20.00 © 2007 IEEE

However, VoIP technology has enabled the mobility of the end user, so it is no longer possible to guarantee the interception of calls based on tapping a physical line.

3 Future Trends

In this section, we analyze future requirements of different stakeholders and services in the first step. We then look at the trends for future networking technologies.

3.1 The Stakeholders

When identifying requirements of future networks, it is difficult to define a final list of consolidated requirements since each stakeholder has its own list of demands, which may in part contradict those of other stakeholders. In this section, the breadth of high-level requirements are detailed from the perspective of six distinct stakeholder positions that have an important part to play in the development of future autonomic networking.

3.1.1 End Users (Consumers)

The group of end users is a quite diverse group. The major part of this stakeholder is composed today of consumers of telecommunications services, that typically comprises residential consumers (e.g. today's ADSL customers) or corporate employees (or similar). The main difference between them is that residential consumers are responsible for the purchase of telecommunications services, while the corporate employees merely consume services provided by the corporate (or academic) network.

With the increasing complexity found in today's multi-PC home-networking environments, combined with ever more complex and demanding applications, the issue of ease of configuration and customer support is likely to become ever more important for this stakeholder. The complexity for the end user must be severely reduced in future networking architectures if the true wealth of disparate applications is to be realized. The future network architecture should aim to solve configuration faults automatically or where physical user or operator intervention is necessary should provide assistance in diagnosing faults.

The networking requirements posed by consumers can be very diverse. Typical Internet users' main concerns are with price and performance while more recently reliability/technical support has become important in reducing customer churn. On the one hand, these users range from co-operate users who are willing to pay a high price for quality services since loosing access to network applications may result in severe business losses. On the other hand, this can be someone using a public open network, e.g. open WLAN networks that could be free of charge to use, but at the same time the provision of a secure transmission is left completely to the users responsibility. Since the service is not paid for, few expectations can be made in terms of performance and quality.

Not all users require connectivity to national and international networks (i.e. the public Internet), many applications require only local (LAN) or personal area (PAN) networking. In this case a network for a particular person comprises all the devices, gadgets, and household appliances owned by this person. These applications probably have high security requirements but with bandwidth demands ranging from a few bit/s till many Mbit/s. An example is ad-hoc gaming networks, often established ad-hoc for the playing of a specific game during a "LAN party". These networks may have quite diverse security requirements, in a gaming scenario users may want to form network instances that are accessible only by certain individuals or devices and therefore strict access control is required. In other cases the game may be open to all users in a given area. Another recent example is IPTV; due to the concerns of the major Hollywood studios combined with national based content licensing schemes and high IP transmission requirements (multi-Mbit/s stream, IP multicast etc.), current IPTV solutions are based on physically separate IP infrastructures than the public Internet, thereby providing a guaranteed quality within a physical locality (e.g. a city-based cable network or national ADSL footprint).

Given all these types of users it is difficult to generalize about the requirements consumers make on future autonomic networks. However, it is assumed given the diverse users, that the network should support a range of performance at different price points, thereby covering mission critical applications as well as "casual" web surfers. A single public Internet with full global addressability can also not be assumed since many application scenarios are confined to a specific location (be it geographic, e.g. the body, an apartment, or political, e.g. a country, or organization, e.g. a company). Autonomic networks should not attempt to force the reinstatement of the holly grail of full end-to-end connectivity now destroyed by firewalls, NATs and VPNs; our analysis shows that many consumers and applications do not require this scenario, future networks should constrain connectivity to the domain required.

3.1.2 Telecommunications Service Provider (Network Operator)

Telecommunications Service Providers either own or lease through wholesale purchase transmission capacity and sell it to end users. The main goal of telecommunications service providers is to provide their shareholders with increasing revenue returns and profit based on: 1) Increasing revenue through the provision of novel services which can be

billed, 2) reduced operational expenditures through simplified management processes enabling greater degree of process automation and less manual intensive procedures, and 3) reduced capital expenditures through networks which require less or cheaper components and which can carry multiple traffic types.

The Telecommunications Service Providers business model is based on multiplexing to enable overbooking. Thus, a given transmission capacity can be resold many times. The history of technological advances in the industry has been chiefly concerned with increasing the efficiency of multiplexing. The future Internet should lead to even more efficient multiplexing of traffic types. Given the use of overbooking to provision circuit-switched traffic (e.g. voice and video) over a packet-switched architecture, there is a demand for more efficient provisioning for mixed circuit and packet switched traffic over an underlying packet-switched core.

End users consider security as integrity of themselves and their data, while for operators security means safety of the network from the end user. The costs of securing a network, including buying, running, and maintaining firewalls, and investigating and dealing with security breaches is enormous. Given the trend, the costs of security may soon make up the largest proportion of the price paid by the end user. The future network architecture should reduce the overhead of securing the network in order to bring down costs.

3.1.3 Regulator & Other Government Agencies

Although the provision of telecommunications services has been liberalized, they remain subject to stringent regulation by government agencies and telecommunications laws. Considering the future network architecture, it is essential not to ignore the legal and regulation requirements influencing its development. The most important issues are wholesale provision and USO. Since we have already discussed USO in Section 2.3, we focus on wholesale provision in the following.

Given the natural monopoly (or at best oligopoly) in the provision of access to telecommunications services, governments have insisted on the granting of access by the owner of the network to third parties (i.e. competitors) on non-preferential basis. This has often meant that, from a technical perspective, traffic has to be routed suboptimally, since both the PSTN and Internet were not originally designed to support a wholesale business model. While the PSTN has evolved to support a range of wholesale services, the Internet still offers no real wholesale model beyond simple Wholesale Broadband Layer 2 Tunnelling Protocol (L2TP) access services. The future network architecture should contain provisions for supporting the dividing network ownership and telecommunications service provision.

3.1.4 Protocol Developer / Standardization Bodies / Hardware & Software Manufacturers

These stakeholders are interested in developing networking solutions that are based upon open standards that all can access. Typically, standardization will take place at the IETF, ETSI, IEEE or ITU although there exist today a multitude of small standardization fora.

The reason for standardized solutions is to reduce development costs while enabling as wide a market for the resulting products and services as possible. These stakeholders pay considerable attention to the requirements of their customers, the Telecommunications Service Providers and end users.

However, there are exceptions, where stakeholders have such market dominance that they may follow propriety solutions to protect their market and can hence dictate requirements to their customers.

3.1.5 Application Developers

The application developer requires from the network a well-defined API that offers a rich set of transmission capabilities. Until today, many features that are not provided by the Internet have been developed by application developers and deployed on top of the Internet, e.g. as overlay in CDNs and Skype, which results in inefficient and redundant solutions.

The Internet provides an open interface, the concept of BSD style sockets has enabled every application programmer to develop end user applications which has resulted in a wealth of innovation. Unfortunately, only the end user interfaces are well specified and open to end users. Network provision, management and routing are all performed over a wide range of protocols some standardized other proprietary which are totally controlled by the network operator. It is difficult for service providers (Google, MSN, etc.) to influence the provision of the communication service. Future network architecture should enable end users, operators, and service providers to (at least partially) control the network in order to optimize their service provision and enable innovation.

Current Internet communication is limited to a very basic API enabling the sending of packets without guarantee of arrival between two static addresses. Other addressing models, e.g. multicast and anycast, are very poorly supported; terminal mobility has not been commercially implemented; Quality of Service can now be provisioned by the network operator (e.g. MPLS) within the scope of its own network, but there is still no interface to the end user or service provider. Future network architecture should extend the richness of communication services beyond the basic IP function set.

3.1.6 Military

The requirements of military differ mostly in that security and resilience concerns are of first priority. Military forces may need to use networking devices in potentially very hostile environments. In addition, end-to-end connectivity for devices cannot always be guaranteed due to very diverse locations that communication needs to take place. For certain military applications, like object tracking, target recognition and self-defense real-time support is needed.

3.2 Future Workload Through Services and Applications

Although it is difficult to predict the future "killer applications", it is meaningful to reason about the potential workload that the future network will have to handle.

Wide scale of throughput: Recently, the miniaturization of computing devices, like sensor nodes, etc., has lead to networking applications that work over wireless networks with very limited bandwidth. On the other hand, it seems that high bandwidth applications, like streaming of High Definition TV, high resolution Virtual Environment and exchange of terabyte large files in Grid computing will also play an important role.

Low and high latencies: Recently, delay tolerant networks and applications are a very important research domain [17], with promising applications for communication in remote areas, like DakNet [25], for emergency and rescue operations in areas without infrastructure, and for interplanetary communication. On the other hand, interactive services either between human and machine or between humans require low latency. An emerging application domain is here for example haptics with tactile feedback.

Varying levels of reliability, availability, and resilience: We see already today the trend that more and more organizations are using the Internet as the core technology for their central operations and, thus, are dependent on the availability of the network and services. However, achieving high reliability, availability and resilience is costly. In certain cases many customers prefer reduced quality with lower costs over expensive high quality service, which is the case for Skype, for instance, as opposed to traditional telephone service. On the other hand, we notice that when both video and voice are delivered over IP but consumed over traditional devices (e.g. TV with STB or VoIP phone), the customer expects the same quality that they received with the old dedicated networks.

Heterogeneous addressing and naming: Addresses are a vital element in networking and will very likely continue to be so in the future. However, IP addressing is inefficient with mobile and multi-homed networking devices (PDAs with WLAN, 3G and BlueTooth). HIP [23] is one approach to improve the situation. Another area in which the concept of traditional addressing might change is context aware computing in which messages might be sent to nodes that are in a particular context.

Different security and access control policies: Network owners define their own security policies and will do so also in the future. Some organizations, like military may require closed high security networks with highly hierarchical policy structures, while it is reasonable to assume that public institutions will provide more open networks.

3.3 Future Technological Developments

The requirements on the future Internet are not only caused by existing and new applications and stakeholders, but also those technologies that need to be supported. Again, it is impossible to exactly predict the future technological developments, but it is possible to identify several important trends.

The number of networking capable computing devices will increase and these devices will be very heterogeneous, ranging from RFID tags, smart dust and sensors to high-end servers and super-computers [14, 27]. Furthermore, many devices will be able to use multiple networks (i.e. multihomed), for example advanced PDAs are already today equipped with WLAN, 3G and BlueTooth interfaces.

We observe an increasing number of networking technologies with heterogeneous properties. Some of today's networking technologies - especially those tied to fixed infrastructure, like cables, will exist for some time. At the same time, new technologies will emerge which may be low power consuming wireless networks with low bandwidth, but also high-speed wireless networks as well as very high speed optical networks [20] or so called "challenged" networks with no guarantee of end-to-end connectivity [17]. Not only the bandwidth will differ in these networks, but also their reliability, like bit error rate.

In recent years, we have seen an increase in wireless networking and mobile computing devices. Thus, computing devices are no longer bound to a physical location and operated only in a device "friendly" environment, e.g. through air conditioning. Important research areas include underwater communication, e.g. for sensor networks to control drilling for oil in the North Sea. Environments like this are not "friendly", and they challenge and compromise the reliability and resilience of the devices and the network. Mobility is not limited to persons walking, but also higher speeds in vehicles [16], airplanes, etc.

4 Fundamental Concepts for A Future Architecture

In the previous sections, we have shown that there are many problems with today's Internet and that there are

Proceedings of the 11th IEEE International Workshop
on Future Trends of Distributed Computing Systems (FTDCS'07)
0-7695-2810-4/07 $20.00 © 2007 IEEE

many additional challenges to meet in the future. In [8], it is also shown that single patches to the Internet are not sufficient to fix all of today's problems. The goal of the ANA project is to develop a new autonomic networking architecture that does not only solve many of today's problems with the Internet, but also addresses future long term requirements. To develop such a future proof autonomic network architecture, it is important to realize that simply a revision of the existing Internet protocols is not sufficient. A long-term solution for future networking must provide support for evolution in the sense that new requirements can be addressed by new solutions without breaking the architecture. In today's Internet architecture, the IP protocol, IP addresses, DNS, and BGP implement core architectural principles that restrict the Internet's ability to adapt itself to new requirements and environments. Therefore, we make the case that a few fundamental architectural concepts are needed to support both, today's needs and future and unknown needs. In this section, we identify these fundamental concepts and analyze the properties they must contain. We claim that these concepts are able to fulfill most of the future Internet requirements that we identified throughout the previous sections. We use an example to illustrate the need and the application of the fundamental architectural concepts we identify.

4.1 Realms

The future network architecture must support different types of networks, including legacy networks, like the Internet; networks with different privacy and security policies; networks that are revised versions of today's Internet; networks that are designed for interplanetary communication; or networks that work without addressing. We use in this document the term realm to capture these "network instances". A realm is essentially defined by a set of policies and rules that determine, for instance, the data transmission protocols and addressing scheme used and security and privacy policies.

There are two main reasons why the ANA architecture has to be based on an architectural concept that supports different realms: First, the history of the (planned) transition from IPv4 to IPv6 has shown that an instantaneous transition from one Internet protocol to another is very difficult to realize [21]. Obviously, an instantaneous transition from today's Internet to another network that is based on new architectural principles will be even more unrealistic. Thus, any new network should be able to co-exist with the Internet. Second, the networking world is prone to be very heterogeneous today and will be even more so in the future with respect to technologies, application requirements, requirements from providers, etc. This extreme heterogeneity leads us to the conclusion that no single solution can fit all networking environments and requirements, i.e., new realms must be established when necessary and co-exist with the existing ones.

4.1.1 Bootstrap

The layering approach of the Internet architecture gives flexibility in the sense that many different transport layer and application layer protocols can be deployed over IP without breaking the architecture. However, IP itself requires a particular addressing scheme, introduces the end-to-end approach, and leads to some of the problems we discussed in Section 2.2, for instance. An architecture that supports different realms with for example entirely different addressing schemes needs to achieve more flexibility than the Internet. Enabling more flexibility means standardizing less in the architecture. In this context, operating systems are a good example. There are many different operating systems developed for many different purposes that can run on the same hardware. To achieve this flexibility, only the first part of the bootstrapping process is standardized for all operating systems. When the computer is switched on, the processor jumps to a small piece of code that is at a standardized location in ROM (read-only memory) and executes it. This small piece of code jumps to a well-defined location on a stable storage, like disk, to load and execute the bootstrap code from the boot block. The bootstrap code is already operating-system specific in the sense that it boots a particular operating system. Thus, the architectural concept that enables different realms needs to provide a "bootstrap process" to boot realms which makes only a minimal set of assumptions and requires only a minimal set of standards.

4.1.2 Inter-communication of Realms

There are many logical networks within the Internet in form of overlays, Virtual Private Networks (VPNs), etc. These networks share a basic set of policies, e.g. IP, DNS, BGP, etc., but apply at a certain level different policies, e.g., overlay networks specify protocols to build virtual links on top of the transport layer and VPNs use particular security policies. This concept of structuring one large network to share only a sub-set of all policies needs also to be supported by the future architecture. Thus, realms must be able to host other realms in the way that they can overlap partially or entirely.

Co-existence of realms does not mean that only closed and independent networks are supported. This is obvious for structured realms, i.e., those that share some nodes, but it is also important for realms that do not share nodes. Thus, the future architecture needs to provide means for inter-communication of different realms. This property will be achieved through certain types of intermediate systems, i.e., gateways.

Proceedings of the 11th IEEE International Workshop
on Future Trends of Distributed Computing Systems (FTDCS'07)
0-7695-2810-4/07 $20.00 © 2007 IEEE

4.1.3 Self-* Properties

We have shown in the previous sections that the complexity of Internet-like realms will increase. This complexity is a threat to such realms, because their management will be either very expensive in terms of resources, not reactive enough, or even impossible. Therefore, it is the aim of the ANA project to target an autonomic network architecture that has the so-called self-* properties [19] that include self-configuration (able to adapt to changes in the system), self-healing (able to recover from detected errors), self-optimizing (able to improve use of resources), and self-protecting (able to anticipate and cure intrusions).

To achieve these properties for realms, and for nodes within them, it is necessary for the realm, respectively the node, to gain some knowledge about its state. For example, if the load is too high in a certain part of a network, the network might perform load balancing. However, first it has to gain knowledge about its load, which is part of its state. Therefore, monitoring must be an inherent part of the architecture that - which we discuss in detail in the next section - provides input data to processes that enable the self-* properties. Continuing with the example of load balancing, we can identify that it is not sufficient that only local monitoring data is used. Once a high load has been detected, it must also be determined how the load should be then redistributed. Thus, information has to be exchanged between different entities in the realm. Autonomic networking requires also that networking entities are able to explore their environment, for instance, to find out which policies, including protocols are applied, which services are available, or what the load is on other nodes. In order to do this, information has to be exchanged among networking entities. Finally, some kind of intelligence is needed that is able to reason about the state information and the information from other entities and perform certain actions, such as adaptations, if needed. Therefore, information and knowledge management needs to be supported.

4.2 Monitoring

Monitoring is the key element to achieve awareness, one of the essential properties of a system that aims for the self-* properties. The monitoring architecture should be such that it provides the necessary input for those components of the architecture that enable the self-* properties. However, the requirements of a specific network instance can be widely different from those of another network instance. Hence, also the knowledge required from monitoring to sustain the self-* properties of different network instances can vary a lot. For example, one network instance may require a certain level of QoS, while another one requires none. Therefore, the question of what exactly is to be monitored should not be answered here. Instead, the monitoring architecture should be generic enough to make it possible to monitor whatever is needed.

A notion of a monitoring session will be necessary. Given the potential volumes of monitoring data and variety of monitoring tasks, it is not feasible to assume "always on" type of services. The notion of a monitoring session enables on-demand monitoring services that are triggered only when necessary.

For extensibility reasons, it would be desirable to have programmable monitoring components to tell the monitoring nodes not only what to monitor, but also how to monitor.

The monitoring architecture should support different storage solutions for the data collected by the sensors. We need both, volatile and persistent monitoring data. Standard MIBs and data stream management systems (DSMS) [15, 26] are examples of using volatile storage of monitoring data. An example of persistent storage system for monitoring data using a database management system (DBMS) is described in [30]. Persistent storage enables evaluation of historical monitoring data in order to identify trends or "go back in time" to understand reasons for certain events , for instance. Dissemination of the monitoring data is tightly coupled with the information and knowledge management which we discuss in the next section.

4.3 Information and Knowledge Management

We have discussed the requirements for monitoring to enable an autonomic network to sense its operating environment and to monitor its state. However, achieving self-* properties requires more than just acquiring raw data about the operating environment and state. This data has to be used, i.e., transformed to information and knowledge to be applied to achieve these properties. For instance, self-healing requires that a system is able to define or to learn what the normal condition is and compare it with monitoring results in order to recognize deviations from the normal condition.

In many cases a single source of data and information is not sufficient. Therefore, it is important for the future architecture to develop proper abstractions for providing information and sharing information between entities. This information can be derived from monitoring data or it is part of a description of an entity, which could be a service description, but also the description of the entities configuration, available components, available resources, etc.

4.4 Example: Content Distribution Realm

The term content distribution network (CDN) covers many different ways of moving data between computers.

There are three main categories. The first is download based, where content is accessed only after having been completely downloaded. The second is broadcast based, where all receivers receive the same data more or less simultaneously. The third is CoD (Content-on-Demand) based streaming, where data is accessed as it is being received. Content is typically located in one of two ways; the identifier based approach used on the WWW, and the message digest based file identification used for file sharing in many P2P networks.

A CDN represents a realm for the autonomic architecture. Such realms might organize content in different ways, but for scalability reasons, it is important to be able to optimize the data transfers. Autonomic architecture is able to self-optimize using the knowledge gathered through the monitoring infrastructure and disseminated by processes of information and knowledge management. Thus, a CDN realm could self-optimize the content distribution in terms of latency or available bandwidth. An autonomic CDN realm would have also the ability to self-organize the nodes within the realm, which means that management of the CDN, e.g. joining and leaving of nodes and locating nodes that share particular content, are basic functionalities provided by the architecture.

An obvious additional advantage from the use of autonomic realms is the ease of application development. Each application no longer necessarily needs to individually optimize the data transfers, e.g. select peers based on bandwidth and/or delay measurements in a P2P-based CDN, or locate content. Application could simply join the CDN realm and request content from it. In this way the application would use a sort of anycast addressing for requests. In addition, the optimization tasks (e.g. available bandwidth measurements), currently performed by each individual application itself, would not need to be done in such a redundant manner, and the information could be reused efficiently within the realm or even shared across realms.

5 Conclusions

In this paper, we first summarized architectural problems that exist with the Internet. Afterwards, we studied the future developments in two stages. First, we took a top-down approach and analyzed the future workload of the Internet. It should be emphasized that this study is based on an analysis of the different stakeholders of the Internet today and in the future. Besides a long list of requirements, it mainly showed that requirements from different stakeholders and applications are quite different and often contradicting. The heterogeneity with respect to the needs of a future Internet architecture is also documented in our bottom-up study, where we analyze the future developments of networking devices and technologies. We conclude that there will be not a "single Internet" in the future. In particular, we have identified a set of fundamental architectural concepts for a future network architecture:

- Multiple realms must be supported concurrently. To achieve this, a simple bootstrap like process that allows to bootstrap multiple different realms needs to be developed.
- Realms must be able to host other (partial) realms and provide means for inter-realm communication.
- A monitoring facility has to be part of the architecture. It must be able to potentially monitor all data and events of interests and provide means for storing monitoring data.
- Information and knowledge management is needed in order for the network to be able to provide a kind of "information plane".

An architecture that implements these fundamental architectural concepts should be able to address all requirements that have been mentioned in this document. Obviously, we do not claim that the ANA project can solve all the problems and meet all the needs of today and of the future. However, we hope to be able to demonstrate within the scope of ANA that an autonomic network architecture can lead to realms that have the following properties:

- It is possible to design highly scalable networks with the ANA architecture that are not restricted by address spaces or other means.
- The ANA architecture can support different realms that fulfill entirely different sets of application and end user requirements.
- Realms can be designed in such a way that they are extensible and can easily include new solutions without breaking the network architecture and design.
- Monitoring infrastructure and knowledge management support the need for resilient networks.
- The information and knowledge plane enables easier self-configuration and self-organization, e.g. through better service discovery solutions.
- Information and knowledge plane leads to better solutions for self-optimization, e.g. for the case of structuring communication into uni-cast, any-cast, and multi-cast communication for efficient resource utilization from networks viewpoint.
- Mobile users and mobile terminals can be better supported than with the Internet today.

Proceedings of the 11th IEEE International Workshop
on Future Trends of Distributed Computing Systems (FTDCS'07)
0-7695-2810-4/07 $20.00 © 2007 IEEE

References

[1] Austrian universal service regulations. http://www.bmvit.gv.at/telekommunikation/recht/aut/verordnungen/udv.html.

[2] Internet traffic report. http://www.internettrafficreport.com/.

[3] Nsf nets program's future internet design (find) initiative. http://find.isi.edu/.

[4] Study of the security of Voice over IP. http://www.bsi.bund.de/literat/studien/VoIP/index.htm.

[5] Autonomic network architecture (ANA). Integrated Project FP6-IST-27489. Sixth Framework Programme - Situated and Autonomic Communications (SAC). http://www.ana-project.org, Oct. 2005.

[6] Future internet & networking infrastructure. Project Description, http://www.hpl.hp.com/research/issl/projects/network/index.html, 2006.

[7] Requirements – Deliverable D.1.2 of Autonomic Network Architecture Project. Integrated Project FP6-IST-27489. Sixth Framework Programme - Situated and Autonomic Communications (SAC), Nov. 2006.

[8] State of the Art – Deliverable D.1.1 of Autonomic Network Architecture Project. Integrated Project FP6-IST-27489. Sixth Framework Programme - Situated and Autonomic Communications (SAC), Aug. 2006.

[9] B. Ahlgren, M. Brunner, L. Eggert, R. Hancock, and S. Schmid. Invariants: a new design methodology for network architectures. In *FDNA '04: Proceedings of the ACM SIGCOMM workshop on Future directions in network architecture*, pages 65–70, New York, NY, USA, 2004. ACM Press.

[10] D. Andersen, H. Balakrishnan, F. Kaashoek, and R. Morris. Resilient overlay networks. In *SOSP '01: Proceedings of the eighteenth ACM symposium on Operating systems principles*, pages 131–145, New York, NY, USA, 2001. ACM Press.

[11] R. J. Anderson. *Security Engineering: A Guide to Building Dependable Distributed Systems*. John Wiley & Sons, Inc., New York, NY, USA, 2001.

[12] T. Anderson, L. Peterson, S. Shenker, and J. Turner. Overcoming the internet impasse through virtualization. *Computer*, 38(4):34–41, 2005.

[13] W. R. Cheswick, S. M. Bellovin, and A. D. Rubin. *Firewalls and Internet Security: Repelling the Wily Hacker*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2003.

[14] D. D. Clark, C. Partridge, R. T. Braden, B. Davie, S. Floyd, V. Jacobson, D. Katabi, G. Minshall, K. K. Ramakrishnan, T. Roscoe, I. Stoica, J. Wroclawski, and L. Zhang. Making the world (of communications) a different place. *SIGCOMM Comput. Commun. Rev.*, 35(3):91–96, 2005.

[15] C. D. Cranor, T. Johnson, O. Spatscheck, and V. Shkapenyuk. The gigascope stream database. *IEEE Data Eng. Bull., 26(1)*, 2003.

[16] T. Ernst. The information technology era of the vehicular industry. *SIGCOMM Comput. Commun. Rev.*, 36(2):49–52, 2006.

[17] K. Fall. A delay-tolerant network architecture for challenged internets. In *SIGCOMM '03: Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 27–34, New York, NY, USA, 2003. ACM Press.

[18] R. Gold, P. Gunningberg, and C. Tschudin. A virtualized link layer with support for indirection. In *FDNA '04: Proceedings of the ACM SIGCOMM workshop on Future directions in network architecture*, pages 28–34, New York, NY, USA, 2004. ACM Press.

[19] B. Jacob, R. Lanyon-Hogg, D. Nadgir, and A. F. Yassin. A practical guide to the ibm autonomic computing toolkit. Technical report, IBM International Technical Support Organization, 2004.

[20] N. McKeown. How emerging optical technologies will affect the future internet. Talk at NSF FIND Meeting, Dec. 2005.

[21] I. Miladinovic, K. Umschaden, T. Höher, M. Banfield, W. Bauer, and P. Tschulik. Ipv6 deployment scenarios for large isp networks. In *Proceedings of the Third IASTED International Conference on Communications and Computer Networks (CCN 2005)*. ACTA Press, Oct. 2005.

[22] A. Mishra and K. M. Nadkarni. Security in wireless ad hoc networks. pages 499–549, 2003.

[23] R. Moskowitz and P. Nikander. Host Identity Protocol (HIP) Architecture. RFC 4423 (Informational), May 2006.

[24] V. Paxson. Addressing the threat of internet worms. UCLA Jon Postel Distinguished Lecturer Series, Feb. 2005.

[25] A. S. Pentland, R. Fletcher, and A. Hasson. Daknet: Rethinking connectivity in developing nations. *Computer*, 37(1):78–83, 2004.

[26] T. Plagemann, V. Goebel, A. Bergamini, G. Tolu, G. Urvoy-Keller, and E. W. Biersack. Using data stream management systems for traffic analysis - a case study. In *Passive and Active Measurements 2004*, April 2004.

[27] D. Raychaudhuri, J. Evans, , and D. Estrin. Future internet and experimental facility design: Wireless, mobile & sensor aspects. Talk at NSF Wireless/Mobile Planning Group Workshop, July 2006.

[28] S. Schmid, M. Sifalakis, and D. Hutchison. Towards autonomic networks. In *Proceedings of the 3rd Workshop on Autonomic Communication (WAC 2006)*, Sept. 2006.

[29] F. Sestini. Situated and autonomic communication an EC FET European initiative. *SIGCOMM Comput. Commun. Rev.*, 36(2):17–20, 2006.

[30] M. Siekkinen, E. W. Biersack, V. Goebel, T. Plagemann, and G. Urvoy-Keller. Intrabase: Integrated traffic analysis based on a database management system. In *Proceedings of IEEE/IFIP Workshop on End-to-End Monitoring Techniques and Services*, May 2005.

[31] J. P. G. Sterbenz, R. Krishnan, R. R. Hain, A. W. Jackson, D. Levin, R. Ramanathan, and J. Zao. Survivable mobile wireless networks: issues, challenges, and research directions. In *WiSE '02: Proceedings of the 3rd ACM workshop on Wireless security*, pages 31–40, New York, NY, USA, 2002. ACM Press.

[32] U. Varshney, A. Snow, M. McGivern, and C. Howard. Voice over ip. *Commun. ACM*, 45(1):89–96, 2002.

Proceedings of the 11th IEEE International Workshop
on Future Trends of Distributed Computing Systems (FTDCS'07)
0-7695-2810-4/07 $20.00 © 2007 IEEE