A hierarchical approach to autonomic service management

Jeroen Famaey

Supervisor(s): Filip De Turck, Bart Dhoedt

I. INTRODUCTION

As computer networks grow more complex, heterogeneous and larger in scale, it becomes increasingly difficult for human administrators to manage and configure such systems. Additionally, services offered across the Internet (e.g. YouTube, Facebook, Twitter) have drastically increased in numbers and diversity.

To alleviate the problems associated with managing large, complex and diverse computer networks, researchers have proposed the principle of autonomic communication networks [1]. The goal of autonomic networking is to allow networks to manage themselves, based on high-level business goals set by human administrators. The autonomic elements use knowledge gathered from the network to adapt their behaviour under changing conditions, in order to keep achieving the predefined business goals.

Although some progress has been made towards a completely self-governing autonomic service delivery network, several important challenges remain to be overcome. Among them is scalability, which is defined as the ability of the system to keep functioning properly in face of increasing network size, number of services, and end-users count. In this paper, we introduce a novel autonomic architecture for deliving services across the Internet, which aims to solve scalability issues of current autonomic network architectures.

J. Famaey is with the Department of Information Technology, Ghent University (UGent), Gent, Belgium. E-mail: jeroen.famaey@intec.ugent.be

II. SERVICE DELIVERY ARCHITECTURE

The proposed architecture allows service providers to deliver a set of services with diverse requirements to their clients across the Internet. By way of autonomic networking principles we plan to reduce complexity for human administrators and allow them to manage the system by defining high-level business goals. This hides unnecessary low-level details from the administrators, while allowing them to focus on more important high-level objectives.

As the amount and diversity of network devices and services increases, the autonomic system will require more knowledge to model the network, and more low-level policies to keep operating within the bounds of the predefined goals. To contain this growth and thus improve scalability, we propose to group servers and other network devices in a hierarchy of logical units or clusters. Within a cluster, knowledge, management elements, policies, and services are shared freely. However, only relevant information is propogated to parent or child clusters. This approach limits the amount of detail available in each cluster. A possible hierarchy is shown in Fig. 1. The servers in a single datacenter are clustered together, while the organisation (e.g. a service provider) consists of a cluster of datacenters.

The architecture consists of several important components, which are shown in Fig. 2. The *policy repository* maintains the set of policies that govern the behaviour of the autonomic system. A policy consists of a set of conditions and actions [2]. The actions are triggered when the conditions are met. High-level business

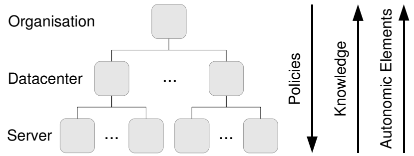

Figure 1. The proposed hierarchical architecture

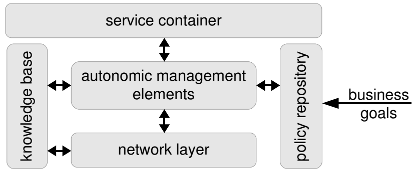

Figure 2. The main components of the architecture

goals are injected by administrators as policies at the higher layers of the cluster hierarchy. They propagate downwards and are translated to more specific policies at each level. For example, the high-level business goal "deliver video in high quality to client X", could be translated to "make sure packet-loss remains below Y for client X" at lower levels.

The *knowledge base* contains the information necessary for the autonomic elements to model the context and make sure everything is working correctly. At the lowest levels of the cluster hierarchy exact information about the network layer and its devices is measured and inserted into the knowledge base. This information is aggregated and summarised before it is propagated to higher level clusters. Additionally, autonomic elements may use reasoning techniques to infer new knowledge from existing information.

The *autonomic management elements* together shape the self-governing behaviour of the autonomic system. Using the knowledge in the knowledge base, they make sure the system operates within the bounds set forth by the policies. Additionally, autonomic elements may themselves infer new knowledge and define new policies, as long as they do not conflict with the predefined business goals. As the elements continuously monitor the state, the system is able to dynamically adapt to changes in the environment and to new or changing business goals. Important tasks of the autonomic elements include maintaining the services running in the *service container* and making sure the requested quality of the delivered services is achieved.

Our current and future work focuses on devising protocols and algorithms for creating and managing the cluster hierarchies, and propagation of knowledge, policies and autonomic elements. In addition, we are developing several autonomic elements necessary for the correct operation of the service delivery architecture. Examples include algorithms for service placement and selection [3], server state management [4], and video quality adaptation [5].

III. SUMMARY

In summary, this paper introduces a novel approach to managing services in a scalable manner. Network devices are grouped in a logical hierarchy of clusters, to contain the exponential growth of knowledge and policies as the system grows in size and complexity.

ACKNOWLEDGMENT

J. Famaey is funded by the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT).

REFERENCES

[1] S. Dobson, *et al.*, "A survey of autonomic communications," *ACM Trans. Auton. Adapt. Syst.*, vol. 1, no. 2, pp. 223–259, 2006.

[2] D. Agrawal, *et al.*, "Policy-based management of networked computing systems," *IEEE Commun. Mag.*, vol. 43, no. 10, pp. 69–75, 2005.

[3] J. Famaey, *et al.*, "A latency-aware algorithm for dynamic service placement in large-scale overlays," in *Proceedings of IM'09*, 2009, pp. 414–421.

[4] J. Famaey, *et al.*, "Dynamic overlay node activation algorithms for large-scale service deployments," in *Proceedings of DSOM'08*, 2008, pp. 14–27.

[5] J. Famaey, *et al.*, "Dynamic QoE optimisation for streaming content in large-scale future networks," in *Proceedings of ManFI'09*, 2009, pp. 128–134.