

Design Principles of Providing Dependable Services and Service Awareness

Moderators:

- Vito Morreale, ENGINEERING, Italy (previous moderator)
- Matteo Melideo, ENGINEERING, Italy (previous moderator)
- Francesco Torelli, ENGINEERING, Italy (current moderator)

Contributors:

- Luciano Baresi, Politecnico di Milano, Italy (previous expert)
- Marco Pistore, SAYSERVICE, Italy (previous expert)
- Andrew Edmonds, Intel, Ireland. (current expert)
- Massimo Villari, University of Messina, (current expert)
- Dimosthenis Kyriazis, National Technical University of Athens (current expert)

Self-aware Dependability

i) Determine (arguments) WHY you are proposing this new principle, the motivations/incentives/objectives for these additions;

In the current Internet there is a lack of methods and means for dependable, trustworthy processing and handling of network and systems infrastructure with respect to the services they host. This is particularly important in many critical environments, such as health care, transportation and manufacturing where compliance with legal regulations, performance requirements, etc. are of critical importance and must be guaranteed and verifiable.

In future, large-scale Internet-based service deployments, the current and contemporary issues beneath it encountered by consumers will and must be tackled.

- Current Internet services are not a "cure-all" and are not fully cognisant of end-user requirements, especially for enterprises and mission critical applications.
- Current Internet services operate on a "best-effort" basis: there is little consideration for quality, e.g. of service, of experience, etc. Indeed in certain parts of a service stack there is little or no comprehension of quality guarantees.
- Current Internet services are modeled prior to their deployment in any environment and according to the aforementioned modeling, scalability rules and policies are enforced during runtime. Nevertheless and given that infrastructures are application-unaware, the enforced scalability rules and policies are not always adequate to meet the application requirements in terms of efficiency (e.g. in cases of multi-tenancy), performance (e.g. scaling after a specific level doesn't lead to better performance), etc.
- There are little or no service guarantees to the consumer: if any, they are static, inflexible and not negotiable. Often it is left up to the consumer to implement their own systems to ensure guaranteed service.
- Dynamic environments and networks in the Future Internet ask for management policies able to deal intelligently and autonomously with problems, emergent situations, tasks, and other circumstances not necessarily envisaged at the design time.

To address this, the Future Internet and its services must be imbued with the principle of Dependability. This includes the ability of self-management and self-adaptation to cope with changes in the operating conditions and to fulfill the dependability requirements.

ii) Describe accurately the seeds/elements that would compose this new principle (definition, properties/characteristics, etc.)

For all Internet services to be dependable then they share common requirements or constraints and as Eames said “Design depends largely on constraints” and “design is an expression of the purpose”¹ where the purpose is dependability. Dependability is a set of attributes that a service SHOULD (requirement) exhibit in order to be deemed dependable. It should be noted that only a subset of those attributes MUST be exhibited (Reliability and Availability) as these are quantifiable. To be dependable a service must exhibit/implement the following attributes as defined by [1], which is the embodiment of 24 years of work in the IEEE:

- **Availability:** readiness for correct² service.
- **Reliability:** continuity of correct service.
- **Safety:** absence of catastrophic consequences on the user(s) and the environment.
- **Integrity:** absence of improper system alteration.
- **Maintainability:** ability for a process to undergo modifications and repairs.
- **Confidentiality:** absence of unauthorised disclosure of information.

This taxonomy was formed in 2004, however in the age of the Internet of services, where end-users are service- and not product-oriented, this list needs to be updated to better reflect the need for trustworthiness through such capabilities as service inspection, introspection and ultimately a service-aware Future Internet. For this we propose the additional attribute of Transparency:

- **Transparency:** the ability to inspect and introspect a service so that the delivered and guaranteed quality of the service agreement can be verified and observed.

This is essential to tackle the issue of lack of trust many enterprises have with placing their workloads on today's service offerings (be they IaaS, PaaS or SaaS). What these service-oriented customers (e.g. service broker or an end-user) require are all of the above comprised by guarantees that the service provider offers. Service providers may offer and many do today guarantees however currently it is very difficult for the customer to verify these guarantees. It is why the ability to inspect and introspect a service should be offered by the provider, (embodied as transparency) to the customer so that they can indeed validate the guarantee that has been agreed upon between the two parties and compare the agreed versus the observed behaviours of a service. It is also recognized by the UK that such needs should be satisfied especially in the areas of converged Future Internet services [2].

Moreover, the network and the services should be aware of their dependability status versus the agreed/desired level and autonomously manage the gap among them.

¹ <http://www.scielo.cl/pdf/arq/n49/art11.pdf>

² The definition of “correctness” is per [1].

iii) Provide a detailed analysis on impact and consequences on current design and its possible evolution, the implications on the applicability of other design principles, etc. resulting from this new principle

Along with the existing aspects of dependability, this design principle requires a new introspection capability inside the network and at the end-point level (i.e. services) in order to make dependability relevant for today's Internet of Services. It also requires a certain degree of intelligence and self-control aimed at the minimizing the gap between the desired level of dependability properties and the current status of them.

The distribution of these capabilities among the network infrastructure and the end-points should be done with the other design principles in mind.

Indeed some other design principles are related to dependability. For instance, "modularization by layering" principle implies two or more levels of dependability management, i.e. at the single-layer level and at the overall stack level (holistic view).

Moreover "loose coupling" principle leads to the opportunity to manage part of dependability at the end-points level while requiring the management of the rest at the network level.

Finally, the "polymorphism" principle is expected to enable services to self-adapt their offerings according to different attributes (e.g. performance metrics, input data set, management policies, etc). Self-aware dependability will enhance polymorphism by allowing the provision of different characteristics, which will be considered as parameters for the "polymorphism behavior" of a service. Besides, the abstraction and autonomous loose coupling of the services (achieved through polymorphism) will be considered as a means to enable the aforementioned self-aware dependability.

References

[1] A. Avizienis, J.-C. Laprie, B. Randell, and C. Landwehr, "Basic concepts and taxonomy of dependable and secure computing," Dependable and Secure Computing, IEEE Transactions on, vol. 1, no. 1, pp. 11–33, 2004.

[2] E. Townsend, UK Future Internet Strategy Group, "Future Internet Report", May 2011.

Service Awareness

i) Determine (arguments) WHY you are proposing this new principle, the motivations/incentives/objectives for these additions;

The proposed new principle is necessary to fully address two limitations of the current Internet (ref. FIArch Current Internet Limitations March 2011).

The first limitation is related to the requirement that the FI architecture should become **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.

This requires replacing (or complementing) the current data-oriented management with solutions more oriented to services, **even at the layers below application**. 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, which are currently seen just as URIs and/or APIs to be invoked over the network. 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.

The second limitation is related to one specific aspect of service-awareness, i.e., the capability of the FI to handle **multiple Qualities of Service** (QoS) or, as we saw in the previous Design Principle specific declaration of the QoS (e.g. dependability). Also this capability requires that the whole network collaborates to guarantee the the achievement of the QoS, and in particular that **QoS targets established at application level can be transformed automatically into guarantees that must be satisfied/offered at each level.**

Solving these two limitations requires (establishing design principles that support) the definition of suitable abstractions and mechanisms for allowing the cooperation across network levels.

ii) Describe accurately the seeds/elements that would compose this new principle (definition, properties/characteristics, etc.)

Each layer of the stack should be a self-aware and self-managed element that works in isolation to provide its functionalities along with promised guarantees, but at the same time it must cooperate with the others for achieving an holistic service provision.

Cooperation modalities should be defined in order to

1. guarantee that each level is aware, at a certain level of abstraction, of the effects on the levels above of achieving or not its guarantee, and
2. allow each level to negotiate and agree certain guarantees with the levels below.

iii) Provide a detailed analysis on impact and consequences on current design and its possible evolution, the implications on the applicability of other design principles, etc. resulting from this new principle

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.

This principle is strongly related to “modularization by layering”, and should complements it by specifying requirements on the functionalities that each module exposes for supporting cross-layer cooperation.

This principle has to be applied in combination with the “loose coupling” principle, in particular in order to understand and evaluate the effects of cross-layer awareness and cooperation to avoid or minimize unwanted interactions and non-linear effects.

Another principle that needs to be considered refers to “locality”. Given the need to reduce the distance from a process (i.e. service) to the corresponding data, service awareness will contribute by allowing the development of delivery models (applied both to services and data), being enabled through self-management and cross-layer cooperation approaches.