Towards Autonomic and Situation-Aware Communication Services: the CASCADAS Vision

Antonio Manzalini[(1)], Franco Zambonelli[(2)]

[(1)] *Telecom Italia*
antonio.manzalini@telecomitalia.it

[(2)] *DISMI – Università di Modena e Reggio Emilia*
franco.zambonelli@unimore.it

Abstract

The complexity and dynamism of modern network raise several challenges in the design and development of communication services. The unbearable costs in configuration and management call for autonomic approaches, in which services are able to self-configure and self-adapt their activities without human intervention. The need for ubiquity of service provisioning calls for the capability of services of adapting their behavior depending on the current situation (social and spatial) in which they are used. In this paper, after having discussed the need for innovative approaches facilitating the design, development, and execution of autonomic and situation-aware services, we try to analyze the key features that should underlies such a general approach, proposes a general-purpose architecture centered around the abstraction of "agent communication elements", and sketch the main research thrusts that should be pursued for the realization of the vision.

1. The Vision

The Internet as we know it today will have to become like an immense ecology of composite, highly distributed, pervasive, communication-intensive services [KepC03, Zam05]. Such services should be able to: (i) autonomously detect and organize the knowledge necessary to understand the general context – physical, technological, social, user-specific and request-specific – in which they operate; (ii) self-adapt and self-configure their functioning to get the best from any situation, so as to meet the needs of diverse users in diverse situation without explicit human intervention. These features will enable a wide range of new activities that are simply not possible or impractical now. For instance, we expect future generation of communication services to be able to:

- improve our interactions with the physical world by, e.g., providing us with any needed information about our surrounding physical environment and exploiting such information to adapt/enrich their behavior on the basis of the actual environmental characteristics (e.g., consider adapting the behavior of a tourist information service network on the basis of the location from which the service is invoked and of the current weather and traffic conditions) [Est02, Rav05];
- get the best of the network infrastructure and resources upon which they operate, being able to adaptively ensure sufficient quality of service, guarantee their security, and tune to user needs and preferences, independently of the actual network characteristics (e.g., independently of the fact that we require them from a Wi-Fi PDA in a MANET context, from a GPRS phone, from a Bluetooth eye-glass monitor, or from whatever connectivity and connected devices will be available at that time) [CapEM03, MikM04];
- facilitate our social interactions, by properly reflecting and exploiting the social context in which we are currently employing a service, e.g. for mere entertainment, or socialization, or in the context of business activities. Today, many opportunities for social communication and interaction are simply not realized due to a lack of information. Although acquiring and using that information raises security and privacy issues, their careful exploitation will open up a wide range of valuable possibilities for communication services (e.g. simply imagine a number of individual tourists that can be supported in forming a group to obtain discounts or other benefits) [ChoP03].

Turning the above vision into reality is very challenging. It requires a deep re-thinking of our current way of developing and deploying distributed systems and applications, i.e., by conceiving them as to be parts of a sort of ecology and by enabling them to prosper and thrive in it at the service of users. However, it is worth outlining that striving for the vision is not only a necessity for giving better services to end-users, but it is also becoming a compulsory economic urge for service providers and system managers. In fact, the increasing dynamism and variability of communication systems, due to the increasingly unreliable nature of communication links, network nodes, and service nodes (as induced by increasing decentralization and mobility) and to the increase in the number of means via which services can be accessed, calls for re-thinking the rigidity of traditional stack-oriented communication models, by considering that applications offering services will likely need also to exploit knowledge and the lower levels, and that will be possibly supported by dynamically reconfigurable network components that can – at their turn – "understand" the implications of dynamic system changes on applications, and adapt themselves (and/or the overall network structure and policies) accordingly.

As challenging as this can be, proving that the above vision can be effectively realized is the key goal of the CASCADAS project (www.cascadas-project-org, *C*omponentware for *A*utonomic *S*ituation-Aware *C*ommunications *A*nd *D*ynamically *A*daptable Services), started January 1st 2006, and funded by the European Commission. Indeed, CASCADAS attempts at defining a general-purpose paradigm for the development of autonomic and situation-aware communication services, and at showing its feasibility via development of associated tools and demonstrators.

In this paper, we intend to share the results of the thorough analysis work that, during the preparation phases of the project, we have undertaken to reach an assessed and rationale understanding of several aspects related to the above vision. This include: *(i)* identifying a few guiding features that any new general-purpose proposal in the area of autonomic communication and self-adaptive services should properly provide; *(ii)* Identifying a unifying abstraction on which to base a new paradigm, and the necessary software tools revolving around this abstraction; *(iii)* Organizing the above into a practical reference architecture for the design, development, and execution of situation-aware and adaptive communication services.

2. Founding Features

We have identified a few complementary founding features that we consider as general key enablers for the above vision, and around which any communication services infrastructures of the future should be conceived. The identification of these features starts from rather assessed concepts in the area of modern distributed systems, and generalize them to properly account the specific characteristics of the autonomic and situation-aware communication services vision. Thus: context-awareness must become situation-awareness; self-organization and self-adaptation must converge into a concept of semantic self-organization; scalability must assume the form of self-similarity; modularity must take the form of a new autonomic component-ware paradigm.

2.1 Situation Awareness

The capability of services to autonomously adapt to the context from which they are requested and in which they execute demands the technologies to capture contextual data and at the same time the ability of the system and of applications to effectively exploit this data at the best.

Much of the technology to acquire contextual information is already becoming available, and it will soon become pervasive with the increasingly frequent deployment of sensors, location systems, users and organization profiles, and run-time systems for the monitoring of computational and network resources [Est02, Phi04]. What is still in its infancy and still needs to be properly resolved, however, is the investigation of the principles and the algorithms with which this growing amount of distributed information can be properly organized, aggregated, and made more meaningful, so as to facilitate their exploitation by services.

In other words, we think there must be an evolution from a model of simple context-awareness, in which services are given access to isolated pieces of contextual data, to a model of "situation-awareness", in which services are given access to properly elaborated and organized information representing, in much more expressive yet still simple to be exploited ways, comprehensive knowledge related to a "situation" [BouSZ05, Tum05].

2.2 Semantic Self-organization

There exist basically two complimentary approaches to enforce autonomic behaviors. On the one hand, self-adaptive systems work in a top-down manner. They have a sort of semantic representation of their state, and can evaluate their own behavior and change it when the evaluation indicates that they are not accomplishing what they were intended to do, or when better functionality or performance is possible. On the other hand, self-organizing systems work bottom-up without any high-level representation, based on a large number of components that interact according to simple and local rules and in which a global adaptive behavior of the system emerges from these local interactions.

Both self-adaptive and self-organizing approaches are being extensively studied [KepC03, BonDT99]. In our opinion, self-organization is to highly preferable in highly distributed and decentralized scenarios. Also, self-organization and the algorithms underlying the emergence of adaptive patterns in complex systems have been extensively studied in communications, e.g., in P2P computing [BabMM02, Rat01], ant-based optimization [BonDT99], social networks [AlbB02]. Self-organization algorithms has the potential to act as enablers for service composition and aggregation, employing proven techniques to abstract from their "organic" implementation and derive design principles adapted to the requirements of artificial systems. At the same time, the presence of self-adaptive systems capable of understanding what's happening and proper reacting accordingly (as in the canonical "autonomic computing perspective [KepC03] can hardly be disregarded to ensure proper reactions and adaptations to various situations.

Accordingly, we think that a major advance with respect to most of the prior art is to provide a way to exploit self-organization approaches and enrich self-organizing components with more "semantic" and/or "cognitive" abilities, in the direction of self-adaptation. This raises the important question of evaluating the amount of information that has to be processed individually by system components, versus collectively by the self-organizing group. Our key goal is to preserve the simplicity and robustness of self-organization phenomena while simultaneously bringing the benefits of semantics self-adapation and situation-awareness, to achieve what can be defined as "semantic self-organization".

2.3 Self-similarity

To realize the vision and make its embodiment manageable, any proposed approach must be fully scalable, i.e., its chosen design principles should be practically applicable to small systems (e.g., a few number of homogeneous nodes), as well as to very large systems (i.e., systems possibly made by thousands of heterogeneous nodes and service components).

While traditional approaches to distributed systems mostly focus on performance scalability, when the focus is on the development of autonomic services, one should also consider architectural scalability, i.e., the possibility for the adopted approach to scale in the without any increase in conceptual (and consequently in design and development) complexity.

In this direction, one promising option is to explore the potential of self-similarity, where any complex service can be realized by individual atomic components that self-organize and self-aggregate so as to reproduce nearly identical structures over multiple scales [AlbB02], and eventually to make an aggregated service appear again as if it were atomic. Self-similarity, which have been so far investigated only with regard to the structure and properties of complex social and technological networks [Dil03], may indeed represent be a key enabler also for the composition of complex communication services, as well as for the structuring of complex situational knowledge.

A successful use of self-similarity would carry on two closely related and complementary advantages: *(i)* it would facilitate understanding, description and management of services (due to the same structural and organizational principles being in force at different scales); *(ii)* it would allow "diving" into specific sub-systems whenever necessary, without having to modify abstractions and tools to work at finer levels of granularity.

2.4 Autonomic Componentware

All the above features should federated by a sound "autonomic component" model [KepC03, LiuP04], which should provide both a robust and dynamic modular conceptual framework for building autonomic, self-organizing, semantic services, and to act as abstract and generic reference model for the production of a new generation of programmable communication elements that can be reused at different stack layers (i.e., for the implementation of communication services at both the network layer and at the application layers) .

This component model has to supply proper abstractions and tools to support self-similarity, self-organization and situation awareness. Therefore, autonomic service components will have to be explicitly conceived as situated in a world of situational knowledge, fitted with mechanisms for semantic self-aggregation and composition, and designed so as to promote the emergence of high-level ensembles that exhibit self-similarity independently of scale.

Identifying the specific nature and structure of such a dynamic autonomic component model is not an easy task. A number of and well-established research areas, such as multi-agent systems [ZamJW03], programmable networks, "traditional" component-oriented engineering, as well as more novel service-oriented architectures can provide useful insights and sources of inspiration, but requires leveraging the level of abstraction and the intrinsic support for dynamisms.

3. Abstractions and Tools

Identifying the basic features, does not solve the issue of building a conceptual and practical framework supporting the design, development, and execution of communication services in line with such features. In the CASCADAS project, we propose to face this via the introduction of a specific software engineering abstraction, i.e., that of "autonomic communication element" (ACE), on which to rely for the flexible component-based design and development of any complex communication services, and of all the associated tools.

3.1 The ACE Abstraction

The ACE abstraction represents the cornerstone of our component model, in which the four driving scientific principle will properly converge and around which proper tools can be developed (Figure 1). ACEs should represent the basic unifying component abstraction on which to rely for the development of communication services. ACEs acts as entities that can implement (typically in a distributed way) communication services, and also act and are perceived as service access points.

While we expect service-specific behavior to be integrated in specific ACE classes, the basic ACE model should integrate in all ACEs the capability to autonomously aggregate with each other to provide composite services at their best; it should promote self-similarity in composition through a set of appropriate interfaces ; they should exhibit self-organization capabilities, possibly of a "semantic" (i.e., meaningful) type; to this end, ACEs should be both loci and consumers properly organized multi-faceted knowledge, coming from a variety of sources and sensors, overall leading to situation-awareness.

The ACE abstraction should be the basis for implementing application-level communication services, as well as the basis on which to implement network-level and middleware level services. We expect ACEs to be able to operate with only a very minimal support infrastructure. This includes: the support for the automation of the ACE service life cycle, i.e., the post-development life span of communication-intensive services, including the autonomic and situation-aware deployment, the internal configuration of ACEs; the monitoring of its internal activities and the basic mechanisms for handling internal ACEs events; the provisioning of the basic mechanism to enable inter-agent communication (while the policies and the routing strategies for inter-ACEs communication are expected to be ACE-specific).

The specific definition of the basic ACE infrastructure is naturally tightly related to the investigation of the nature and structure of ACEs, and will be dealt with together with those issues, through a set of appropriate interfaces and mechanisms native to ACEs.

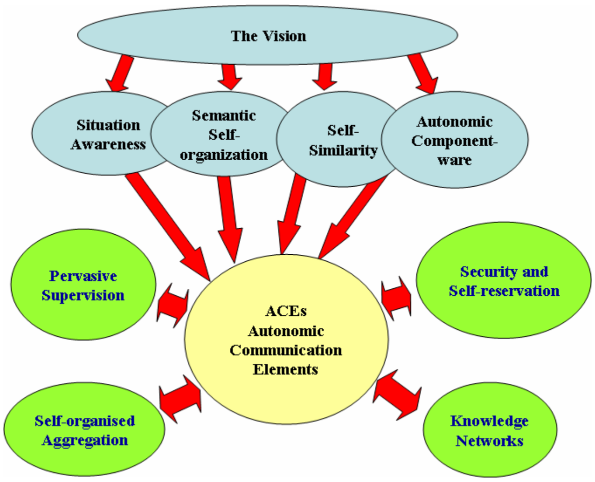

Figure 1: ACEs as the Central Abstraction of a new Paradigm for Autonomic and Situation-Aware Services, Around which Proper Tools can be Organized.

3.2 ACE-based Tools

The key idea of the ACE abstraction is that, beside the described minimal support, any kind of communication service can be implemented via proper dynamic composition of ACEs. In other words, we clearly expect that application-level ACEs will provided with the necessary algorithmic tools, security tools, knowledge tools, and with any needed infrastructural services. However, we also envision that all of these tools and infrastructural services can realized in terms of ACE-based services in their turn, and that all of them will lead to a practical and trust-worth paradigm.

While we expect to be able to implement any needed tools and services via ACEs, within the CASCADAS project we intend to focus on the definition of a few specific services and tools that we consider of a basic paramount importance.

We intent to develop pervasive supervision functionalities across an ensemble of interacting ACEs. Pervasive supervision addresses the runtime construction of an ad hoc and dynamic runtime structure that encompasses a set of cooperating ACEs, and exerts a fully automated and de-centralized control of the communication-intensive service provisioned collectively by those ACEs. This research thrust is primarily relevant to the founding features of self-organization and self-similarity, but clearly also relates to situation-awareness.

We intend to develop algorithms and techniques to achieve dynamic QoS adaptation and enforce given service properties through automated aggregation of ACEs. Smart aggregation will be the basis for identifying and exploring opportunities for co-operation within an ensemble of ACEs, which would allow the collective system to exhibit certain desired properties and to hit situation-dependent QoS targets. This research thrust is primarily relevant to the features of self-organization and situation awareness.

We intend to develop trust, security and self-preservation techniques, an aspect which is of are paramount importance because of the very assumptions upon which the idea of ACEs relies: the heterogeneous nature of the network, the varied capabilities of ACEs, their ability to self-organize and cooperatively supervise each other, which implies the lack of centralized administrative control. Since an ensemble of ACEs possesses those highly dynamic adaptation characteristics, we intend to exploit them to make sure that the resulting system is highly robust and secure, and trust-worth. This research thrust is primarily relevant to the founding features of self-organization and situation-awareness.

We intend to identify and implement models and tools for the organization, correlation and composition of knowledge networks**,** according to to which ACEs can exploit all the available information about their situation, however sparse and diverse. Situation is intended here as a generalization of context, relating to both *(i)* the social-organizational context from which services are invoked (i.e., by a specific users living in a specific social context and accessing the network with specific devices and network technologies); *(ii)* the technological and physical environment in which ACEs live and execute, primarily their networked environment. This research thrust is obviously primarily relevant to the principle of situation awareness, but also represents a common substrate upon which all the other activities will rely, to different degrees.

The scientific and technological objectives devoted to the definition of the ACE model and to the implementation of the associated tools will also be complemented by activities devoted to *(i)* validating the developed approaches and techniques against a set of selected application scenarios; *(ii)* evaluating the social impact of the proposed techniques; *(iii)* properly disseminating the project results and promoting their exploitation.

4. The Architectural Perspective

To acquire a more "operational" perspective of the CASCADAS approach and of its objectives, one can refer to Figure 2, sketching the overall architecture envisioned for our ACE-based autonomic communication network and services.

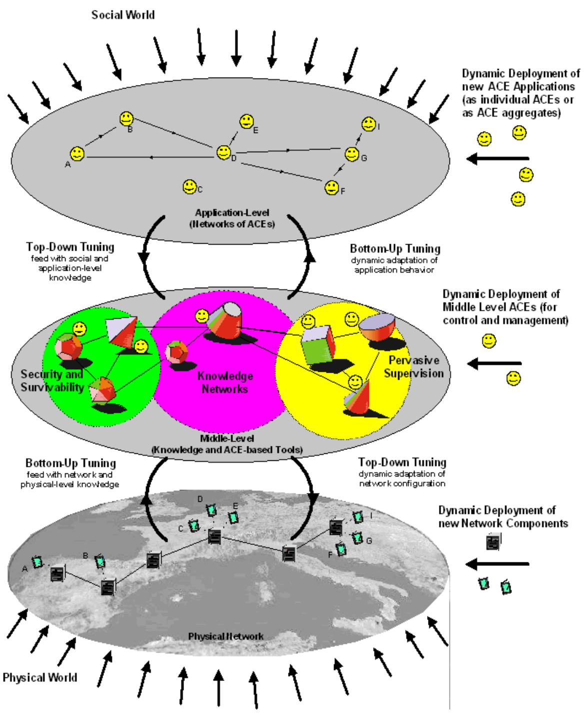

Figure 2. The Architectural Perspective

CASCADAS considers a scenario in which dynamic and heterogeneous networks, possibly enriched with sensors and devices connecting with the physical world, have to host the dynamic deployment and execution of applications and services. Such applications and services have to serve users according to both their social situation and the current network and physical situations. To this end, at the application level, CASCADAS considers developing and deploying application and services (by individuals users as well as by software companies and system managers) in terms of ACE components or of ACE aggregates. These components dynamically self-organize as needed with each other and with the already deployed ones, and will start interacting so as to provide the desired functionality in a situation-aware way without (or with very limited) configuration efforts.

Below the application level, a sort of "middle-level" hosts knowledge (properly organized in knowledge networks) and ACE-based tools to enforce specific properties such as situation-awareness via knowledge networks, semantic self-organization, adaptive QoS, and security. This middle level is fed both by application-level and social-level knowledge (coming from the upper levels) and by network-level and physical-level knowledge (coming from the lower levels), and continuously interact with these levels, in a sort of continuous tuning feedback that ensures adaptability and, thanks to the connection with the lower-levels, also cross-layer tuning. The power of dynamically influencing and controlling the behavior of the network and of the application is guaranteed by the possibility of dynamically injecting in the middle-level proper ACEs components to exert such influence.

The lower levels, i.e., those concerned with actual network architectures and with physical sensing and embedded systems, are not directly within the CASCADAS scope. Still, CASCADAS will take into account the network-level and the physical-level in terms of the information that, from such level, can reach the higher levels and can be exploited to enforce situation-awareness.

5. Conclusions

Evolution of the telecommunications market is likely to be characterized by a dramatic increase of traffic volumes (mainly due to video-based services) whilst Network/Service Providers' revenues are expected to grow very modestly. As a matter of fact competition and traffic growth will determine the need of enhancing current infrastructure whilst reducing costs in order to maintain the telecommunications business sustainable. In order to match these requirements, Network/Service Providers are evaluating to evolve service platforms introducing service-awareness and autonomics features (e.g. self-healing/self-protection, self-optimization).

The purpose of embracing those research thrusts in the project and bringing that kind of advancement to the area of communication-intensive services is multifold. In the first place, and with the shortest-term outlook, we aim at overcoming service platforms main bottlenecks and simplifying the handling of, interconnection of and interaction with the portfolio of existing communication-intensive services, substantially reducing labor and costs. For example two of the current bottlenecks that an autonomic NGN service platform may overcome are fault recovery and dynamic load balancing of the Service Logic Execution Environment. Furthermore, we aim at facilitating the assembly and management of new forms and types of services that are currently too complicated or costly to implement, because of their inherent complexity and situational dynamism.

In the longer run, we aim to help laying some of the necessary foundations and mechanisms that will enable the construction of a repertoire of innovative services that cannot yet be envisaged in the current communication environment, but that will become part of the fabric of a Connected Society in the years to come.

Finally, another paramount benefit that applies to the current landscape as well as any future outlook of communication-intensive service will be the enhanced trust by users/citizens, as well as by the eco-systems of service providers, in the reliability of the services they use, in particular related to the crucial aspects of privacy, security, correctness and guarantee of quality levels.

Acknowledgments. The Authors would like to acknowledge the European Commission for funding the Integrated Project CASCADAS (IST-027807) "Component-ware for Autonomic, Situation-aware Communications, And Dynamically Adaptable Services" (FET Proactive Initiative, IST-2004-2.3.4 Situated and Autonomic Communications). The paper represents the work and contribution of individual parties involved in the project. The authors also wish to thank other project partners for excellent and fruitful discussions.

References

[AlbB02] R. Albert, A. Barabasi, "Statistical Mechanics of Complex Networks", Rev. Mod. Phys. 74(47), 2002.

[AndS04] S. Androutsellis-Theotokis, D. Spinellis, "A Survey of P2P Content Distribution Techniques", ACM Computing Surveys, 36(4):335-371, Dec. 2004.

[BabMM02] O. Babaoglu, H. Meling, A. Montresor, "Messor: Load-Balancing through a Swarm of Autonomous Agents", Proceedings of the 1st Workshop on Agent and Peer-to-Peer Systems, 2002.

[BonDT99] E. Bonabeau, M. Dorigo, G. Theraulaz, "Swarm Intelligence: From Natural to Artificial Systems", New York, NY, Oxford University Press, 1999.

[BouSZ05] P. Bouquet, L. Serafini, S. Zanobini, "Peer-to-Peer Semantic Coordination", Journal of Web Semantics, 2(1), 2005.

[CapEM03] L. Capra, W. Emmerich, C. Mascolo, "CARISMA: Context-Aware Reflective mIddleware System for Mobile Applications", IEEE Transactions of Software Engineering Journal, 29(10):929-945, 2003.

[ChoP03] T. Choudhury, A. Pentland, "Modeling Face-to-Face Communication Using the Sociometer", ACM

Conference on Ubiquitous Computing, Seattle, WA, USA, 2003.

[Dil03] S. Dill, R. Kumar, K. Mccurley, S. Rajagopalan, D. Sivakumar, A. Tomkins, "Self-Similarity in the web", ACM Transactions on Internet Technology 2(3):205-223, 2003.

[Est02] D. Estrin, D. Culler, K. Pister, G. Sukjatme, "Connecting the Physical World with Pervasive Networks", IEEE Pervasive Computing, 1(1):59-69, Jan. 2002.

[KepC03] J. Kephart, D. Chess, "The Vision of Autonomic Computing", IEEE Computer, 36(1), 2003.

[LiuP04] H. Liu, M. Parashar, "Component-based Programming Model for Autonomic Applications", International Conference on Autonomic Computing, New York, NY, USA, 2004.

[MikM04] M. Mikic-Rakic, N. Medvidovic, "Support for Disconnected Operation via Architectural Self-Reconfiguration", 1st International Conference on Autonomic Computing, New York, NY, USA, 2004.

[Phi04] M. Philipose, K. Fishkin, M. Perkowitz, D. Patterson, D. Fox, H. Kautz, D. Hahnel, "Inferring Activities from Interactions with Objects", IEEE Pervasive Computing, 3(4):50-57, 2004.

[Rat01] S. Ratsanamy,, P. Francis, M. Handley, R. Karp, "A Scalable Content-Addressable Network", ACM SIGCOMM Conference 2001, Aug. 2001.

[Rav05] N. Ravi, P. Stern, N. Desai, L. Iftode, "Accessing Ubiquitous Services using Smart Phones", Proceedings of the IEEE International Conference on Pervasive Computing and Communications (PerCom), Kauai Island, Hawaii, USA, March 2005.

[Tum05] L. Tummolini, C. Castelfranchi, A. Ricci, M. Viroli, A. Omicini, "Exhibitionists and Voyeurs do it better: A Shared Environment Approach for Flexible Coordination with Tacit Messages", Environments for MultiAgent Systems. LNAI 3374, Springer-Verlag, January 2005.

[ZamJW03] F. Zambonelli, N. Jennings, M. Wooldridge, "Developing Multiagent Systems: the Gaia Methodology", ACM Transactions on Software Engineering and Methodology, 12(3):317-370, 2003.

[Zam05] F. Zambonelli, M.P. Gleizes, M. Mamei, R. Tolksdorf, "Spray Computers: Explorations in Self-Organization", Journal of Pervasive and Mobile Computing 1(1):1-20, May 2005.