Towards Carrier-Grade Next Generation Networks

Cornelis Hoogendoorn, Karl Schrodi, Manfred Huber, Christian Winkler and Joachim Charzinski
Siemens AG, Munich, Germany

Abstract

This paper describes the objectives, concepts and solution approach of the KING research project, which is carried out jointly by a number of leading research organisations in Germany. Our overall objective is to develop efficient solutions for carrier-grade IP networks. The distinguishing characteristic of our approach is that we pursue an integrated solution to achieve the carrier-grade requirements for Quality of Service (QoS) and resilience as well as low-cost efficient operation.

1 Introduction

Next generation networks will need to support a variety of services, ranging from high-quality interactive real-time services (e.g. for voice and video applications) to best effort-service as known in today's Internet. The term 'carrier-grade' refers to the properties that are expected of such a network [1]:

- a high QoS for interactive real-time services, being the most demanding class of applications, as well as high-value data services. QoS refers to the combination of low delay and delay jitter as well as low packet loss essential for high-quality transmission
- a high resilience and fast recovery from failures, for example for high QoS services a connection restoration time of less than 300ms after a network failure, as well as other mechanisms that ensure availability comparable to existing telephone networks
- simple management, ease of operation and operational cost reduction through automation of traffic engineering and management tasks

Today's best effort IP networks do not fulfill these requirements. It is our aim to offer an integral solution for the above properties while maintaining the simplicity of established Internet principles.

Acknowledgement. This work was partially funded by the Bundesministerium für Bildung und Forschung of the Federal Republic of Germany under contract 01AK045. The authors alone are responsible for the content of the paper.

2 Key Concepts

The goal is to satisfy QoS and resilience requirements by means of a common approach while at the same time minimizing operational overhead. QoS can of course be provided by resource reservations along the path through the network, but this conflicts with the requirement for fast failure reaction. The reason is that all resource reservations along paths affected by a failure have to be reallocated to other available resources along alternative paths, which is a time-consuming task. Moreover, a non-local reaction to a failure as in today's routing protocols (i.e. finding a new path) is too slow. Our approach to resolving these conflicts is to react locally to failures and to keep the core network stateless by banishing resource-related states to the edge of the network. In case of failure, a stateless core will avoid time-consuming updates of state information, while immediate and autonomous local reaction will shorten the time for selecting an alternative path. Resource management is performed by admission control at the network borders only, no resources are reserved along paths inside the network. A further key feature of our approach is that network control, i.e. the rules for admission control at the edge and traffic engineering within the network, is automated. This relieves the operator from routine network and traffic management tasks and thus reduces operational cost significantly.

The goal to provide QoS in routed IP networks is of course not unique (compare e.g. Diffserv, Bandwidth Broker concepts, IntServ) [2]. The distinguishing aspect of our approach is the simultaneous consideration of QoS, resilience and ease of operation. An overlay network using connection oriented techniques like MPLS [3] can be also used to provide traffic engineering capabilities (but not necessarily QoS) as well as protection switching using statically configured alternate paths. However, abandoning routed IP also means losing the inherent advantages of stateless and connectionless packet networks. By adhering to the connectionless paradigm, we intend to continue and extend the success of routed IP in particular in the areas of resilience and network operation.

Our solution comprises the following components (refer to figure 1):

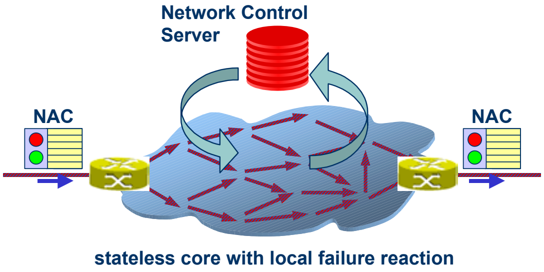

Figure 1: Key components of the KING solution approach

- **Fast failure reaction.** A prerequisite for fast local failure reaction is fast failure detection locally within each node, covering interfaces, links, adjacent nodes etc. This should be in the sub-second range to permit sufficiently fast reaction to failures with minimum impact on QoS traffic.

- **Multipath routing.** Multipath routing has to provide at least 2 alternative paths from each network node towards the destination. This enables immediate local traffic redirection along alternative paths in reaction to link or node failures without having to wait for routing updates.

- **Load distribution.** In addition, multipath routing can be exploited to improve load distribution and prevent hot spots in normal operation as well as after link or node failures.

- **Autonomous network nodes with diffserv-like per-hop-behaviour.** Application requirements and traffic profiles are categorized into a small number of defined classes called network services, and mapped onto suitable traffic treatment profiles.

- **A network admission control entity (NAC), located at network borders only.** The NAC is the termination point for external resource request signaling. NAC resource budgets are calculated such that successful admission implies sufficient capacity in the network core.

- **Redundancy.** The admission control algorithm takes potential failures into account by providing "spare" distributed network capacity for QoS traffic. During normal failure-free operation this spare capacity is of course available for best effort traffic.

- **A network control server (NCS).** The NCS has the task of arranging network-wide traffic distribution and (re-)computing NAC budgets, based on statistical traffic data, network topology information and route information. This task, which requires a global network view, can be performed periodically, for example every 15 minutes, and/or prompted by network status changes (e.g. link or node failure) or changes in observed traffic. NCS will thus relieve Network Management from routine network operation tasks. NCS does not deal with resource reservations and does not keep reservation states, i.e. it does not perform any bandwidth broker functions or other real-time traffic-related tasks. It can therefore easily be replicated for availability purposes.

An illustration of how the desired carrier-grade properties are achieved by means of these components is given in the following example.

3 A Communication Example

Assume a particular network flow (e.g. a real-time video communication) which requires a higher class of service. It must first be assigned to a network service and must then be admitted into this service class before its Quality of Service can be assured. The corresponding admission control communication is sketched in figure 2.

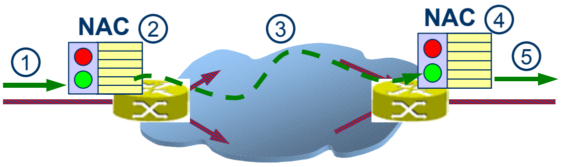

Figure 2: Resource request signaling in a KING network

A resource request (1) from a partner resource signaling agent is received by the ingress NAC of the network domain, indicating the requested service. The ingress NAC instance checks (2) if there is sufficient corresponding network service budget left for the request to be admitted without compromising the QoS of existing flows. If the flow can be admitted, the request will be forwarded (3) to the NAC instance at the egress connected to the next network. This NAC instance checks its egress budget before it will further forward the request to the resource signaling agent of the next network (5). Note that the NCS is not involved in this real-time process.

After the flow has been admitted, user traffic can be distributed within the network as shown below. At the ingress border router, incoming packets are marked according to the network service they will be treated with. Within the network, different packets towards the same destination may take different paths according to traffic distribution weights, as indicated in figure 3 (top).

However, single flows can easily be forced on the same path. By calculating a hash value based on IP source and destination addresses and locally selecting the next hop using this value, packets are still delivered within sequence.

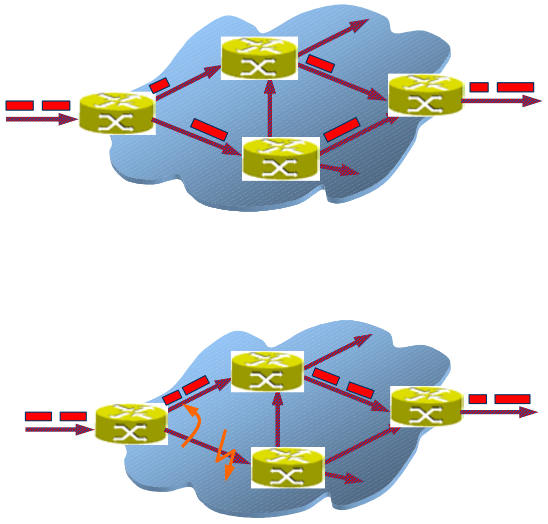

Figure 3: Traffic Distribution before and after a link failure

When a link fails, the nodes adjacent to the link detect this failure using an IP level failure detection technique. The next hop behind the defective link is then taken out of the local forwarding table. The technology of loop-free local alternative paths enables each node to react locally and select one of the remaining next hops towards the destination, as depicted in Figure 3 (bottom). Thus, multi-path routing offers a faster local reaction than conventional IP routing, as routing tables do not have to be re-computed after a link or node failure to restore connectivity.

The NCS will react later (when necessary) by re-computing traffic distribution parameters or NAC budgets if for example the local failure reaction caused significantly unbalanced traffic load in the network.

4 Solution Highlights

The key concepts introduced above outline the KING approach to next generation networks. The following paragraphs discuss selected research topics and summarise results already obtained.

4.1 Network Services

The various applications sending traffic over a network generate very different traffic profiles and at the same time have distinct quality of service demands. As a result the network needs to support differentiated treatment of traffic according to its respective requirements.

To keep the network manageable and to prevent tailoring it to specific applications, the basic mechanisms should be kept as simple as possible. This leads to the well-known concept of network services, which involves categorizing the application requirements and traffic profiles into a small number of defined classes and mapping these classes to suitable traffic treatment profiles.

The chosen treatment regime is strict priorities for traffic scheduling. Accordingly, the highest class network service suffers the lowest delay. In contrast to more complex weighted scheduling schemes like Weighted Round Robin or Weighted Fair Queuing, strict priorities deliver a deterministic behaviour in particular under high load. In normal operation a managed network with admission control allocates the shares of bandwidth to the individual classes and prevents the network from overload. However, a link or node failure can create a bandwidth shortage. With strict priorities the lowest priority class, e.g. a best effort class without explicit guarantees, suffers packet loss first and the quality of the higher classes is entirely preserved.

4.2 Multipath Routing

Multipath routing is the prerequisite for fast local failure reaction and additionally enables fine-grained load distribution. A minimum requirement for resilience is to offer at least two different paths towards the same destination at each traversed node.

Multipath routing has the inherent problem of potential routing loops. In principle there are several ways to prevent loops:

- Restrict the set of valid routes to equal cost shortest paths (e.g. minimum number of hops). This option is applied in OSPF Equal Cost Multi-Path (ECMP) [4].
- Consider not only the destination address but also the source address of the traffic in the routing decision.
- Invent new routing algorithms for destination based routing that avoids loops.

Multipath routes created with ECMP usually do not satisfy even the minimal resilience requirement of at least two outgoing paths at each node. Therefore this scheme is too restrictive to meet our objectives. The main disadvantage of the second approach is its incompatibility with today's IP routing equipment.

For these reasons, we pursue the third option. The challenge is to avoid loops and at the same time observe some given resilience targets. First results indicate that for realistic networks multipath routes with at least two paths per node can be found (in some adverse cases small changes to the network topology may be required).

4.3 Fast Failure Detection and Local Failure Reaction

A fundamental property designed into IP networks is their inherent tolerance to outages. The routing protocols re-establish connectivity, i.e. calculate new routes, whenever links or nodes fail. Because many nodes in a network are involved, this becomes an increasingly time consuming process taking between about half a minute up to several minutes in large networks. In fact there are protocol dependent timers involved which cannot be reduced below a certain limit without introducing the risk of routing instabilities, putting a lower bound of some seconds on achievable rerouting time even with much faster processing.

The major limitation of this traditional approach is the tight coupling between failure detection and (re-)routing. To speed up failure reaction and to enable local failure handling, we separate failure detection/processing from route calculation. This way, failures can be treated locally and immediately without reconfiguration of the whole network.

A second measure is to introduce a fast failure detection scheme. Since routers are operating on the IP layer, failure detection should also monitor the IP layer. "IP-keep-alive" messages frequently exchanged between adjacent routers [5] together with router internal supervision procedures guarantee fast detection and correlation of malfunctions.

The failure handling process triggers a fast local reaction based on the multipath routing approach described above. Connectivity is re-established in some hundred milliseconds thus preserving the quality of service of at least the higher traffic classes.

4.4 Network Admission Control

In contrast to link or path based admission control schemes, the KING network admission control manages resource budgets only for the network as a whole. This method avoids reservation-related state information in the network. Hence flows can take alternative paths inside the network without requiring state reallocation. This is a significant advantage in failure situations or other conditions leading to route changes.

To protect the high-class traffic in case of outages, the calculation of admission budgets (which are defined per network service) also considers defined failure scenarios. The resulting distributed "spare" capacity, which can in fact be used by lower-class traffic under normal network conditions, enables the network to survive failures while maintaining the agreed QoS. The flexible and configurable inclusion of resilience aspects is an important benefit of our admission control process.

4.5 Network Control Server

The purpose of the Network Control Server is to relieve the operator from the burden of permanently observing and regulating the network.

Realistic networks experience rapid and frequent fluctuations in traffic load and other operational conditions. This precludes instantly adjusting the network to maintain some optimal point of operation. The objective should rather be to prevent the network from leaving the zone of dependable operation.

The Network Control Server continuously monitors the network operational parameters (e.g. link load and traffic matrix, statistics of blocked admission requests, failure events, etc.). Its task is to recognise when the network approaches an undesirable state. A decision engine determines the appropriate reaction to avoid undesirable operational states. Possible reactions include redistribution of traffic, reallocation of NAC budgets, optimisation of routing and other parameters. It is important to note that the Network Control Server is not involved in resource reservation or other real-time tasks. Therefore the Network Control Server may fail without affecting short-term network behaviour.

The Network Control Server supports network management by automatically taking care of substantial parts of traffic, performance, configuration and fault management.

5 Summary

The combined application of the key KING concepts provides an integral solution which is especially suited to fulfill the requirements of high QoS and high resilience demanded by interactive real-time services. Research results obtained so far have confirmed the viability of this approach. The next step is to verify operation in real live networks; prototypes of the key control elements NAC and NCS are therefore being developed for inclusion in both lab and field trials.

References

[1] K. Schrodi: ‚High Speed Networks for Carriers', in *Proc. IFIP/IEEE PfHSN*, Apr. 2002, pp. 229 – 242

[2] G. Armitage, 'Quality of Service in IP Networks', MacMillan, Indianapolis, USA, 2000.

[3] E. Rosen, A. Viswanathan, and R. Callon, 'Multiprotocol Label Switching Architecture', RFC 3031, Jan. 2001.

[4] J. Moy, 'OSPF Version 2'. IETF RFC 2328, 1998, Section 16.8

[5] D. Stamatelakis, W. Grover, 'IP Layer Restoration and Network Planning Based on Virtual Protection Cycles', *IEEE J. Select. Areas Commun.*, vol. 18, pp. 1938–1949 (2000)