

Diss. ETH No. 16450

Composition and Deployment of Services in Heterogeneous Programmable Networks

A dissertation submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY
ZURICH

for the degree of
Doctor of Technical Sciences

presented by
MATTHIAS K. J. BOSSARDT
Ing. dipl. en électricité, EPF Lausanne
born March 24, 1974
citizen of Willisau (LU)
and Schötz (LU)

accepted on the recommendation of
Prof. Dr. Bernhard Plattner, examiner
Prof. Dr. John W. Lockwood, co-examiner

2006

Abstract

Starting out as a research tool for a small community of experts, the Internet network architecture has experienced a tremendous success. An important reason is its open communication interface, which allows anyone with basic know-how of the technology to build a wide range of applications and services at the end-systems. The Internet has become a business-critical communication infrastructure and is used as a basis to provide services far beyond the anticipations of its inventors. These new services and the accompanying change of the user community reveal the limitations of the original Internet network architecture. For instance, the Internet core offers a quasi static functionality. Thus, the evolvement of the Internet technology is practically limited to increasing data forwarding performance among and within network nodes.

Research in the field of active and programmable networks has aims at a more flexible network infrastructure. A lot of research efforts target programmable nodes and execution environments for dynamically loadable network services, but address the deployment of such services only partially. In particular, they lack concepts and architectures to manage such services in heterogeneous networks of programmable nodes. Management support for heterogeneous networks is, however, a prerequisite for the acceptance of the technology by network providers.

The main concern of service deployment is to orchestrate the distributed network nodes in a way as to deliver the requested network service. In programmable networks this means not only configuring the devices consistently, but also distributing and installing execution environment-specific code in the form of service components. In this thesis, we develop a service deployment architecture for programmable networks that is able to cope with heterogeneity .

Our architecture takes into account network and node level deployment. At the network level, we introduce the concept of “traffic ownership”, which defines a trust relationship for the different roles involved in service deployment. The role of a traffic control service provider enforces traffic ownership in networks consisting of different administrative domains.

At the node level, our key concept is the “service creation engine”, which maps a generic service description into a node-specific, i.e. native, service implementation. Therefore, the service creation engine processes service descriptors and a “node descriptor”, which is a document specifying the facilities a particular node offers. This mapping mechanism allows coping with the heterogeneous environment, while exploiting the particular facilities of a node.

To validate the architecture and the description languages, we implemented a proof-of-concept system and provide a number of detailed application scenarios in the areas of quality of service monitoring and network security. The validation shows that our architecture is applicable to different realisations of programmable nodes and covers a wide range of application scenarios.

We argue that some security issues of the current Internet architecture cannot be mitigated at the end-systems alone. Countermeasures must be taken within the networks in order to be effective. The legitimate interests of all parties involved in Internet communication must be taken into account by balancing their power to influence the communication process. The application scenarios show that our architecture enhances programmable network technology with necessary concepts to use it in a secure way, to balance what we call the “sender-receiver disequilibrium” in the Internet, and to provide a basis for business models for new dynamically deployable network services.

Kurzfassung

Anfänglich ein Forschungswerkzeug für Experten, ist das Internet und die dazugehörige Architektur zu einem phänomenalen Erfolg geworden. Ein wichtiger Grund dafür ist die offene Kommunikationsschnittstelle, welche es jedermann mit rudimentären Technologiekenntnissen erlaubt, verschiedenste Applikationen und Dienste auf Endsystemen zu entwickeln. Das Internet hat sich in der Zwischenzeit zu einer geschäftskritischen Kommunikationsinfrastruktur entwickelt und bildet die Basis für Dienste, die weit über die ursprünglichen Vorstellungen seiner Erfinder hinausgehen. Diese neuen Dienste und die damit einhergehende Veränderung der Internetnutzer decken die Einschränkungen der ursprünglichen Internetarchitektur auf. So kann beispielsweise die Basisfunktionalität des Internets nur mit Mühe erweitert werden. Folglich beschränkt sich die Weiterentwicklung der Internettechnologie in der Praxis auf die Erhöhung der Datenweiterleitungsleistung in und zwischen Netzknoten.

Die Forschung im Gebiet der aktiven und programmierbaren Netze zielt darauf ab, die Netzinfrastruktur flexibler zu machen. Schwerpunktmässig hat sich die Forschung mit programmierbaren Netzknoten und Laufzeitumgebungen für dynamisch ladbare Netzdienste befasst. Die Handhabung solcher Dienste, welche unter anderem deren Verteilung und Installation umfasst, ist nur partiell untersucht worden. Insbesondere fehlen Konzepte und Architekturen, welche es erlauben, Dienste in heterogenen Netzen programmierbarer Knoten zu handhaben. Managementunterstützung für heterogene Netze ist allerdings eine Voraussetzung für die Akzeptanz der Technologie durch Netzbetreiber, da viele verschiedene Laufzeitumgebungen für unterschiedliche Anwendungsfälle entwickelt worden sind.

Die Bereitstellung von Diensten umfasst die koordinierte Steuerung der verteilten Netzknoten, mit dem Ziel den angefragten Dienst zu erbringen.

In programmierbaren Netzen erfordert dies nicht nur eine konsistente Konfiguration der Geräte, sondern auch die Verteilung und Installation von Programmcode in der Form von Dienstkomponten, welche von der Laufzeitumgebung abhängen. Diese Arbeit beschreibt eine "Dienstbereitstellungsarchitektur" für programmierbare Netze, die mit heterogenen Netzen umgehen kann.

Unsere Architektur berücksichtigt die Bereitstellung von Diensten sowohl auf der Netz- als auch auf der Knotenebene. Auf der Netzebene führen wir das Konzept des "Traffic Ownership" ein, welches die Vertrauensbeziehung zwischen den verschiedenen in der Dienstbereitstellung involvierten Rollen regelt. Die Rolle des "Traffic Control Service Providers" erlaubt es, diese Beziehungen in Netzen mit mehreren Verwaltungsbereichen umzusetzen.

Auf der Knotenebene bildet die "Service Creation Engine" unser Schlüsselkonzept, welches es erlaubt, eine generische Dienstbeschreibung in eine knotenspezifische Dienstimplementation abzubilden. Dazu verarbeitet die Service Creation Engine Dienst- und Knotenbeschreibungen. Bei der Knotenbeschreibung handelt es sich um ein Dokument, das die technischen Einrichtungen eines bestimmten Knoten spezifiziert. Der Abbildungsmechanismus erlaubt es somit, sowohl eine heterogene Umgebung zu handhaben als auch die spezifischen Eigenheiten eines Knotens auszunutzen.

Um unsere Architektur und die Beschreibungssprachen zu evaluieren, haben wir einen Prototypen entwickelt und beschreiben verschiedene, detaillierte Applikationsszenarien im Bereich der Dienstgüteüberwachung und der Netzsicherheit. Die Evaluation zeigt, dass unsere Architektur auf verschiedene Realisierungen von programmierbaren Knoten angewendet werden kann und dass ein breiter Bereich von Applikationsszenarien abgedeckt wird.

Wir legen dar, dass gewisse Sicherheitsprobleme der aktuellen Internetarchitektur nicht alleine durch Massnahmen auf den Endsystemen bewältigt werden können. Um effektiv zu sein, müssen Massnahmen im Netz getroffen werden können. Weiter müssen die legitimen Interessen der verschiedenen in die Kommunikation involvierten Parteien berücksichtigt werden, indem deren Möglichkeiten den Kommunikationsprozess zu beeinflussen ausgeglichen werden. Unsere Architektur erweitert programmierbare Netze mit den notwendigen Konzepten, um diese auf sichere Art einzusetzen, das Ungleichgewicht zwischen Sendern und Empfängern auszugleichen und die Basis für Geschäftsmodelle für neue, dynamische Netzdienste anzubieten.