Design Principles for a Service-Aware Future Internet

Position paper by SOFI project

Contributors:

- V. Morreale, ENGINEERING, Italy
- M. Melideo, ENGINEERING, Italy
- R. Krummenacher, STI - University of Innsbruck, Austria
- L. Baresi, Politecnico di Milano, Italy
- M. Pistore, SAYSERVICE, Italy

1 INTRODUCTION

The expansion of the Internet, as worldwide network of interconnected computer networks based on the TCP/IP standard communication protocol, was driven over the last 30 years by the exchange of data between hosts such as personal computers (PCs). Today, the Internet has become an essential enabler for data and contents, flows and information and knowledge exchanges all over the world, enabling in turn a wide range of applications and services.

The Internet of the future is expected to evolve towards a holistic communication, information and knowledge exchange ecosystem. It is supposed to provide efficiently, transparently, interoperably, flexibly, timely and securely provisioned services (including essential and critical services) to humans and systems, allowing for both collaboration and competition among the various stakeholders without restricting considerably their choices.

Indeed, although the original design principles of the Internet [1] are still valuable, there is growing evidence that its architecture and components specified therein face certain objective limits in view of today's understanding and use of the Internet [2]. Certain objectives of the Internet are not sufficiently addressed to cope with the users' new expectations and behaviours (in particular, in terms of flexibility and adaptation) as well as new technological challenges (particularly in terms of mobility and manageability) and socio-economical challenges. The "end-to-end arguments" [3] as one of few general architectural principles of the Internet, though still valid, require some deep analysis with respect to their application in the Future Internet (FI), such as in [4].

In addition to the underlying architectural limitations of the infrastructure, the way services are offered today to human end-users suffers from further obvious limitations in regards to their overall awareness and dynamic adaptability to evolving contexts, whose elements include people, devices, situations, cultures and time.

The evolution of the current Internet towards the FI should be considered from various inter-related perspectives, such as the networks and infrastructure perspective, the services perspective and the media and information perspective. This document is intended to discuss some design principles of the FI with a focus on the service perspective. Thus, the next section briefly describes the vision of the Internet of Services (IoS).

2 Internet of Services (IoS)

Service-oriented computing has gained increasing attention in recent years as the evolution of component-based software. In the Internet context, services have been radically changing the way Internet applications are engineered, executed and maintained (i.e. Internet-scale service oriented computing).

The term Internet of Services (IoS) is an umbrella term to describe several interacting phenomena that will shape the future of how services are provided and operated on the Internet [5].

Throughout this document the term "***service***" refers to any action or set of actions performed by a provider in fulfilment of a request, which occurs through the Internet (i.e. by exploiting data communication) with the ultimate aim of creating and/or providing *added value* or benefits to the requester(s).

In the IoS the access to complex physical computational resources, data, knowledge or software functionality should be in the form of "lose coupled" services, regardless where the software components providing the services are (i.e. "somewhere in the cloud"): personalized services can be flexibly detected and invoked based on the properties describing the context. The IoS will also allow proactive services and not only reactive services as currently enabled on today's Internet [5].

In the IoS interaction will occur through a much broader and interoperating variety of user interface types, which span diverse/heterogeneous networks, organizations and computing platforms/devices, allowing users to dynamically select the most appropriate mode of interaction according to their specific needs, providing input via speech, handwriting, and keystrokes, and getting output via displays, pre-recorded and synthetic speech, audio, and so on. Moreover in the IoS it must be easier for users to design, configure, and mush-up their own services (i.e self-servicing).

The above-mentioned requirements represents only some of the desired properties of the envisaged IoS, which is then supposed to emerge in several layers, from fundamental infrastructures to applications. Thus, the rest of this paper reports and discusses some design principles that would enable the IoS.

3 Additional design principles of service-aware Future Internet

The FI will not only allow access to services based on technical characteristics such as IP-location or web service identifiers, but also based on contextual information (e.g. using geographical context or business context).

The FI architecture should be service-aware and include services as first order abstractions. Indeed services are the obvious means for users (organizations, public bodies, companies and people) to get *controlled* access to the available information and functionalities provided through the Internet. A service-aware Future Internet should enable and support a permanent, transparent, seamless, context-aware, empowering, and trustworthy interactivity, i.e. the "Perfect interactivity" [5].

The design principles presented and discussed in the rest of this section move around the concept of “service awareness” of the Future Internet (see section 3.1) and cope with the topics of quality (see section 3.2), naming (see section 3.3), addressing (see section 3.4), delivery (see section 3.5) of services as well as shared access to them (see section 3.6).

3.1 Service-awareness

In the current Internet, services are implemented through data exchanges based on suitable protocols. Almost all the management of services is performed at the application level. This approach, however, strongly limits the capability to achieve the “perfect interactivity” requirements mentioned above, since their achievements critically depend on the possibility to control and manage functionalities that are implemented at levels below applications. Achieving these requirements of the application level would require replicating at this level information belonging to lower levels, and forcing the application level to control the operation of these lower levels. This is clearly an impractical approach.

In our opinion, the right approach is that the Future Internet should become a **“service-aware” infrastructure** and include services as *first order abstractions*: core and enabling functionalities for service request and service delivery should be provided by the infrastructure on top of which applications run. The Network of the Future should replace (or complement) the current data-oriented management with solutions more oriented to services and things, **even at the layers below application**. In other words, services and things should be first order concepts modelled and managed at the layers between IP and application (included).

In order to achieve this, it is necessary to provide the application level with more complex or meaningful abstractions than the ones currently adopted for services and things, which are currently seen just as URIs and/or APIs to be invoked over the network. These abstractions should cover, for instance, aspects discussed in detail in the next sessions, such as quality of services (see Section 3.2) or contextualization (see Section 3.3).

It is however also necessary to guarantee that **the network as a whole is able to manage these abstractions**. This means that these abstractions should influence the behaviour of network operations based on those lower levels. This service awareness of the network as a whole will benefit service delivery, and in particular the achievement of perfect interactivity. Moreover, network operations based on the lower levels (e.g. routing) will benefit from being able to understand services as first order abstractions, and to optimize their behaviour according to them.

3.2 Enabling multiple Qualities of Service (QoS)

If we think of services as first-class entities of the Future Internet, we must also think of how consumers (i.e. people or other systems and services) can exploit them. Besides the “usual” functional requirements, non-functional ones also play a fundamental role. Oftentimes known as Qualities of Service (QoS), these agreed commitments can be mandatory for the actual and effective exploitation of provided services. The single best-effort approach offered by the current Internet to transport packets is not enough anymore since clients want to exploit services with the guaranteed qualities they want and not just the best way the network allows for in a certain moment.

This means the network must be equipped at all layers to provide consumers with the services they want and with the qualities of service they need. Indeed guarantees at application level do not come for free, but they are the “results” of what provided by the entire stack. This is why holistic approaches that address the problem at all layers, from the infrastructure up to the application level, are deserving more and more attention.

The provision of "quality" service is also more and more based on Service Level Agreements (SLA) established between the parties. These contracts are usually established at application level, but we need means to transform them automatically into the guarantees that must be satisfied/offered at each level. Each layer of the stack should be a self-aware and self-managed element that works in isolation to provide its "services" in isolation along with promised guarantees, but at the same time it must cooperate with the others for the holistic provision.

3.3 Naming of services

The current Internet has only one "architected" name space, i.e. the set of domain names [6], which were initially devised to basically name entities such as physical machines. Though the domain name system has proved robust and scalable, domain names are not optimal for the naming of either services or information objects.

Services (and service definitions) need to refer different entities (e.g. software components, including other services, and other ICT infrastructures, devices, sensors, and entities in the real world and IoT, users, people and communities, organizational structures and roles, etc.).

The infrastructure offered by domain names is not sufficient to solve the naming problem is such a complex context. Just to make one example, services can be implemented on more than one physical machine[1]. In some cases, the party invoking the service does not care which instance of the service is selected or in which machine (and specific address) it is. In other cases, only one machine is suitable for a given party, e.g., due to the context of this party, to access restrictions, or to quality levels to be reached; the identification of this machine should however be transparent to the user. Finally, in some cases the invoking party wants to exploit a service on a specific machine.

Accessing to services requires infrastructural, global solutions to manage identities and names for all these entities at a global level. These solutions should support for instance: assignment of unique identities for these entities at the different levels of abstraction (e.g., service, service machine, service instance); identification of different names corresponding to the same entity; possibility to define virtual names and hide the real identity; manage changes in names and identities.

3.4 Addressing services

Since users are more and more nomadic and services are more and more provided through a cloud, services must become addressable in a unique way. Future Internet needs service addressing mechanisms that are independent of the physical location (and if possible, technology) of the services and things: current service-to-service addressing is strongly based on physical location (is part of the URI!) and technology (how shall I address a service in a mobile phone?). In the current Internet, interactions between distributed components are limited by technologies that are static (in structure and even on physical set-up), difficult to change and prone to failure. Development is strongly influenced by the technology platform (web, mobile, etc.), while devel-

[1] The reasons to replicate a service on multiple physical machines range from providing that service at locations distributed around the network, to offering low latency and higher performance when the service is invoked, to spreading the load over multiple machines to avoid overload on one physical machine, and so forth.

opment (and most of deployment) of distributed solutions and systems should be made in a way that is independent of the physical location and underlying technology used (e.g. in the cloud, we should be able to move services from one location to another and the existing compositions should keep working). Detachment of service addressing from the URIs is especially important if the other computational entities in the FI (things, objects, sensors, resources) are going to be described as services.

Services should not be addressable because of the servers that provide them or because of their physical location, but they should be identifiable through “abstract” and unique names (identifiers) associated with the services themselves from the beginning and do not change during the whole lifecycle of the service (see section 3.3). This approach would simply the actual retrieval, selection, and usage of services, but it also requires a proper layer that is able to mask locations and technologies and provide the right abstraction level.

If we consider routing, nowadays we have IP-based routing and there are proposals for content-based routing. The same applies to service identification: the IP/URL-based approach should be replaced by a more abstract, “content-oriented”, and uniform solution.

3.5 Flexibility, dynamicity and adaptability in service delivery

A key element of service design is the separation of the request and delivery protocols and APIs from the actual implementation of services. This opens up the possibility to change and adapt the implementation dynamically. For the time being, however, the desired flexibility and adaptability is realized either at the service requester or the service provider side.

In order to increase the offered level of these non-functional guarantees, there is the need to move this flexibility and adaptability into the service-aware Future Internet. This allows for service infrastructures that are fully independent of service consumers and providers, and opens up space for more comprehensive service computing solutions. The architecture design idea is thus transform service delivery features from being purely application layer components to being – at least conceptually – integrated Future Internet elements.

In other words, the Internet as a platform is further emphasized and the service delivery platform concept is moved closer to the actual data delivery network. This design principle matches the general trend of moving computation closer to data, as for example in computational databases or more recently in comprehensive cloud computing frameworks too.

Eventually, this requires an Internet-scale, open-ended solution to service discovery, composition and management. All together, a Future Internet designed according to this principle allows for a much more flexible and agile computational network that can react more quickly to the dynamics and diversity of today’s Internet working requirements.

3.6 “Shared” access to distributed resource and services

As claimed in the previous section, the Future Internet architecture must be designed with dynamics and diversity in mind. Additionally, the Internet is naturally distributed and no longer only with respect to addressable resources, but also with respect to changing content and functionality, such as large databases and services. The former, in the context of the World Wide Web, was sometimes referred to the Deep Web. A similar effect will also reign the Future Internet when it comes to networked economies or sensor networks.

Having the Internet designed with distributed storage in mind is no longer enough. It will be necessary to incorporate sophisticated notions of content delivery networks, which optimize the access to resources, both static and dynamic – such as services or sensor data aggregators. In other words, the Internet of Services and the Internet of Things, in particular, have advanced needs in terms of novel data management architectures.

Furthermore, certainly manifested through networked economies and service-oriented infrastructures – and hence eventually also Smart Grids and clouds – is the notion of collaborative activities and multi-tenancy. In both cases, although from different perspectives, the data storage and provisioning infrastructure must cope with shared access to distributed and heterogeneous data resources and services (on the fly). In other words, the Future Internet architecture must be designed on the basis of a convergent data routing, data delivery and data management infrastructure.

4 Conclusions

In this paper, with respect to the original design principles of Internet, we presented and discussed some additional design principles enabling the Internet of the Future as envisioned within the Service Offering for the Future Internet (SOFI) community and projects.

The goal of SOFI[2] is to complement EU R&D projects in the area of Internet of Services, Software and Virtualisation (Objective 1.2) through specific support activities. SOFI aims to ensure the position of European research as a leader in the definition and realisation of the theoretical and technological foundations of the Future Internet of Services, as well as European industry's competitive advantage in the creation of value and new opportunities from its use. SOFI will build upon and complement the current efforts around the Future Internet Assembly, and particularly the service related working groups, most specifically the Future Internet Service Offer WG (FISO).

The presented design principles move around the concept of “service awareness” of the Future Internet and cope with quality, naming, addressing, delivery of services as well as shared access to them.

References

[1] B.Carpenter, Architectural Principles of the Internet, Internet Engineering Task Force (IETF), RFC 1958, July 1996.

[2] FIArch document: FIArch Group, “Fundamental Limitations of Current Internet and the path to Future Internet,” December 2010.

[3] J. Saltzer, D. Reed and D. Clark: “End-to-end arguments in system design”, ACM Trans. Comp. Sys., 2(4):277-88, Nov. 1984

[4] T.Moors, A critical review of "End to end arguments in system design, Proceedings of IEEE International Conference on Communications (ICC) 2002, New York City (New Jersey), USA, April/May 2002.

[5] Cross-ETP Vision Document. Available at <http://www.future-internet.eu/fileadmin/documents/reports/Cross-ETPs_FI_Vision_Document_v1_0.pdf>

[6] David D. Clark, Toward the design of a Future Internet. October 10, 2009, http://groups.csail.mit.edu/ana/People/DDC/Future%20Internet%207-0.pdf.

[2] http://www.sofi-project.eu/