October 2007

ICT SHOK Future Internet — Research Agenda

Executive summary

The ICT SHOK Future Internet programme forms a vital, integral part of the global research effort taking place in Europe, US, China, and elsewhere. We expect that by becoming early developers and adopters of the Future Internet, Finnish companies and academia will gain significant benefits in new markets and scientific results.

Two major interlinked changes in the Internet are presently challenging its structure and reshaping its future: the move from the end-to-end principle towards a trust-to-trust principle and from interconnecting nodes towards interconnecting information. Other main change drivers include connectivity, scalability and autonomous resilience.

Several present bottlenecks constrict the future progress of the Internet: unwanted traffic, choking of the routing system, mobility and multi-homing, compensation and congestion, privacy and attribution, and trust and reputation. The major challenges blocking its future development include information networking, energy consumption, changing usage patterns, and the impact of shifting bottlenecks.

By solving these problems and addressing these challenges, we progress towards the vision of **Future Internet as critical backbone of society connecting users and information**. Hence, the mission of the programme is to combine efforts of Finnish companies and academia to make a significant contribution towards **developing the Future Internet technology and ecology as a platform for innovation**, especially focusing on network and information governance and leveraging mobility as a key source of competitiveness and global added value.

To reach the vision, the Future Internet programme must be implemented as a coordinated effort of the participating companies and research groups to ensure maintenance of shared vision and focus, efficient resource utilisation, information flow, and impact. For this, we also propose a number of horizontal actions of the entire programme, designed to complement the individual research lines. These include work on key application drivers and deployment scenarios, and also joint testbeds and larger-scale experiments.

Contributors

The editors of the present version are Pekka Nikander and Martti Mäntylä, with valuable help from Christer Carlsson, Jyrki Huusko, and Pasi Lassila. The technical background comes mainly from Hannu Flinck, Pekka Nikander, and Pasi Sarolahti, with contributions from Christer Carlsson, Tapio Frantti, Matthias Grossglauser, Jussi Kangasharju, Raimo Kantola, Jouni Korhonen, Kari Kuutti, Pasi Lassila, Jukka Manner, and Pekka Savola. Some pieces of the text were adapted from a joint paper written by Dirk Trossen, Sasu Tarkoma, Mikko Särelä and Pekka Nikander.

Table of Contents

Introduction

During the years when the technology that later became called the Internet was initially created, Finland was its early adopter both in academia and in industry. As a result, Finland's size on the Net has ever since been disproportional to its population or geography. The direct and especially indirect impacts of this attractive and serendipitous circumstance have been critical to the country's recent performance in information and communications technology based economy.

Today, Internet technology and applications have thoroughly penetrated business and personal life of nearly all developed countries, and are making inroads everywhere else. Yet the progress we have witnessed so far is just the beginning. What looms ahead is an Internet that will merge with mobile communications and the "real world", for the first time opening the door of a truly global information network reachable by the whole population of the planet.

For Finland, which can count itself as one of the winners of globalisation (a phenomenon largely driven by the progress of the Internet) this situation is both attractive and challenging. Finnish research and industry are in a good position to become early developers and adopters also of the next generation of the Internet, and thereby to continue harvesting the rich rewards it will bring. The challenge stems from the success of the present Net: the entire global academic and industrial Internet community is likely to get involved in the effort to create the successor Internet. Thus, stakes will be higher this time.

The Future Internet — just like the Internet today — will be a global network. Consequently, research on the area will necessarily be a collaborative effort, drawing ideas from the rest of the world and pushing aggressively its own views and innovations abroad. Consequently, the ICT SHOK Future Internet programme must not be a standalone effort but a vital, integral part of the work taking place elsewhere in Europe, as well an in the US, China, and the rest of the world. In practical terms, our well established connections to the NSF NeTS FIND and GENI programmes, the related EU FP7 projects, the New Generation Network (NWGN) programme in Japan, and the Tsinghua University and Chinese next generation Internet programme must be tightly integrated within the ICT SHOK umbrella, and reinforced with strong, personal-level contacts with other major programmes around the world. To achieve this, **Finnish research groups in companies and academia must be combined in a co-ordinated programme that reaches the critical mass and can maintain the close long-term partnership with related activities elsewhere**. This is the core of the research strategy of the Future Internet programme.

The rest of this document is organised as follows. First, we outline the main drivers for the currently perceived need for change in the Internet. Next, we formulate the vision and mission of the programme, and also discuss the key elements of the research strategy needed for achieving the mission. After that, we discuss a number of present major problems in more detail, and suggest research actions of different time spans for addressing each. In the next section, we outline some anticipated major future challenges that are likely to require long-term fundamental research. Finally, we describe an action plan, designed to make progress towards the mission and vision while maintaining the focus and transparency of the programme, and facilitating the impact of its results.

Background

The original Internet was designed for connecting computing devices into a global web of computers. Ever since its dawn, the advances of the computing technology and its applications have guided the evolution of the Internet. This section reviews key current computing trends that will have a major impact on the Future Internet.

End-to-end connectivity

At the present, mobile computing is transforming mobile phones into multimedia computers that need high bandwidth access to content, Web 2.0 applications build on the assumption of always available connectivity to network embedded storage and server applications, and the emerging sensor applications are broadening the scope of the Internet into the physical world. Unfortunately, present Internet hardly lives up any longer to the assumption of end-to-end connectivity that underlies these developments.

The "classical" Internet was built on the **end-to-end principle (E2E)**, where the network connecting the end-hosts and users was only performing packet routing. The shortage of IPv4 addresses, lack of mutual trust, and various business reasons have resulted in the rise of network middle boxes,

i.e., devices, other than bridges and routers, that lie on the communication path between end points. Examples include Network Address Translators (NATs) and firewalls.

While middle boxes have benefits, they also have drawbacks, often architectural ones. Firstly, middle boxes break the end-to-end principle. In a typical case, the network implements functionality that in principle could have been implemented in the end host; however, following that principle typically would require changes to the overall architecture, and therefore may not be economically and practically feasible. From the end-to-end viewpoint, the middle box creates a point of failure in the communication path: if the middle box crashes, end-to-end communication cannot continue even if an alternative communication path were available. Each middle box also creates a potential performance bottleneck and may make deployment of new applications and services harder. Some middle boxes plainly and simply break protocols. While this may be intentional and beneficial to the network owner, it is typically harmful to the end user.

Secondly, middle boxes mix poorly with security. For end-to-end security, either a middle box must allow encrypted traffic to flow essentially unchanged through itself, thereby reducing its benefit, or the trust model has to be changed in a fundamental way so that the middle box may know the relevant cryptographic keys. In general, middle boxes change the trust model, especially if the middle box is located within another domain.

Thirdly, a huge amount of work has been invested within the IETF and elsewhere to counter the problems of middle boxes. Practically all protocols designed today, and most of the old specifications, need to consider middle box traversal, at a considerable cost in more complicated engineering. With the available IPv4 addresses running out in a few years, middle boxes will only increase in number, and probably in complexity. The global deployment of IPv6, when it finally happens, will not help: the IETF proposal of stateful IPv6 firewalls introduces a new middle box to harm IPv6-based protocols, effectively requiring solutions similar to IPv4 NAT traversal.

From end-to-end to trust-to-trust

Trust-to-trust

As mentioned above, the end-to-end principle stipulates the execution of functionality in endpoints of communication, following secondary principles like minimality, generality, simplicity, and openness.

When looking at many deployments of communication services in today's Internet, it becomes apparent that placement of the execution of functionality is crucial when designing solutions to be deployed. This, in effect, has led to the appearance not only of middle boxes, but also services that are largely implemented at trusted 3rd parties rather than in the endpoints of the actual service (from a user's perspective).

The realisation that the place of executing functionality properly and in a trustworthy manner is crucial when designing an overall solution to a given communication problem has been clarified by David Clark as the **trust-to-trust (T2T)** principle:

> *"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."* — David Clark, MIT Communications Futures Program, Bi-annual meeting, May 30-31, 2007, Philadelphia, PA.

For instance, looking at a firewall, in the light of T2T, as a component that implements traversal reception in a large enterprise network in a trustworthy manner (from the perspective of the enterprise) makes it a part of the entire end-to-end solution rather than a hack that was inserted after the fact.

However, there is another, more subtle, shift that occurs when moving from E2E to T2T. The original end-to-end principle has a strong focus on the notion of endpoints. There is an underlying, usually implicit assumption that the recipient is willing to receive whatever the sender is sending, or else the sender would not send it in the first place or will cease very quickly. Obviously, this assumption is no longer true.

As stated before, the careful clarification of E2E — in its form of trust-to-trust — seemingly shifts the focus away from (network) endpoints to (trustworthy) points of execution. This opens up a discussion whether a network-endpoint-oriented paradigm, such as IP, is still well suited under the refined T2T principle. Also, as argued for instance by Van Jacobson, many communication scenar-

ios are not focused on endpoint communication rather than the retrieval of information from (trustworthy) sources that can provide it. This observation, together with the T2T clarification, seems to call for revisiting the design of the network endpoint oriented IP solution.

Against this background, one cannot overemphasise the importance of new, controlled openness. Only by moving towards a network that is simultaneously open and genuinely trustworthy can one efficiently support new applications and services. In order to reach such a situation, it is necessary to understand the economic and other tensions between providers and consumers on one hand and between different types of providers on the other hand. Such tensions must be made transparent, preferably in the form of creating new open and level market places, so that they can be internalised in the form of market-based pricing or other compensation mechanisms.

Nature of networking

Changing nature of networking

Apart from trust-to-trust, the second crucial challenge that motivates our work is related to the changing nature of networking.

Instead of mostly being a point-to-point network, aimed at conversational applications between hosts such as telnet, file transfer, or generally client-server applications, the present Internet is dominated by informational applications where users are concerned with finding and accessing certain content, and often do not really care which network host the content happens to reside in. Thus, we are witnessing a transition of the Internet from connecting nodes to connecting information. Unfortunately, the present Internet is at odds with this changed usage pattern; for instance, securing the path between two communicating nodes is solving the wrong problem if the user is primarily concerned with the authenticity of a piece of content.

Ubiquity

Ubiquity

A further challenge results from the ubiquity and penetration of the Internet, with most users being non-professionals using the Internet for the widest variety of personal and professional purposes. Critical themes brought to the limelight by this include usability issues such as zero-configuration (users must be able to install new services themselves without assistance from a service vendor or operator), personalisation, and turning the complexity of the Internet and its services to a resource that users can manage and adapt for their variable goals. Also the wider issues related to social impacts of networked communication and information exchange on people, communities, and societies are equally critical.

Scalability

Scalability

The Internet as we know it is one of the most complex artefacts humanity has ever built. Nevertheless, its reach is still continuing to grow with the expected convergence of voice and data (Telco and IT) traffic that is setting the capacity, performance and usability requirements to a new level. In the future, when moving towards more ubiquity and shifting attention from computing platforms to information objects, the level of complexity is likely to grow by several orders of magnitude. Hence, the network must be built from the beginning with scalability in mind. Everything must scale; therefore, we do not explicitly mention scalability below but expect that each and every aspect of the network must necessarily scale. To give some ballpark figures to get an understanding of the scale of the complexity, we assume in the order of 10^{10} autonomous domains and order of 10^{15} information objects.

Scalability is likely to be the main challenge in creating the future Internet. As mentioned above, scalability does not solely refer to the number of computers in the network, but instead encompasses all the computers, network equipment, user devices, users, and all the information stored anywhere in the network. In particular, owing to the increasing amount of user generated content, the number of information objects on the future Internet will grow at tremendous speeds. This places severe challenges on how to locate all this information, and requires scalable methods for naming and addressing of objects, as well as locating them in the network. Experience has shown that a hierarchical system of naming, as practised on the current Internet, does work well in solving

many scalability problems. An open question is how much of that can still be applied and what needs to be created anew.

A related challenge is the apparent end of the applicability of the famous Moore's law at least for processor cores. For some years, a combination of technical and business reasons have driven the development of processor chips no longer towards faster CPU cores, but instead to multi-core architectures that can benefit from the continuation of Moore's law on transistor level. This is likely to change the content of the toolbox by which scalability issues must be dealt with from faster single-core algorithms facilitated by fast caches to parallel algorithms that map well on multi-core architectures.

Availability, reliability and dependability

Availability, reliability, dependability

New business models are emerging that build on the access independency, global reach, huge existing content, and the wide application and developer base of the Internet. With this, it is becoming a critical resource that public bodies, companies and institutions rely upon in performing many of their vital functions. As popularised by Chris Anderson[1], even new forms of business can emerge based on the almost cost-free inventory of digital products which can be accessed through the Internet. This requires completely new levels of availability, reliability and dependability.

Internet was designed based on the idea of distributed control and adaptivity, expressed in the routing protocols and TCP, which make the system rather robust. Indeed, for Internet routing it is known that the properties of the AS level graph guarantee that the likelihood of network failure is very small. While this provides some assurance of operability, it is still an open question how to define more precisely the reliability properties of the whole system, even including the services and their vulnerability to hostile attacks. Furthermore, in the future it must be possible to design and engineer systems (network topologies, services) to meet given levels of reliability and dependability.

Many original Internet services were designed as federated applications where autonomous nodes co-operate to provide a service infrastructure; examples include e-mail, news, and IRC. While individual nodes are prone to service breaks, this design principle makes the whole system very robust and scalable. More recently, many Internet services, including ones that might be considered infrastructure level, have departed from this tradition and assumed the client-server model instead. While there are sound business reasons for this, from availability, reliability, and dependability viewpoint this is counterproductive. Therefore, rethinking the underlying design model may be needed.

Vision and mission

Internet as critical backbone of society

With the above developments in mind, we formulate the underlying vision of the Future Internet, motivating the work detailed in this research agenda, as follows:

> *Future Internet will become a mission critical backbone of global information society with billions of mobile and wire line users instantly connected to information and each other, and using the Internet to communicate, conduct business, manage their everyday lives, express themselves, and enjoy entertainment.*

[1] *The Long Tail*, RH Business Books, Cambridge 2007.

As described, Internet will be a backbone for the society on even a larger scale than today, and thus have profound effects on the economy by driving down the transaction costs (especially search and trust establishment related costs), increasing the agility of markets, facilitating innovations for a growing diversity of products and services, and generally transforming almost every conceivable activity or phenomenon into an informational entity that can be tracked, memorised, and acted upon. This means that the Internet becomes a critical infrastructure, a complete collapse of which would have dire consequences. In a sense, the future Internet will bring about an updated version of Marc Weiser's original vision of ubiquitous computing. In this updated view, the Internet cannot be outside of the society anymore; it must become a full member.

Consequently, the mission of the ICT SHOK Future Internet programme is stated as follows:

> *Enhance the Internet technology and ecology as a platform for innovation while providing strong governance over the use of the network resources and information in such a way that especially mobile use of the network and its services will be natively supported.*

Innovations, governance, mobility

We view **innovation** as the key driver for a market-based economy and the key enabler for a robust and durable development of the new Internet, creating a beneficial circle where societal gains in productivity fuel the investments. At the same time, **strong governance** is critical for securing future investments (network part) and creating prerequisites for new services (information part). Under Finnish circumstances, understanding and leveraging **mobility** is a key source of competitiveness and global added value.

To make future Internet an attractive platform for innovations, control, and mobility, especially the following aspects are critical, and thus drive the entire research agenda:

- *Multifaceted trust:* The original Internet was based on the assumption that various co-operating parties are trustworthy and behave accordingly. Unfortunately, this is no longer true, nor will it be in the Future Internet. Lack of trust leads to losses of opportunities, higher transaction costs, and ultimately loss of potential welfare. Therefore, trust in its various forms must be a fundamental thread that runs through the entire research programme.
- *Leveraging situational complexity:* In the past, mobility was often equalled with the concepts of universal reachability of mobile terminals and universal access to essentially the same services independent of time and location. In the future characterised by the ubiquitousness of the Internet and digital services, we expect mobility to express itself in widely more personalised and localised forms, enriching the variety of times and places rather than making them increasingly the same.
- *Emergent properties - reliability and surprising new invariants:* Data analysis and modelling are crucial for discovering new invariants in the global-scale behaviour of the Internet. By identifying invariants that, *e.g.*, limit the scalability of the system, the architecture and design of the Future Internet can be improved. As the Internet is also becoming a critical resource for the whole society, the reliability properties of complex distributed systems, such as the Internet, must be better understood. The goal is that the Future Internet can be designed and engineered to meet given levels of reliability and dependability.

Research strategy

The overall vision and mission of the Future Internet programme are challenging and wide-reaching. To make progress while maintaining the focus and impact of the work requires a balanced portfolio of activities aiming at short and medium term goals that pave the way towards the long term goals.

Therefore, in the subsequent sections we describe the research efforts by detailing themes and objectives with three time ranges:

- *Short term:* This indicates work whose results should be available in 1-2 years. These activities show that the work is progressing and facilitate its impact, confirm that the basic technology needed for the longer term goals is feasible to implement in practice, gather experimental data on

current leading edge technologies, and gain additional understanding of the related issues and side effects.

- *Medium term:* The main focus of the program will be on research which should give applicable results in 3-5 years. Topic for a PhD thesis (at least from university perspective). This work develops larger architectural parts and concepts, demonstrates their feasibility, and tests their applicability in naturalistic settings.
- *Long term:* This work provides the overall vision, "where do we want to be in 10 years?", sets the goals and direction for the short and medium term projects, and provides fundamental understanding, theories, models, and methods for the rest of the work.

In particular, the ICT SHOK Future Internet portfolio should include work that identifies the real life problems of the current Internet and its usage, contributes to the standardisation, and partially verifies and provides feedback to the new mid and long term concepts developed in parallel.

As detailed below in the action plan section, the research programme must be complemented with various "horizontal actions" to maintain the focus, transparency, and competence basis of the programme. With this, **Finnish research groups in companies and academia will be combined in a co-ordinated programme that reaches the critical mass and can maintain the close long-term partnership with related activities elsewhere in the world**, thereby giving the programme the best chances of success.

Present problems

Present problems

1. Unwanted traffic
2. Choking of the routing system
3. Mobility and multi-homing
4. Compensation, Resource consumption, and Congestion
5. Privacy and Attribution
6. Trust and reputation

The issues discussed in this section are present problems of Internet that already act as bottlenecks of its continuous deployment and growth.

Unwanted traffic

The various forms of unwanted traffic, including spam, distributed denial of service attacks, and phishing, is arguably the biggest problem in the current Internet. The root reasons to unwanted traffic seem to be best characterised with economics.

We can view the current Internet as a global, distributed system where the main cost of communication is paid by the recipient. This is a direct (though certainly unintentional) consequence of the network architecture. By explicitly and directly naming all the potential recipients, we create a system where the senders can easily express their desire to send data to any recipient in the network. Given that under the typical contracts the marginal cost of sending additional packets is very close to zero (up to some capacity limit), there are few or no incentives for refraining from sending unwanted traffic; sending some packets just for fun costs so little that it doesn't matter. At the same time, even a marginal response rate creates a strong incentive for sending unsolicited advertisements, and even a small success rate creates a strong incentive for DDoS-based extortion.

The current practice to fight unwanted traffic is to add more middle boxes. However, they cause a number of reachability problems, creating a significant challenge in today's network design. Instead, we need some form of controlled reachability, preventing the unwanted traffic from consuming network resources, especially over wireless link, but at the same time allowing flexible use of different servers, services, and protocols also behind the wireless links. Thus, unwanted traffic is not merely a sender-recipient issue, but also a fairness issue among the traffic flows, *i.e.*, how and to whom to distribute the cost fairly, and thereby connected with the congestion control and compensation issues.

- Near term: protocols to interact directly with NATs and firewalls; unwanted traffic on other layers, such as spam mail or unwanted IM spam messages
- Medium term: communication frameworks for preventing unwanted traffic arriving from upstream; end-hosts to signal what services they provide and what traffic they are interested in.
- Long term: economic solutions; completely new architectures.

Choking of the routing system

The second big problem can be described as relative choking of the routing system. The current Internet routing system relies solely on the Border Gateway Protocol (BGP), a protocol that has received some facelifts but internally has remained the same for the last decade or so. At the same time, the business environment where the Internet Service Providers (ISPs) compete has become immensely more complex and competitive. From the operational point of view, the main technical term characterising the current routing complications is traffic engineering. A basic fact is that the current routing system, *i.e.*, primarily BGP, does not offer any good facilities for it — nearly all of the various ways that the ISPs attempt to perform traffic engineering can be considered as protocol violations or hacks relying on obscure side effects of the routing mechanisms.

A potentially even bigger problem appears to be routing table convergence. For example, there are indications that the routing system may fluctuate days or even weeks after major events affecting the links, such as a recent undersea earthquake near Taiwan that cut a handful of undersea communication cables. At the same time the non allocated IPv4 address space is projected to exhaust in some two-to-three years, creating urgency for migrating to IPv6. The transition period will stress the routing infrastructure even further as two different address spaces need to operational. Consequently, the Internet research community has started to develop new approaches to meet the scalability and traffic engineering needs of the core Internet.

Together, these issues should be taken as early warnings, indicating that our current routing system may be near its inherent capacity limits. With further growth, the ISPs may no longer be able to perform effective traffic engineering, probably leading to some market consolidation and loss of competition, and the routing system may start to experience global-scale instabilities making large fractions of the Internet unavailable for excessive time periods.

Specific research items in this area include the following:

- Near term: IPv6 migration; Renews role of NATs; BGP housekeeping to deal with the immediate scalability needs.
- Medium term: Extended link layers; cross-layer interactions; pressure from mobility and multi-homing; pressure from overlay networks.
- Long term: Compact routing; Hierarchical or recursive routing; Understanding the consequences of power law or long tail topologies.

Mobility and multi-homing

The current Internet architecture was not developed with mobility in mind. Today, there is clear need for combined mobility and multi-homing support at different levels of granularities ranging from users and applications to complex subnetworks. Effective mobility support requires a level of indirection, something that the current Internet architecture is gravely missing at most potential mobility points. Adding suitable indirection points is likely to require architectural changes.

Already today, a typical wireless end-host can use several network access technologies to get connected to the network. In addition, rather than just a single publicly open “Internet”, the world consists of a number of isolated network islands. A wireless terminal may see a number of available network access points using different technologies, and without making an initial attachment by which it gets additional information, a wireless host has very little information about the identity and characteristics of the networks behind the access points. A further complexity is that in addition to the physical access interfaces, different types of virtual interfaces or VPNs can be used to access a specific network domain.

Furthermore, many of the current services contain several layers of network protocols each with their own attachment mechanisms. Starting from link-layer procedures and IP address resolution, a large number of layers might each require some number of hand-shake messages to establish connection. These handshakes could be parallelized. This, nevertheless, involves security challenges and requires heuristics to roll back established communication state if the handshake is terminated on some layer for some reason.

However, all these solutions target physical nodes moving on the Internet, and as a consequence also tie information and devices together. There are many other entities that would like to enjoy the same level of nomadic behaviour, for example, users, communication sessions, transport protocols, user profiles, software agents, user interfaces, and even application processes. People want to

switch between different computers and still have the same profiles and data available, VoIP sessions and the media carried could migrate from device to device, or application processes or even single threads could be run on other devices. The current networking model of the Internet and the protocols tie the physical node and data together, prohibiting this wider view of mobility, and multihoming.

- Near term: performance on mobility, hand-offs, and multi-homing using the existing protocols; means for the end hosts and networks to select the best possible access interface; fast hand-offs.
- Medium term: context awareness; controllable disconnections and reconnections, intermittent connectivity; isolated network domains; different network characteristics; seamless behaviour; working seamlessly in a mixed IPv4 / IPv6 environment; new forms of mobility beyond device/host mobility.
- Long term: infrastructureless networks; new compensation mechanisms.

Compensation, resource consumption, and congestion

The current Internet architecture is not well equipped to deal with short-term resource shortages. That is, while the TCP and other related well-behaving protocols courteously back up and reduce their resource consumption in the face of congestion, there is nothing in the architecture itself that enforces such behaviour. Our basic claim is that the lack of resource and congestion control in the Internet is inherently related to the lack of compensation methods.

The Internet was fundamentally built upon the idea of at-least minimally co-operating agents. In the very core of the Internet architecture lies the assumption that if a host does not want to receive traffic, the sending nodes will cease submitting it. Similarly, almost as close to the core, lies another strong assumption: the hosts and routers will co-operate in getting the maximal amount of traffic through for the maximum number of hosts. These two assumptions are engraved deep into the architecture and the implementations.

Looking from an economic point of view, and assuming selfish rather than co-operating agents, both of the core assumptions start to appear silly. If the senders are not forced to somehow compensate for the resources they use, or at least overuse, there is no way of creating incentives for stopping them.

Looking at the two problems (lack of resource control and lack of congestion control) more closely, we can see that they are actually the same problem, only working on different time scales. The resource control mechanisms attempt to make sure that there are sufficient resources at all times. The congestion control mechanisms attempt to make sure that the available resources are allocated according to some "fairness" or "relative utility" principle in those cases where the demand exceeds the supply.

From a more technical point of view, the challenges include the following:

i) how to satisfy performance requirements of different kinds of new networking applications, such as Internet telephony, video streaming, and games, and at the same time implement feasible and fair congestion control in a best effort network,

ii) how to make the congestion and rate control algorithms more reactive to the changing network characteristics, and

iii) how to address other types of resource limitations than available bandwidth and buffer space, such as available power in a battery-powered device.

The possible research work includes:

- Near term: external triggers, make the congestion control more reactive to the sudden changes in the network environment; challenges related to using newly proposed congestion control algorithms.
- Medium term: incentives related to congestion control; ways to encourage correct behaviour; mechanisms to provide increased flexibility to the applications
- Long term: fine-grain adjustments in the flow transmission rates; new types of compensation methods; disentanglement of the functions of money; currencies based on other models besides

fractional reserve banking; structured market places to internalise the most obvious tension points.

Privacy and attribution

The final two current problems are somewhat different than the ones above. Here, we are interested in preventing bad things from happening, in one hand by imposing restrictions on information flow, and in the other hand by creating explicit incentives for trustworthy behaviour.

The privacy problem can be described from a three different starting points. First, from the Orwellian point of view, the question hinges on freedom of speech and governmental control. Second, the Kafkaesque aspect of privacy focuses on citizen's ability to retain their autonomy without fear of unfounded harassment. Thirdly, the economic aspect of privacy relates to the fine balance between socially beneficial differentiated pricing vs. socially harmful price discrimination. From these points of view, it seems a necessity to provide a reasonable base-level of privacy as a built-in feature in future networks.

Thus, privacy can be seen as regulating the interaction between an individual and the various actors in her environment, such as protecting an individual against unsolicited messages, or from seeing offending contents. Thus privacy protection can be seen to be linked to the issue of dealing with unwanted traffic. Indeed, unwanted traffic protection at network layer will be sufficient to guard some forms of privacy, while anti-spam protection will provide more extensive coverage.

The flip side of privacy is attribution. Unbounded privacy encourages unwanted side effects, such as rampant advertising (spam). To counter these, increased privacy requires increased accountability. How to provide them at the same time, and in correct balance, appears to be a hard problem.

- Near term: Experimental/empirical research to expose privacy issues in various domains
- Medium term: Identities with balanced and strengthened privacy and attribution
- Long term: Economics of identity, privacy, and attribution. The technical solutions are relatively easy (but not trivial); the hard part is to understand how they will interact with the rest of the system.

Trust and reputation

The final problem we consider is the lack of trust and reputation. The original Internet architecture was built with a fairly homogenous, mutually-trusting community in mind, thus assuming that all users, nodes and software enjoy a high level of mutual trust. Today the user community is very large and diverse. The hosts cannot be trusted to respect protocol specifications any more, due to prevalence of botnets and other malware. More generally, we postulate that the current network and applications suffer from the lack of standardised, wide spread mechanisms for asserting trust and reputation.

Technically, facilitating trust and reputation can be based on applying suitable communications and information security methods to establish the authenticity and integrity of communicating parties and the information they transfer. Histories of such transactions can be collected and analysed to facilitate inferences on parties' past behaviour and their level of trustworthiness.

While the technology for these purposes is not trivial, the problems of trust and reputation are largely non-technical. Monitoring and recording the behaviour of parties obviously creates risks of privacy problems that users must be made aware of and which they should have the means to regulate. Another legal / economical issue is that data may need to be collected by one party to be applied by another. Even when this is possible from the regulation viewpoint, it remains unclear how the parties should be given incentives to co-operate.

The most difficult problems of trust and reputation relate to human factors. How do people make decisions involving various kinds of risks? How do they use available information to assess their options? What information, in which form, and at which time is most likely to be useful in the sense of improving the user-perceived success of the decisions? The answers are likely to be different in various trust domains where the risk profiles and incentives are different.

- Near term: Experimentation with various approaches to trust and reputation management in various trust domains, especially when reinforced with social networks. Application of the prin-

ciples of usable security to trust and reputation systems: how should trust-related information be exposed to users, so that they can best act on it?

- Medium term: New concepts and technologies for expressing and inferring on reputation and trust in various trust domains.
- Long term: Understanding and modelling human decision-making in different domains involving trust; using this to design informational support most effective for decision making and compensations that should be distributed fairly amongst the participants to create a positive feedback loop.

Major future challenges

This section discusses themes that seem to form major roadblocks on the possible future evolution of the Internet in the sense that unless they can be addressed, major potential business and societal benefits cannot be fully realised. Unlike the previous section, we do not suggest individual research topics here, as the major future challenges are likely to require fundamental research.

Major challenges

1. Information networking
2. Usability and usage patterns
3. Network socio-economics
4. Autonomy and resilience
5. Energy consumption
6. Shifting bottlenecks

Information networking

The network today is no longer used just to convey written or otherwise recorded messages or pass real-time voice or video, but also for entertainment, *e.g.*, interactive gaming, and more generally for creating a new kind of social consciousness. The younger generations consider it normal to be "online" all the time. More recently, the usage patterns have shifted towards sharing presence and experiences; *i.e.*, using the network to keep track of the whereabouts of friends and acquaintances, allowing new kinds of social structures to emerge.

At the same time, the proliferation of intelligent, networked devices and search services is gradually making it impractical to identify information by the device hosting them. In many cases it is immaterial if the information is found in a given format or another, or where within the global Internet the information is located. As can be seen from the usage statistics of peer-to-peer networks, there are strong incentives for both acquiring data and making it available even in the case where the legal property holders of that information would like to inhibit such practises.

From the networking point of view, this new viewpoint imposes a huge challenge. Once again, the whole architecture needs to be rethought. Once again, the concepts of naming, addressing, and routing will need to be completely re-scrutinised, and most probably some completely new concepts and technologies will emerge, similar to the emergence of routing protocols as a result of the first revolution.

The task is particularly challenging in naming and addressing of content from the point of view of the actual users. While searching is a natural way of discovering information (as shown by Google), every object must have a unique identifier to allow for retrieving the same object later (*e.g.*, a bookmark in a browser). The new architecture must support both access paradigms at the same time transparently to the user.

On the other hand, the problems of privacy, accountability, trust, and reputation still pertain, though in a different form. For example, since the object of interest is now data instead of connections, privacy problems will now condense around data creation and consumption instead of connection tracking and eavesdropping. Similarly, while the problems of node trustworthiness and reputation become less obvious (though no less important), a new set of problems related to data content and its reliability emerge.

The ability to connect to the network from anywhere at any time presents several challenges for managing the content in the network. Information in the network may therefore need to be migrated on the fly, according to where it is currently being accessed. This is in stark contrast to the current client-server-paradigm and calls for new kinds of content distribution mechanisms. In these mechanisms, it seems useful to consider the (possibly mutable) semantic metadata describing a piece of content (*e.g.*, title of a song, presenters, album track number, publisher, genre, copyright information, ...) separately from the (immutable) actual bulk data (*e.g.*, megabytes of MP3 coded

music). For instance, it may be worthwhile to distribute the metadata widely to facilitate search, while storing the bulk data in fewer dedicated nodes.

As the storage capacities of mobile devices increase, the possibility of carrying around most, if not all, of the information relevant to the user at a given time gets closer to a reality. This storage will act as a cache for the permanent information in the network. If the user is allowed to modify such cached information, we need efficient mechanisms for keeping the permanent information in the network up to date and require some conflict resolution algorithms. This consistency problem is not unique to mobile devices, but instead affects the whole network since information is replicated. A very relevant related theme is Delay-Tolerant Networking (DTN), seeking networking solutions even for scenarios where a true end-to-end connection between endpoints might never exist. This is likely have an impact on how the stack and maybe services are typically implemented.

Changing usage patterns

Given that Mark Weiser's original and influential vision of transparent, ubiquitous computing is turning out not to be where technology actually is heading to nor what people want, the complex two-way interactions between users' behaviour and social structures, including co-evolution of institutions and technology, and the emergence of new patterns from the changing individual behaviour, we postulate that it is virtually impossible, at this stage, to anticipate what kind of traffic patterns the evolving usage patterns will produce. Consequently, it appears even more important than before to bridge the gap between applications and networking researchers. For the future Internet, the networking people are looking at radical ideas, such as information networking. At the same time, the applications people are largely assuming that the basic form of networking will continue to exist at a host-to-host basis. This disconnect is likely to lead to largely misleading projections about the evolution at both fronts.

How would users create, search, and share information in the future Internet based on novel concepts such as information networking? How would they maintain their social networks and communicate with others? How would they plan and carry out their daily activities in work and leisure in a world where every act is an information act? How do they manage the inherent complexity of the information, communication, and social spaces exposed through Internet (possibly enhanced by embedded network computing)? Which information, in which form, and at which moment is most useful to users to help them deal with the related tasks? How does it really *feel* like being connected to the new ubiquitous Internet?

Questions such as these belong traditionally to the domain of usability research. Unfortunately, its established methods are not well suited to inform future research and development of novel, ubiquitous Internet technologies, systems, and applications, and are in fact harming the present work by their myopia and narrowness. To make progress, it seems necessary to re-conceptualise ubiquitous interaction and how it should inform the design of related technologies, systems, and applications by studying how people build, use, and share complex artefacts such as information and social networks enabled by future Internet.

A possible theoretical framework for this is suggested by Susan Leigh Star based on the concepts of *infrastructure*, *configurability* and *practices*. She views infrastructures as relational concepts: infrastructures are human-built and provide different things for different people. As every human is a part of the infrastructure of the other people, the technological and social factors are closely interrelated. In particular, technological complexity is related to system configurability, while social complexity relates to the practices enabled by technology, including questions on trust and privacy between people, the role of communication, emergent properties of communication practises in groups and communities of users, and technological opportunities and difficulties in innovating and maintaining such practices. Exploring and understanding the relationship of these two kinds of complexity is a key theoretical goal of future research.

At an even higher abstraction level, the new types of applications and social interactions are likely to impose people and businesses to new types of risks. Hence, in order to facilitate models for risk management and assessment, the insights for new interaction models and resulting traffic patterns need to be enhanced with strategic[2], trust related information and models of how people handle it.

[2] We use the term "strategic information" here in its ethnographic meaning, denoting information that people collect about other people in order to be able to interact socially.

Finally, at a more macro level we anticipate that the changes in this arena will necessarily lead to formation of new business models. The immediate causes are likely to spawn from internalising the reduced transaction cost and the tensions created by the uneven creation of new wealth. In the longer run, it looks also very likely that completely new forms of compensation will emerge, perhaps more suitable for informal or non-institutional social interactions. From an economic point of view this will mean that part of the total economic activity will move from the "dollar markets" to local, social markets. Among other things, this creates new challenges for measuring the total amount of economic activity taking place in the society.

Modelling and data analysis for performance and reliability

IP based networks are becoming a critical resource that public bodies, companies and institutions rely upon to deliver information and to communicate. This requires completely new levels of service guarantees from the network, both in terms of security and reliable performance.

The Internet is subject to many external influences that cannot be controlled, such as fluctuations of the characteristics of network traffic, hardware and software failures, link and power interruptions, malicious attacks, and so on. An additional source of uncertainty in wireless networks is due to node mobility, which leads to unpredictable changes in network topology.

While sound engineering and protocol design are crucial to mitigate these problems, these are not always sufficient to understand large-scale systems. To give an example, the discovery of self-similar structure in network traffic and in the occurrence of link failures had a very significant impact on the practice of capacity planning and provisioning.

Reasoning about large-scale systems that go beyond what can be built and experimented with in the lab force us to resort to models. This is especially true for disruptive technologies that simply do not exist yet. For example, during the early years of the Internet, the networking community relied heavily on models to discuss the pros and cons of traditional reservation-based telecommunication networks versus best-effort IP-style networks.

The programme therefore intends to study novel models of salient features of the future Internet, as detailed below, to complement and inform the architecture and protocol engineering aspects of the project. We foresee the following areas as particularly promising for performance modelling:

1. *Mobility:* Mobility is the chief additional source of uncertainty in wireless networks, and an important parameter for the their performance, because mobility requires network and end-point mechanisms to deal with changes in network topology. A solid understanding of real mobility processes allows us to design better algorithms for routing, resource control, and handover. Therefore, we seek models of mobility in different application scenarios (*e.g.*, vehicular, human in urban environment).

2. *Overlay networks:* Overlay networks provide a means for increasing the scalability of the current Internet, at least in terms of addressing. However, overlay structures on top of the IP network topology create new challenges especially for traffic management. The problems may be in the future exacerbated by the widespread adoption of peer-to-peer based content streaming technologies, such as in the Joost project. This will require new innovative solutions to achieve load balancing in the system. Also, the management of the distribution tree will be an issue.

3. *Wireless access:* Wireless networks are always fundamentally resource limited. In this domain, advanced physical layer techniques such as OFDMA together with adaptive coding and modulation schemes have enabled the systems to achieve high user data rates and a high efficiency. In the near future, physical layer technologies will also use MIMO techniques and, perhaps in the more distant future, Dynamic Spectrum Access, to further increase system efficiency. Above the physical layer the key challenges are in devising radio resource management and traffic management schemes to achieve load balancing and efficient utilisation of the underlying wireless resources.

4. *Traffic:* Network traffic is inherently unpredictable, and a lot of effort has already been invested in developing accurate performance models. However, in the future Internet, traffic characteristics may well be quite different, because of very different usage patterns and applications. Also, as the future Internet may be much more resource-constrained (power, disrupted operation, *etc.*), traffic may have to be modelled at a finer scale, and performance trade-offs to be made more explicit. From the point of view of data applications, traffic needs to be modelled at the so called flow level. Flows correspond roughly to file transfers and users experience the

performance as the total delay of the file transfer. This requires a flow-level modelling approach that captures the impact of the dynamic sharing of the network capacity between the stochastically varying number of flows.

Network socio-economics

Internet is a network of networks. Its main benefit is in social and economic networking, *i.e.*, in reducing the transaction costs when communicating with other people and businesses. From that point of view, a key point of the Future Internet programme is to investigate the networks' ability to deliver useful services to its users. The work shall be tuned towards systems where "the user is happy and satisfied", *i.e.*, where the user considers getting value for her money. Taking a more economic angle, we can define these current and future services as things that provide more utility to the users than is required for their production by their providers, thereby increasing the total amount of wealth in the world.

The long term goal should be a network where "all of the magic" happens automatically, without human action needed. From that point of view, we can say that the vision is to create the networking version of Adam Smith's "invisible hand" that guides the markets towards increasing wealth, combined with an open, democratic, socially fair approach that will force the created wealth to be distributed more evenly than what a raw, pure capitalism would induce.

To give the invisible hand a good grip, Lawrence Lessig has advocated the "code" approach where we try to develop our technical solutions in such a way that the network would be able to enforce the appropriate rules of appropriate behaviour, and when outside recourse is needed, it would be clear beyond (reasonable) doubt who did what. In more technical terms, this seems to coincide with various policy-driven approaches to service and network configuration and deployment, although its scope must be far wider. In some cases, the concept of "soft law" may be applicable to give such "coded" rules jurisdictional power. Ultimately, though, nothing will eliminate the need for some internationally recognised body with some authority to speak on the matter (*e.g.*, WIPO on trademark issues).

Autonomy and resilience

The current way of joining the inter-connection and roaming community requires a lot of manual work. If we assume that network edges and operators show up and disappear frequently, the process and distribution/removal of contact information should be instantaneous. Same applies also for end users that change their subscription owner but retain their identity. Having the configuration information distributed without a centralised management is also highly desirable.

Dealing with these issues requires completely autonomous resiliency and zero configuration, jointly resulting in something that might be called *configuration agility*. Despite its urgent need and apparent efficiency benefit, there has been relatively little real life progress towards genuinely autonomous and resilient networks. This realisation creates two distinct problems:

- Understanding the socio-economic factors that have largely hindered the adoption of even those few pieces of autonomy or resilience enhancing technologies that exist.
- Search for new approaches towards building autonomous and resilient networks, taking into account the aforementioned socio-economic factors.

As a first approximation, we posit that ideas and techniques from complex adaptive systems (CAS), which have been applied successfully to explain certain aspects of biological, social and economical phenomena, can also form a partial basis for building autonomously resilient networks. However, to get there fundamental research is needed to better understand the mechanisms behind the emergent socio-economic features facilitated by the current Internet. Only such understanding will allow one even postulate the potential micro-level formations that are needed to facilitate the emergence of resiliency at the macro-level. In other words, our initial assumption is that, given the complexity constraints, it is impossible to design networks that are genuinely autonomous and resilient at the same time. Therefore, resilience and genuine autonomy shall be seen as emergent properties, growing from the complex structure of underlying interactions, much in analogy to a living cell being an autonomous and resilient system.

Energy consumption

The energy consumed by high-tech industries and institutions represents an attractive and mostly untapped opportunity for energy savings. Characterised by large base-loads operating 24 hours a

day with energy intensities much larger than typical commercial buildings, high-tech buildings include laboratories, cleanrooms, and data centres. These facilities are essential to various industries important to the national economy. As it relates to data centre operations, establishing measurements and metrics for energy efficiency will guide managers on what results to expect from investing in green technology.

Remedies to energy consumption issues can be sought from cross-layer, link-layer, MAC, and network design. In cross-layer design the fundamental questions include to define the set of information that should be exchanged across protocol layers, how that information should be adapted, and how global system constraints and characteristics should be factored into the protocol designs at each layer. However, experience suggests that over-extensive cross-layer communication design, which increases the complexity of the system, is unlikely to succeed. Therefore, we should study optimal protocol solutions not needing extensive cross-layer communication design and patchwork also for energy consumption optimisation.

Traditionally energy consumption research has focused on the base band and radio technologies. In link layer design the goal is to the achieve fundamental capacity limits of the channel; *i.e.*, to overcome channel impairments with relatively little energy. This may require new coding techniques, drawing ideas from block codes, convolutional codes, trellis codes, and turbo codes. Closer to the physical level, MIMO and multiple antenna designs are important. At the MAC layer, the main problems include how to divide the spectrum into different channels and how to assign them to different users to make best use the available time, frequency, and energy. Last but not least, power control is needed to compensate for random channel variations, reduce the transmit power, minimise the probability of link outage, reduce interference to neighbouring nodes, meet hard delay constraints, and prevent buffer overflows. Adaptive resource allocation needs link transmission scheme transmitted power level, symbol transmission rate, constellation size, and coding rate/scheme.

Although these optimisations are important, the ultimate goal is on seamless integration of radio access networks with the services and core Internet network. Hence, also the energy consumption optimisation must cover interaction between all of the system components and whole protocol architecture, including the service and network interaction framework, transport and network protocol solutions, and flow and congestion control mechanisms and algorithms. In addition to radio technology themes outlined above, also the relevant Ethernet and 802.xx developments are highly relevant.

From the networking point of view, problems include neighbour discovery and network connectivity: probing, transmission range, adaptive rate, power, and coding; routing: flooding: robust, less routing overhead, waste bandwidth and battery. In energy-constrained routing issues to consider include delay constrains, battery lifetime, and routing efficiency.

Shifting bottlenecks

> "*There is an old network saying: Bandwidth problems can be cured with money. Latency problems are harder because the speed of light is fixed — you can't bribe God.*" –Anonymous (quoted by D. A. Patterson in CACM, Oct. 2004)

Internet traffic demand has historically roughly doubled every year. While this rate has varied over time, we believe that the trend will continue for the next 10 years. For the Internet core, this means means that the link capacities must grow 1000-fold within the next 10 years. In the optical domain, capacity has grown and can grow at the pace of the traffic demand. However, there are issues related to how to scale the routing capacity in the electronic domain to meet the demand. Today, and in the foreseeable future, the capacity of relaying nodes is largely expressed in terms of packets per second (pps). Consequently, one way of scaling up the nodes would be an architecture that allowed using very large packets. How to split the scaling problem between the optical and electronic domains is also likely to form an issue.

The increasing capacities of core networks allow for more and richer content to be transmitted over larger distances. However, the speed of light puts a minimum bound on network latencies and limits the usability of certain applications in certain network conditions. Data intensive applications are not affected by the "latency bound" and can benefit from added network capacity, but for many, especially interactive applications (games, telephony, videoconferencing, *etc.*), the speed of light will limit their applicability.

Increasing storage capacities will offer many interesting possibilities for content management and caching. This applies to both fixed and mobile devices. Large amounts of ubiquitously available storage make caching and replication attractive, and in many cases necessary, as mechanisms for improving performance. However, wide-scale content replication brings about consistency problems when the content is modifiable. The requirement to verify that the content is up-to-date may become a severe bottleneck for the system. Falling back to a more centralised solution will incur severe performance penalties, so the natural way forward seems to be the development of efficient, distributed consistency management systems. Because different kinds of content have very different consistency requirements, such systems need to be very flexible. As mentioned above, a separation of metadata from bulk content looks like an attractive approach for this, paving the way towards truly networked file systems.

Action plan

The work in the programme will be organised as a (small) number of parallel projects, working towards their partial missions derived from and contributing to the overall mission of the programme. In addition, the programme will involve a number of cross-programme (or cross-ICT SHOK) horizontal actions intended to maintain the focus, transparency, and impact of the work, and facilitate competence building and maintenance. Of course, the decision-making, coordination, and reporting procedures of the programme must match this portfolio and qualitative objectives.

Projects

The bulk of the work of the programme will take place in research projects, running in parallel in a coordinated and transparent fashion. The projects are likely to vary in size and duration. We foresee the need of at least two kinds of projects:

Projects

1. Targeted development
2. Long-germ research

- Targeted development projects addressing current issues with varying time scales. Short-term projects should focus on testing and experimenting, and should include substantial industrial participation. Longer-term projects should focus on new technologies, platforms, and tools, and are likely to include considerable standardisation activities.
- Long-term research projects addressing future challenges. These are likely to require multidisciplinary work including substantial inputs from various non-technical fields from behavioural sciences to social sciences, in addition to fundamental technical work.

A possible third project type might be devoted to innovations and venturing.

In this document we do not endeavour to describe potential project themes in more detail.

Deployment scenarios

The transition to the new Internet will require substantial investments from the service providers, manufacturers, and various users. Obviously, these investments will only be made if they can be expected to bring in a reasonable return. Can we think of circumstances which make this possible? The work on deployment scenarios is intended to provide answers to this question, and more, to create a deeper understanding on the relevant issues and assumptions.[3] This might include studying questions such as the following:

Deployment scenarios

1. New killer application
2. New mobile service platform
3. Two billion mobile Internet users
4. Corporate global intranet
5. Market disturbance
6. New application class

[3] Note that this subsumes work on so-called "business models". A given "business model" is only viable under certain conditions, such as the market position and relative bargaining power of the various stakeholders, the end-users attitudes, the regulative infrastructure, and various network effects. Studying "business models" in disconnection from this background is uninteresting and ultimately futile.

- Which market events or conditions trigger a scenario?
- Who makes the technology being deployed? Who are the key actors in the changing market
- Who is the customer of the technology? What is the added value the customer is willing to pay for?
- Which kinds of secondary products, services, or infrastructures would be needed to create complete services on the basis of the technology? Who will provide them?

Ideally, deployment studies should provide an orthogonal, application-driven approach that complements the bottom-up, technological view of the Future Internet that characterises the bulk of the work. This should provide the programme realistic and challenging use scenarios of the various technologies to be investigated, including experimentation scenarios for field testing in various scales, and expose possible biases and unfounded assumptions on the nature of future applications that, undetected, might distort the programme to developing tomorrow's solutions to yesterday's problems.

At this stage, the following deployment scenarios have been identified:

New killer application: This scenario is triggered by accumulation of enabling technologies (such as multimedia, location, context sensitivity, sensing and acting, and mobile payment) which in combination make a novel mass market application ("killer app") technically possible and highly desirable by the main market. The initial business model is likely to focus on mobile marketing and advertising, but secondary business models are likely to emerge in concert with the success of the application. Secondary services would include massive scale content distribution, adaptation, and personalisation. This scenario is considered possible within the next 2 to 5 years.

New mobile service platform: The triggering event of this scenario is the emergence of a new type of service platform that can provide highly scalable security and privacy to customers as a sellable, value-adding service, and which causes a mass migration of customers. The core end-user applications of the platform would focus on mobile banking and related mobile value value services. Also business-to-business applications are influenced, and drive the development. As consequence, Internet architecture to support context awareness and contextual scalable security is challenged. This too is a medium term scenario (2 to 5 years).

Two billion mobile Internet users: The trigger for this scenario is that the major incumbents of the telecom market create an optimised GSM/EDGE based wireless access for developing markets (China, India, SE Asia, South America, Africa), along with Internet capable inexpensive (<50€) handsets. This creates a service offering which finds the appeal of the next two billion Internet users. The required cost-cutting innovations are based on deploying mass market Internet technologies, such as Routed Carrier Ethernet transport for easily managed data service. This would obviously create immense secondary markets not only for additional networking services, but also end-user and B2B services. The scenario is technically feasible in medium future (2 to 5 years).

We have also identified several more long-term deployment scenarios. They include the following:

Corporate global intranet: A large international corporation switches to the new Internet architecture in its own network, to drive down maintenance and other operational costs, and to have more flexibility.

Market disturbance: A market entrant with a "killer application" forces the incumbents to follow suit and adopt a radically altered service offering. *E.g.*, a mobile P2P social network application winning instantly the hearts (and wallets) of hundreds of millions of users.

New application class: A new application class creates an initially niche market for a new architecture that subsequently is adopted to other uses as well. Possible examples include large-scale services for energy distribution, environment monitoring, intelligent traffic system, well-being applications, and the so-called "Internet of things" enhanced with embedded network computing. Many of these require new business ecologies, often including the public authorities as stakeholders.

Dual use Internet: Rampant terrorism and war forces the national defence systems in developed countries to scrap the present Internet and install a secure, trusted dual-use Internet that can repel terrorist attacks.

Experiments and testbeds

For chance of impact, the long-term, fundamental research in new communication and networking paradigms has to be tested, at least as a proof-of-concept, in realistic environments of sufficient scale, to assess the feasibility of these new concepts, to verify their large-scale (side) effects, and to derive further requirements, orientations and inputs for the long-term research. In some cases, initial deployments can take place as overlay networks over the current Internet, possibly using platforms such as PlanetLab, PanLab, and OneLab. Some themes (*e.g.*, mobile-focused) are more self-contained and could be tried natively. The reason behind choosing an overlay approach is that it allows for the flexibility to try any kinds of ideas with little additional development effort.

Experience nevertheless shows that many issues and side effects are only discovered when systems are deployed in "real-life" situations. For instance, experimentally driven security research includes experimentation with intentionally and unintentionally misbehaving programs and machines in a large, heterogeneous, real-world-like testbed environment, which nevertheless is isolated from the outside world. The experimental facility on Future Internet technologies will also broadly support medium to long term research on networks and services to compare current and future approaches including IP-layer options, IPv4 and IPv6 interoperability, future IP versions (IPv7), various routing protocols (BGP, OSPF), IP multicast and anycast, as well as completely clean-slate approaches (*e.g.*, publish-subscribe).

Large scale experimental facilities of course cannot be developed anew and repeatedly in every single research project. Therefore there is a strong need to ensure that synergies are achieved, through a sustainable and dynamic evolution of the federation of testbeds. Therefore, we propose that the Future Internet programme will include a substantial experimental testbed part, outlined below.

Future Internet Experimental Testbed

Future Internet Experimental Testbed

Finnish universities and many research institutions are connected by existing Funet network administered by CSC. Funet is being expanded to allow offering an even richer set of connectivity services. The following services facilitate conducting large-scale experiments and gaining deployment experience:

1. Direct IP routed connection. If a research unit is behind a restrictive firewall, a separate essentially unfiltered connection can be provided to work around these "techno-administrative" bottlenecks. As a result, experiments with protocols and protocol extensions that would normally be blocked by university IT administration firewalls become possible.
2. Direct L2 connection. For research units wanting to test substantially different IP-versions, custom routers, or other experiments that are not feasible with current IP network, a separate layer two ("Ethernet lambda") connections can be set up to build either
   A. point-to-point connections between two or more universities, or
   B. to set up a Finnish testbed connecting all the sites.
3. Provide experimental infrastructure services. In most cases, the experimental facility participants themselves provide the novel services used in experiments. If in some circumstances it makes more sense to centralise a part of that infrastructure, the some of such services could also be managed by the network operator.

These experimental testbeds can be connected to other European testbeds under consideration in the FIRE consortium or other possibly emerging Finnish testbeds.

The testbeds (except point-to-point connections under scenario 2.a) are connected to Internet and CSC provides basic services (*e.g.*, DNS, NTP, management and operations) in the testbed in a current Funet like manner. In addition, an agreed level of network usage information is provided. This approach allows researchers to focus on doing research and experiments while CSC provides the infrastructure.

A participating site may use the testbed for "production traffic" with agreement of all the parties.

Showcases

Showcases

An essential element of Future Internet research plan are showcases – largish, co-operative longer-term experiments, where new ideas are tested in practice by real users. They differ from technology testbeds, where technologies are tested in laboratory conditions. Showcases will serve the program in four respects:

Showcases will make possible to study integration issues in real life contexts. The Future Internet will consist of a number of layers and services, part of them legacy ones, and although it is possible to demonstrate the functioning of a novel solution in a laboratory experiment, there is a real challenge to integrate all new and old parts into a functioning whole.

Secondly, the final usefulness of novel services can only be tested with real users in a real-life setting, and preferably during a longer testing period. Such experiments can and should be used in further improvement of proposed solutions and innovation based on a feedback from users like in ”living labs” environments. Although it is not necessary that all showcases are related to existing living labs, the possibilities for such co-operation should be studied where possible.

Thirdly, by pooling the experiment resources together showcases will enable much better and complete experimenting. If each partner would do their own experiments, these would be much smaller and each concentrate only to the corresponding area of development.

Finally, showcases will be the most visible part of the project towards different stakeholders and general audience. They will make it possible to advertise the idea of Future Internet in concrete terms, and potentially to draw more interested partners (public bodies, SME companies) in contact with the program.

There should be more than one showcase, but the number should be limited to make them large enough. Their areas will be selected according to several criteria, like the interests of partners and availability of suitable fields and target groups. Showcases will be co-operative efforts between a number of partners, and they will be organised as separate projects with their own project structure and budget. Some of them are likely to cover themes across the whole ICT SHOK programme, or even across several SHOK’s.

International liaisons

Internationalization

To reach its intended impact, the Future Internet programme must create strong links to key research groups and researchers in the industry and academia worldwide and also maintain a close liaison with the key standardisation activities in IETF, W3C, and other bodies. For this, the existing links of the partners should be consolidated and reinforced. A partial list of the key present contacts include the following:

Industrial:

- BT Research: Dirk Trossen, Bob Briscoe
- DoCoMo US Labs: James Kempf (dormant)
- Microsoft Research: Dr. Tuomas Aura
- Boeing Phantom Works: Dr. Thomas Henderson
- Telecom Italia: Mario Morelli

Academia:

- SICS: Bengt Ahlgren
- Aalborg University: Lars Roost, Gustav Haraldsson, Per Toft
- RWTH: Petri Mähönen
- U. of Hamburg-Harburg: Murugarag Shanmugam, Aarthi Nagarajan
- Darmstadt University of Technology: Max Muehlhaeuser, Ralf Steinmetz
- Cambridge: Ross Anderson, Frank Stajano (somewhat dormant)

- INRIA: Claude Castelluccia (somewhat dormant but will get renewed soon)
- University of Parma (UNIPR): Natalya Fedotova
- Berkeley/UCB and ICSI: Scott Shenker, Sally Floyd, Vern Paxson, Ion Stoica, Randy Katz, Anthony Joseph, Hal Varian, AnnaLee Saxenian, et al.
- Stanford/CyLAB: Allessandro Acquisti, Adrian Perrig (contacts would need to be renewed)
- USCD/Systems and Networking: Alex Snoeren, Stefan Savage, Amin Vahdat, et al.
- MIT/CSAIL: Karen Sollins
- Brooklyn Polytechnic: Keith Ross
- Tsinghua University: Prof. Wu Jianping, Prof. Xu Mingwei, Prof. Xu Ke
- Beijing University of Posts and Telecommunications: Prof. Ji Yang
- University of Sao Paulo, Brazil: Prof. Tereza Cristina, M. B. Carvalho

Standardisation:

- IETF HIP Working Group
- IETF HIP Research Group (co-chair: Dr. Andrei Gurtov, HIIT)

Possible means for maintaining international liaison:

- Programme advisory group involving world-class members from industry and academia
- Intimate co-operation with (a few) selected partners, such as UCB/ICSI (through the Finland-ICSI Center for Novel Internet Architectures) or Tsinghua University/CERNET
    - Projects
    - Joint experiments and demonstrations
    - Technology exchange
    - Personnel exchange
    - Joint events
- Annual road show at selected sites rotating in Europe, USA, and China.

Competences

Competence build-up and maintenance

Finally, the Future Internet programme should include activities intended to build and maintain the skills required for its implementation. This would include activities such as

- Research community-building activities (workshops, conferences, seminars, publicity)
- Doctoral programme facilitating especially researchers in industry to focus on research leading to a doctoral degree
- Other educational initiatives (*e.g.*, international M.Sc. programmes).

Obviously, some of these activities overlap the missions of participating universities or other actors such as the Academy of Finland or the Ministry of Education. Therefore, their implementation will require co-operation and co-ordination between the programme and these other actors, possibly also co-funding. Of course, some of the actions might cover the entire ICT SHOK initiative.