

Future Internet

Shpëtim Latifi

South-East European University
Tetovo, Republic of Macedonia

Abstract

Historically, the Internet architecture developed decades ago has been evolving ever since, and as a result, some issues and problems that could have not been foreseen or predicted a few decades ago are now appearing; some principles like end-to-end connectivity have been violated, and in the meantime new issues are emerging - identity and authentication, security, mobility, energy efficiency, etc. While recently there are calls for changes in the current Internet architecture, we stress the hidden risks and additional issues related to these ideas, including but not limited to network complexity, concerns about privacy, lack of guarantees for security, and discouragement for Internet access to ordinary people, and discouragement for developing new applications for the developers. This paper/study instead, maintains that the approach of patching the current architecture - firewalls, proxies, and other middle boxes are helpful and do not increase the complexity, thus do not endanger the evolvability and simplicity of the network. We believe that by better and more global laws the security issues will be minimized - this is what even the most advanced ideas for new architectures call for, since there are no guarantees for absolute security. The integration with cell phone networks and services can solve majority of existing issues with mobility, and energy efficiency. Also models for naming, multicast and anycast are also proposed, and as the need for mobility is growing, together with the need for better identification of data and services instead of machines, these models will emerge slowly. This industry is in the beginning of its era, and there are some great solutions on issues related to mobile support for smartphone devices, and we believe this field is moving in the right direction, thus rejecting the need for major changes in the current architecture.

1. Introduction

The reasons for having the current Internet architecture as such (and the way how it was developed) are very strong. The historical perspective of over four decades of evolution of this architecture is tremendously rich in research and innovations. While most of the proposed technologies over the years have failed and very few have been adopted, one thing is true – that is the Internet architecture is prone to change very slowly. Some of the arguments for this claim will be stressed shortly. While contrasting arguments for a need for quick and comprehensive change in the architecture (i.e. start from scratch, revolution) vs. arguments for slow change (evolution), one needs to stress the basic principles on which the current Internet was developed, and well as stress the major reasons why the current reality requires a different view, and poses the needs for a possibly new architecture.

This paper barely proposes an innovative idea on the future internet architecture, and merely tries to summarize the differences between the requirements for the Internet of 40 years ago and the nowadays Internet; the experience during this time that had led to the need for making a fundamental change in the current architecture, as well as discuss some of the already proposed solutions to some of the issues.

While there are sound reasons for a call about a new architecture that would address current issues with the Internet, two things must be considered in order to be able to say that the current Internet generation should be replaced by a new architecture: 1. how strong are the reasons for changing it?; this view takes into consideration the costs and benefits for replacing a whole architecture with a new one, and as long as patches can be provided for issues that do not concern a vast majority of Internet users anyway, chances are that patches are more feasible, and 2. if the idea for a completely new architecture is embraced, are there guarantees that these issues will be solved, and new issues related to it will not emerge? For example, one could argue that if intelligence is added to the core of the network, generality cannot be preserved, possibly limiting development of new applications, etc.

Historically, the aims for the DARPA Internet Architecture [1] were to have a suite of protocols (known as TCP/IP) for packet switched networks. Also, this includes multiplexing techniques for interconnecting existing networks. Another alternative to packet switching was circuit switching. Many researchers would argue that circuit switching offers more in the sense of network delivery guarantees, delay guarantees, etc., but on the other hand the networks that were to be integrated in the network of networks were packet switching networks. Not only that, but another aim was to have a stateless network, which during failure would be able to reconstitute quickly, without interruption in the communication. Most of the goals set at the time of beginning of the Internet, are still very sound. For instance most of the new proposals for new architectures include the need for communication non-interruption even if some nodes fail, or are attacked. They also request that any new architecture should support different types of services, and be able to integrate variety of networks. But, being it a network developed primarily for military purposes, the importance of different goals was different from the one that is now, and some of them, understandably, were not even foreseen. For example, few people might have guessed that this will be an internetwork that will host billions of nodes worldwide. This bare fact imposes many issues – trustworthiness and security, network space and addressing availability, location, naming, the end-to-end principle, etc. As a consequence, the Internet evolution has addressed some of this issues, some successfully, some of them without any success, and some of them have introduced new issues in the picture (for example the introduction of private IPs has addressed the addressing issue, but has contributed in violation of the end-to-end argument).

The following part of the paper discusses the abovementioned aspects one by one, with an explanation of the issue followed by ideas that have been proposed as solution to those. We will focus on key issues, without discussing details about things they might raise.

2. The principle of evolvable Internet

One key characteristic of the Internet architecture is that it allows the system to evolve. In other words, the architecture is minimalistic and general, broad to support applications that are not developed yet; support technologies that are different from one another in many ways. This specific of the Internet has made it so robust that it has prevailed drastic technology improvements in all the years.

While some proposals argue whether the New Internet should be built from scratch or change the core of existing architecture [2, 3, 4], one thing that remains clear is that whatever it is, the New Internet should maintain the generality in mind [5], which will allow support for future technologies and applications. After all, this is an easy statement, but it is hard to implement, since it includes non-existent factors (factors that belong to the future, which are unknown). For example, how general should be this architecture? In the meantime, what are the compromises that need to be done for this benefit? If this generality involves too many options to extensibility, which most of the times will not be used, or are difficult to implement and include a big overhead, it will clearly be of no benefit.

From past experience and evidence, one could argue that the IP packets are one of the biggest advantages of Internet architecture [6]. Regardless of the technology or the medium, IP packets have driven the innovations of different applications over the years, and thus should not be replaced with a different implementation. IP packets may not be the most appropriate method that optimizes in a best way the use of a specific technology, but here comes in the argument for generality; clearly this is the price for being able to allow different networks to be integrated together. Technologies should (and do) change rather fast to suite the architecture, not vice versa. For instance, the best effort delivery of IP packets does not include any delivery or delay guarantees, but this has improved a lot over the years with different technologies, starting from 10Mbps, 100Mbps, 1Gbps, to 10+Gbps. This is a proof of how technology can change so it will meet the rising needs better, and yet, it multiplexes packets in a general and mutual infrastructure, without having to change the architecture. By this, we want to show that there are things that can be done with the existing architecture that will address some issues; but there are still some major problems that are much more complicated than this. Some of these elements will be discussed in the following sections.

3. Issues in the existing architecture

This section addresses some of the issues of the current Internet architecture and the question raised by the National Science Foundation about the next Internet generation called Global Environment for Network Innovations [13] (GENI). Taken into consideration that the initiative has encouraged the research community to come up with different ideas on how the next Internet generation might look like, it is our understanding that some of the proposed ideas have very strong arguments and theoretical ground for the suggestions that they make, and yet most of them fall short to take into considerations the risks that those ideas could bring (the issues will be treated one by one in the following section). Without getting into much detail, this paper tries to prove wrong some of them as too optimistic and too risky for a serious consideration until they show practical results in the ground.

3.1. Security

Perhaps one of the most significant reasons for the calls for change of the architecture is the security aspect. One could argue that the current architecture does very little when security is concerned, and concerns have been raised regarding security - Denial of Service attacks, as well as Internet fraud and spoofing (and sometimes all at once). It should also be mentioned that not only individual nodes are attacked, but the network itself as a whole can be a target; not only is the security an issue, but the trustworthiness as well. Over a billion users interact so much on the network nowadays, that in most cases they are not aware of the host node they are in contact with. This is a major philosophical change on how the Internet is now perceived. The original Internet design has not considered this matter. In fact, decades ago the assumption would have been that individual end nodes would be willing to communicate. This is clearly not the case today (who is willing to receive spam email?!).

However, this situation is not without a reason. The network does very little about security, because it is very transparent (although network transparency is broken on many occasions, which will be discussed next). Packets leave the network in the same way they enter it – they are not modified (the end-to-end principle which is now violated; this principle will be discussed in the next section). Many would argue that the stateless nature of the network is not sufficient, and instead routers should be able to keep some state, and also distinguish flows in the network. These demands might go against the principle of network simplicity, and clearly go against the end-to-end principle (section 3.2), and the use of datagrams. Basically, if one node is currently attacked on the network, the node is left on its own to deal with the attack, because no intermediate en route hop does anything to prevent this attack. In particular, when discussing Denial of Service attacks, here the attacker wastes victim's recourses (the victim is often a server) to the point that no other (good) user is able to get any service from that server. One model suggests a proactive approach in defense, i.e. by offense [7]. What this model suggests is to isolate the attackers by requiring the good clients to increase their upload bandwidth. This, so called 'speak up' model, does not guarantee any success, at best. It also makes the good clients busier with upload, which they might not desire. And it also adds total overhead to the network.

Traceback solutions are also proposed to solve the problem of Denial of service attacks [14, 15, 16]. This is also very difficult to rely on, because in such a case, the victim has the IP source address of the attacker, which might be spoofed. One of the flavors of such a technique is Probabilistic Packet Marking. The idea behind this is that the router marks the IP address packet in a probabilistic fashion. Since the attackers send a big deal of packets to attack the victim, eventually the victim should be able to reconstruct the graph to the attacker on basis of those marked packets. One weakness of the model is that the marking fields themselves may be spoofed (it should be as easy for the attacker to spoof the marking field, as the source address). Another weakness is that different spoofing methods, like enhanced random spoofing, and topology aware spoofing can be very successful as attack techniques, with almost no chance to be captured from the attacker. The weaknesses of this model are best described in [8].

These problems (which are much more difficult to deal with) have led the research community to propose completely new models, new architectures, which will try address these issues.

One proposal suggests that the layered architecture is not anymore an adequate foundation for networking architectures, due to the fact that layers have already been violated by middle boxes (NATs, firewalls, caches, proxies, etc.), thus it suggests a non-layered role based model architecture - RBA [2].

In this architecture, the communication would be organized in functional units – roles, instead of hierarchically. The proposal admits that role violations are possible, however.

This idea suggests a completely new architecture, which is very difficult to run for, especially because there is an IP layered architecture in place that has served well for many decades; there should be a good reason to change it now – the fact the role violations are possible here, it is very difficult to prove that this architecture is a better one. But, if we disregard that fact for a moment, perhaps the biggest drawback is that in a non-layered architecture there is basically a huge number of ‘layers’. This type of architecture would run different protocols for different purposes (connection, routing, control, security, management, etc.), which will make the network very complex, that even if it still remains general, it will take too much time and effort to set up connections, deploy new applications and services, most of them users will not be able to adopt easily.

One good reason for the huge increase of Internet usage over the years is the simple and general architecture of the Internet. The hourglass model of TCP/IP is the simplest way of network deployment – all that is needed to use the network is the IP address. Everything else is up to the developer and the user. Just as a comparison, the telephone network, although much older than the Internet, has not been able to evolve into nothing else but a telephone network, exactly due to the complexity that exists in the network itself [9].

So, when discussing the need to change in order to meet new demands from the new era, one should inevitably admit the advantages of the current architecture like the IP model, and engage to keep them as such. Having said this, this paper still acknowledges the major issues of the current architecture, security being one of them, but instead, the solutions should be found within the existing architecture (be it existing or new methods). Two things that will contribute to the security are trustworthiness and stricter laws and law enforcement (section 3.7).

The trustworthiness is not a new concept, but has never been treated thoroughly and with participation from all sides concerned. Recently Microsoft has initiated a larger debate on the principle of end-to-end trust [17]. The strategy gives three directions in which a higher trustworthiness (as a broader concept than security) can be achieved. First is creation of a trusted stack where security is rooted in hardware and where each element in the stack (hardware, software, data and people) can be authenticated in appropriate circumstances. The second one involves managing claims relating to identity attributes, which means that identity claims need to be passed (sometimes names, sometimes an attribute like proof of age or citizenship). Some important technologies, such as public key infrastructure (PKI) and smart cards are now mature enough for broad deployment, and this is meant to ensure higher security in the Internet.

The Internet, in order to be safe, needs to support the option of identities based on in-person proofing, enabling the issuance of credentials that do not depend upon the possession of a shared secret by the person whose identity is being verified. To some extent, government activities and markets themselves are driving ‘in person proofing’ regimes. For example, governments are issuing (or considering issuing) e-ID cards for government functions. This idea does not require any architectural change in the Internet.

Finally a good alignment of technological, social, political and economic forces so that we make real progress is needed. The nontechnical aspect (laws and regulations) is discussed in section 3.7, and is unavoidable aspect of security, even in some future architecture. Besides, ‘intelligent’ future Internet with higher security premises cannot guarantee non-existence of violations in authentication, eavesdropping from authorities (or higher managers in a realm), and preserve the principle of network generality. These issues are discussed next.

3.1 Network transparency

The network transparency has been widely discussed in recent years, and the fact that the network provides no security (hence is transparent) has led to the suggestion of controlled transparency [6]. The basic idea behind this proposal is that due to the fact that network users not always know their ‘interlocutor’ on the other end, sometimes they need to protect themselves better than other times. In such cases, the network transparency would decrease, and (hopefully) the malicious user will not be able to utilize the fully transparent network to attack the good user, since the network will block the bad user. In cases where users are authenticated and there is trust among them, the network will continue to be fully transparent. This idea sounds fine, but there are also questions that are associated with this idea, which make it more complex; we can only guess some of the questions that might be raised with respect to this idea.

First, will this violate the end-to-end argument? (The end-to-end argument is discussed in the next section and the explanation to why the answer to this question is yes); second, are we willing to move from a 'de facto' to a 'de jure' violation of the network transparency? In other words, are we going to institutionalize the violation of the end-to-end argument by deploying the controlled transparency? The final question is how and who is going to perform the controlled transparency? Obviously, behind each network element there is someone (human being, company, government, etc.).

To better address these questions, let us stress the end-to-end principle that has been a very important characteristic of the Internet for many years first.

3.1.1. The 'End-to-end' principle

This principle clearly states that devices communicate directly among each other, and the network provides the means for that communication. The network itself is not and should not intervene in a whatsoever way in the content of the packets sent over it – it remains dumb. All application specific functions should be done on the end nodes, even if it is possible to implement such function in the network, it is not recommended to do so [8]. There are many reasons for this. This way the network remains simple and easy to manage. It is also easier to upgrade the network and different technologies on that network if it is kept simple. It also enables to deploy more new application on the network much easily without having to change the core of the network. This is the principle of generality, which is the main reason for the development of all applications so far over the Internet.

The Internet is not fully transparent today. The use of firewalls, NATs, caches, etc. has been violating this principle for some time now. Firewalls are used to protect one end system or a part of the network (possibly local network) in an on/off mode. All communications that seem insecure are rejected from and to the outside Internet. This is a primitive way of protecting a part of the network, and it still does not protect from internal attacks. Firewalls are clearly implemented in the network, between two end nodes that might want to communicate. However, the principle of evolvable Internet has not given a better answer or a solution to this issue so far in an effective manner. Firewalls are used largely nowadays, and therefore we maintain that the end-to-end principle can be violated in such cases, but this remains an exception, especially since firewalls do not add any complexity or intelligence to the network besides breaking the end-to-end communication.

NAT (Network Address Translator) boxes are another type of element that violates even more the end-to end principle. Originally, the goals for their implementation were to solve the problem of running out of public IP addresses. NATs allow end systems to be assigned private IP addresses, which are recognized and used only within a part of the network, but not on the outer side of the NAT. This way, packets from different end nodes from the local network that need to leave this network, are assigned a new IP address from the NAT that is public and globally unique. Thus, NATs not only interrupt the clean end-to-end communication between two hosts, but they also change the content of the packets. Along with the IP address, the port numbers are also changed from the NAT. Therefore, hosts communication between them, actually communicate with the NAT, and the NAT is the 'entrepreneur' that brings together the two ends nodes, without them being aware of this fact.

The IPv6 and its slow deployment is expected to solve the problem with the IP addresses, and its full implementation will not take much longer [11] (the IPv4 addresses are really running out now), and perhaps this will slow down the further implementation of NATs. As many developers and content providers cannot use NATs due to the fact that their servers cannot be seen from behind the box, it remains to be seen that these boxes will eventually be used less and less on the Internet.

Caches are also elements that violate the end-to-end principle in a more subtle way. One simple way is when one visits one website, some (or all) of the contents might be placed on a different region closer to the user, so future accesses are quicker at users' convenience. One example of this type of implementation is Akamai [3]. The simplified explanation of Akamai is that content is distributed over to different servers in different locations, making access to content quicker and more reliable. Another example is search engines – for a same keyword search, a user might get a different answer from one city (or country) than a user from a different one. This is done based on IP location, and is specific to different factors, like market, people's interest, etc. In both cases just mentioned above the user thinks that they are communicating with a single end system, but the reality is quite different. The type of cache orientated design is a change of the original pattern where exactly two end systems communicate with each other.

Current trends show that we cannot rely on the pure end-to-end principle, nor it seems to be the goal. While the deployment of IPv6 will address some issues, like the local IP addresses introduced with NATs (which is very important), one cannot hope for a pure end-to-end and fully transparent network.

This fact has led to some more drastic proposals for change in the core of the network. One idea proposes a new architecture, where this network transparency is as low as possible at a level, where users need to authenticate first before being able to access the network [2]. In other words, without explicit authentication, one cannot access the network. From what was stated earlier, it is clear that some type of in-person authentication can be made even in the current Internet, without the need to modify the network core. While it is the willingness of all parties concerned that will decide if this trend will take off (ISPs, governments, businesses, users, etc.), it should remain clear that the authentication (which is not binary, but differs upon occasion) should only take place when necessary (like money transactions, personal data, etc). This brings up the other element that authentication brings - privacy. By privacy, we mean that while users should be responsible and accountable for their interactions on the network, there should still be a more subtle way of identity check (if necessary), instead of explicit disclosure. The privacy in general (or even anonymity), cannot be compromised a priori and in its social context is a highly appreciated value in many societies, and should be revealed only when it is necessary, but not always and for no reason.

Two other elements should be taken into consideration when we talk about changes in the network 1. The social factor and 2. Necessity for simple network.

3.3 The social factor

By social factor we mean that there should be persistent effort to bring the Internet closer to the public. The expectations of big increase in access to the Internet are still not being met, and different parts of the world have very little or no access to the Internet – the tool that is meant to bring the world into every home. If, for some reason, these limitations come into picture, they can only discourage the usual users to utilize less the network, which might have substantial consequences.

3.4 Simplicity of the network

The necessary simplicity of the network assumes the simple IP model that encourages and enables a wide range of application to be developed on the network. One undisputable fact is that whatever change that might be made to the current architecture core with respect to increase of its complexity, it make it less flexible in terms of ‘runs over everything’, regarding applications and technologies. This feature has been a driving machine of the Internet for many years, and has led to an economic and social revolution, and it should be given credit for that.

As far as the questions raised in this and in the previous sections are concerned, this paper maintains that the network should continue to be as transparent as possible. By this, we mean in cases where for practical reasons we might need some higher level violation of the end-to-end principle, as in the case of caches, a compromise can be made. This does not affect the privacy, and yet does not add state and much intelligence on the network – the routers will still do their basic job, which is routing, and the servers will not maintain any state. But, even in cases where this principle has to be violated, it should be done on a higher level, not in the network itself. One cannot guarantee that if the network core is changed (by adding state in the routers, leaving the best effort stateless approach, etc.), the generality will be preserved, and the network will remain as technology and protocol friendly as it is now.

This paper, as explained above, is more conservative in its approach, and does not back the ideas for a revolutionary Internet, or architectures that are new and all different. Having stated this, some change is desirable, since the network should provide ground for new and future applications and services; some of them are not completely supported now (like Mobile IP), but there are advanced integrated solutions offered by software and telephone companies that address these problems fairly well (section 3.6); and some are still yet to come, thus the approach used so far – offering patches and repairing problems en route has been successful to a great extent, at least as far as current applications are concerned.

3.5 Quality of service

One issue that the current Internet architecture faces is lack of Quality of Service; that is the network does not make any guarantee about the delay and throughput rate of flows. Some would argue that due to its stateless environment, the network does not even recognize flows.

One proposal suggests is the future Internet should be able to provide QoS guarantees, and sometimes even total isolation [11]. Thus this idea suggests that both datagrams and circuits should be provided in Internet 3.0. Hence, in a shared wire circuits will be offered to those who are interested in delay and bandwidth guarantees, whereas datagram routing will be left for those interested in that.

While this approach is feasible from the technical aspect, some hidden risks are associated with it. First, provided that circuits offer a better service (some guarantees), this must include the higher cost for the costumers interested in it (this is not bad at all – just as a comparison, there are people that fly business class at a higher rate, and people in economy class at a lower rate). Since circuits do not make best use of the resources that are offered, and instead sometimes they are even wasted (and cannot be allocated to someone else as long as the circuit is not disconnected), there is a risk that this might influence the datagram routing seriously. As a result, users that do not have quality of service support, might start to get way worse service that they do now. Practically, it does not cost much big businesses and corporations to buy out high rates of bandwidth and data rates over circuits, and individual costumers could be damaged by it. Thus the combination of circuits and datagrams might be a solution to the existing problem for quality of service, but only if there is a balance such that datagrams do not suffer. Since these two approaches do not use the recourses in a ‘fair’ way (that is use the resources reasonably and responsibly), and weakest of two (which is datagrams) will be left on the mercy of the other one. Besides, having only datagrams (as it is now in the current Internet) sometimes requires higher bandwidth and better lines (even though datagrams utilize the network in a best way), let alone the fact that now datagrams have to share ‘fate’ (medium) with another service, such as circuits. One can always argue that more bandwidth from the ISP can solve this problem, but it is not the ISPs that pay for that, it is the end users.

Thus, without serious analysis and numbers that might prove the opposite, we believe that the approach of shared medium for datagrams and circuits can cause more harm to the net, by satisfying only some users’ needs. This paper maintains that only datagrams is still the best approach, and it should not be changed in the future Internet. Perhaps with some more investment on the Internet infrastructure, and more optical links deployment, the data rates will increase, at most users’ convenience; for those users with higher needs for guarantees, the private line VPNs are also available; A network architecture should not be changed due to the need to replace private VPNs with core circuit switching on the Internet – this only shows that this is not a major problem, and even now there is a solution to that.

3.6 Mobility and naming

It is clear that IP addresses should serve a purpose, and it is addressing. Today, IP addresses express both network location and node identity [25]. IP addresses should represent location, and as one node moves from one place to another, IP addresses are assigned dynamically. This approach is acceptable and should therefore not be changed. The original Internet architecture was designed to provide unicast point-to-point communication between fixed locations. In this basic service, the sending host knows the IP address of the receiver and the job of IP routing and forwarding is simply to deliver packets to the (fixed) location of the desired IP address. The simplicity of this point-to-point communication abstraction contributed greatly to the scalability and efficiency of the Internet. However, many applications would benefit from more general communication abstractions, such as multicast, anycast, and host mobility. These abstractions have proven difficult to implement scalably at the IP layer [18,19,20].

As the number of mobile devices is growing, the need to identify these kinds of devices in a unique way while in move is growing. While Mobile IP has attempted to solve the issue of mobility, it has been difficult to deploy it due to issues including scalability.

More specifically, the need for naming is much bigger with servers than it is with clients, since the demand and server access is much bigger. Basically, this is an issue that is already emerging – naming data and services, instead of nodes.

The current model is a legacy of the first Internet, where Internet applications of that time, like file transfer and remote login were part of the host-to-host communication design [3]. Needs are changing lately, and instead, today most Internet applications include data and service access, and a service (like CNN, MSN, etc.) might include multiple hosts and multiple locations. Certainly, the existing architecture supports this kind of access, although issues like availability might arise. Akamai is an example of a successfully developed idea of mirroring content (usually media content) stored on costumer servers. Although the domain name in this case remains the same, the IP address directs to an Akamai rather than the costumer server. This Akamai server is then picked based on the content and the actual location of the user's network. This example shows that content delivery on the current architecture is not as problematic as once thought, and with additional ideas put in practice, data and service availability can be also reached.

A solution for naming replicas (mobile data and services) has been proposed in [21]. This idea introduces the idea of Human-Friendly-Names (HFN); a scalable HFN-to-URL resolution mechanism that makes use of the Domain Name System (DNS) and the Globe location service to name and locate resources. Using URNs (Uniform resource names) to identify resources and URLs (Uniform resource locator) to access them lets end users use one URN to refer (indirectly) to copies at multiple locations. To access the resource identified by a URN, a way to resolve that URN into access information, such as a URL is needed. Because a URN refers to a resource rather than its location, users can move the resource around without changing its URN. A URN can thus support mobile resources by referring indirectly to a set of URLs that changes over time. Because URNs identify resources to machines, they need not be human-friendly [22]. Unlike URNs, HFNs explicitly allow the use of descriptive, highly usable names. Replicating or moving a resource will not affect its name, for example, and a user can freely change the HFN without affecting replica placement.

A new idea of a network which proposes a completely new, called data oriented network can be found in [3]. As the demand for supporting mobile devices is rising, new ideas which not necessarily require a change in the network architecture are emerging. One example is the integrated Wi-Fi support in smartPhones. Arguably, Apple is in lead with this technology, offering the possibility to connect iPhone devices on available wireless networks, thus making possible direction communication with not only simple email servers, but Exchange servers as well using the push protocol for a direct connection with the server. These devices offer most of the necessary things for IP mobility nowadays: they are ready to support asymmetric protocols, they make use of proxies in case they are turned off or are in sleeping mode, thus enabling high energy efficiency at the same time, support for streaming data (voice and video) [12].

Currently, mobility is still best supported in the application level, although it is not yet clear if the trends will continue in this direction. [23] Thus, as one might argue that the network architecture should change in order to be able to provide for better mobility management, the other side of the coin is that the cell phone industry and the computer industry might integrate both networks into small computers and cell devices, and make it much easier to deploy a whole range of services for these mobile devices. The applications for these devices are in an early stage, and in the following period we could witness a bigger exploitation of the possibilities in this field. It is clear that the emergence of P2P overlay networks has not been exploited enough, and much of the answers arguably may lie in the overlays. This is turn favors the application level mobility and multicast, which might take off in a near future. As demand is growing, especially for mobility, new ideas that can be built on top of current IP (as overlays) are coming in the picture, thus eliminating the need for a new architecture. One such proposal is 'the Overlay-based Internet Indirection Infrastructure (i3) [24]. Authors propose a single new overlay network that serves as a general-purpose Internet Indirection Infrastructure (i3), which offers a powerful and flexible rendezvous-based communication abstraction; here applications can easily implement a variety of communication services, such as multicast, anycast, and mobility, on top of this communication abstraction. This approach provides a general overlay service that avoids both the technical and deployment challenges inherent in IP-layer solutions and the redundancy and lack of synergy in more traditional application-layer approaches.

3.7 Laws and regulations

Perhaps the experience in networking in the last decades, and certainly the experience from other fields of social life prove that one cannot claim absolute security and certainty in anything. Just like there is no guarantee that there won't be crime and robberies in the streets and banks, one cannot hope for a better situation in the networking world.

Having acknowledged this, the networking community and the general public requires at least the enforcement of the principle of accountability for those who decide to violate the law. In other words, there is a need for identification on the network, such that authorities are able to identify possible illegal activities on the network, and take measures to sanction it. This discussion leads to the eternal debate of privacy vs. responsibility. In many countries, certainly including the United States, the principle of privacy is a strong value and has a high position in the value system [10]. In other words, entities in the network should be able to interact freely and without constraints, as long as they do not violate the law. This freedom includes the element of privacy (which not always means anonymity). And privacy is determined based on the type of interaction that is performed on the net: individual subjects, people, businesses, etc. normally have less privacy as they move along the axis individual customer – content provider (with the latter one being less private). This is again, in accordance to social life – ordinary people enjoy more privacy than celebrities. Therefore, wiretapping and other sort of surveillance should be made possible on legal terms (technically it is possible), but it should be made known to the public who can do it, and under what circumstances (most people would not have against wiretapping of possible terrorists). It is clear that laws can be misused for this purpose, but the bottom line is that any kind of violation of this type should be dealt with outside the network.

The opposite of this would be having people to authenticate every time they use the Internet, which may lead to privacy violations. Besides, this type of network would not encourage the growth of number of people with access to the network, but it will rather discourage it. The laws for cyber criminal do not have a long history, and they are too undergoing a process of evolution, with more precise definitions and sanctions for abusive behavior on the net, for Internet infringements, scams, etc. The level of legal progress is different in different countries, some being in the beginning phase of the process. It is governments' responsibility to act as closely as possible to harmonize the laws of this area, and certainly, for countries where it is not possible to treat cyber crime as illegal, other measures are still possible: credit card payments (say VISA) are not possible from certain type of countries that do not meet the standards and regulations for safe e-commerce. This sort of block is lifted once countries meet these standards. Clearly today's Internet needs to address issues like security in future, and as explained so far, this paper maintains that the best way to do this is by changes to the existing architecture (until accountability is ensured) and changes in the legal terms and better law enforcement. Technical changes only will never solve major problems, without causing other related harm.

Conclusion

The original Internet architecture, developed since over 40 years ago has been evolving ever since, and as a result, some issues and problems that could have not been foreseen or predicted a few decades ago are now appearing; some principles like end-to-end connectivity have been violated, and in the meantime new issues are emerging - identity and authentication, security, mobility, energy efficiency, etc. While recently there are calls for changes in the current Internet architecture, we stress the hidden risks and additional issues related to these ideas, including but not limited to network complexity, concerns about privacy, lack of guarantees for security, and discouragement for Internet access to ordinary people, and discouragement for developing new applications for the developers. This paper/study instead, maintains that the approach of patching the current architecture - firewalls, proxies, and other middle boxes are helpful and do not increase the complexity, thus do not endanger the evolvability and simplicity of the network. We believe that by better and more global laws the security issues will be minimized - this is what even the most advanced ideas for new architectures call for, since there are no guarantees for absolute security. The integration with cell phone networks and services can solve majority of existing issues with mobility, and energy efficiency. Also models for naming, multicast and anycast are also proposed, and as the need for mobility is growing, together with the need for better identification of data and services instead of machines, these models will emerge slowly. This industry is in the beginning of its era, and there are some great solutions on issues related to mobile support to smartphone devices, and we believe this field is moving in the right direction, thus rejecting the need for major changes in the current architecture. This industry is in the beginning steps, and there are some great solutions on issues related to mobile support to smartphone devices, and we believe this field is moving in the right direction, thus rejecting the need for major changes in the current architecture.

References

- D. D. Clark, "The Design Philosophy of the DARPA Internet Protocols", SIGCOMM '88, Computer Communication Review Vol. 18, No. 4, August 1988, pp. 106–114
- R. Braden, Ted Faber, Mark Handley, "From Protocol Stack to Protocol Heap – RoleBased Architecture", ACM SIGCOMM Computer Communications Review, HotNets I, Princeton, NJ, USA, October 2002
- T. Koponen, M. Chawla, B. G. Chun, A. Ermolinskiy, K. H. Kim, S. Shenker and I. Stoica, "A Data-Oriented (and Beyond) Network Architecture, SIGCOMM'07, August 27–31, 2007, Kyoto, Japan
- R. Jain, "Internet 3.0: Ten Problems with Current Internet Architecture and Solutions for the Next Generation", Proceedings IEEE Military Communications Conference (Milcom 2006), Washington DC, October 23–25, 2006
- S. Ratnasamy, S. Schenker and S. McCanne, "Towards an Evolvable Internet Architecture", 2005, SIGCOMM'05 August 21–26, 2005, Philadelphia, Pennsylvania, USA, p.2
- D. D. Clark, K. Sollins, J. Wroclawski and T. Faber, "Addressing Reality: An Architectural Response to Real-World Demands on the Evolving Internet", ACM SIGCOMM 2003 Workshops, August 25&27, 2003, Karlsruhe, Germany, p.3
- M. Walfish, M. Vutukuru, H. Balakrishnan, D. Karger and S. Shenker, "DDoS Defense by Offense", SIGCOMM '06, September 11.15, 2006, Pisa, Italy
- V. Paruchuri, A. Duresi and R. Jain, "On the (in)effectiveness of Probabilistic Marking for IP Traceback under DDoS Attacks"
- D. R. Kuhn, "Sources of Failure in the Public Switched Telephone Network", IEEE Computer, Vol. 30, No. 4 (April, 1997).
- M. S. Blumenthal and D. D. Clark, "Rethinking the Design of the Internet: The End-to-End Arguments vs. the Brave New World", ACM Transactions on Internet Technology, Vol. 1, No. 1, August 2001, Pages 70–109. <http://www.potaroo.net/tools/ipv4/index.html> Apple Inc.", <http://www.apple.com/pr/library/2008/03/06iphone.html> "GENI", <http://geni.net>
- D. X. Song and A. Perrig, "Advanced and Authenticated Marking Schemes for IP Traceback," in Proceedings IEEE INFOCOM, 2001.
- D. Dean, M. Franklin, and A. Stubblefield, "An Algebraic Approach to IP Traceback," in Proc. 8th Network and Distributed System Security Symposium, 2001.
- A. Yaar, A. Perrig, and D. Song, "FIT: Fast Internet Traceback", in Proceedings IEEE INFOCOM, 2005
- Establishing End to End Trust, Scott Charney, Microsoft Corp.2008.
- Y. Chu, S. G. Rao, and H. A Zhang, "Case for end system multicast. In *Proc. of ACM SIGMETRICS'00* (Santa Clara, CA, June 2000), pp. 1–12.
- H. Holbrook, and D. Cheriton, "IP multicast channels: Express support for large-scale single-source applications", in *Proc. of ACM SIGCOMM'99* (Cambridge, Massachusetts, Aug. 1999), pp. 65–78.
- I. Stoica, T. Ng, and H. Zhang, "A recursive unicast approach to multicast", in *Proc. of INFOCOM'00* (Tel-Aviv, Israel, Mar. 2000), pp. 1644–1653
- G. Ballintijn, M. v. Steen, and A. S.Tanenbaum, "Scalable Human-Friendly Resource Names" Oct. 2001 "RFC 1737", <http://www.apps.ietf.org/rfc/rfc1737.html>
- A. C. Snoeren, and H. Balakrishnan, "An end-to-end approach to host mobility", in *proc. of ACM/IEEE MOBICOM'99* (Cambridge, MA, Aug. 1999).
- I. Stoica, D. Adkins, S. Zhuang, S. Shenker and S. Surana "Internet Indirection Infrastructure", SIGCOMM '02 Pittsburgh, Pennsylvania USA
- J. Saltzer, "On the Naming and Binding of Network Destinations In Local Computer Networks", North-Holland Publishing Company, Amsterdam, 1982, pp. 311-317. Reprinted as RFC 1398, August 1993.