Department of Communications and Networking

On providing reliable and economical intranet connectivity

Antti Mäkelä

Aalto University

DOCTORAL DISSERTATIONS

Aalto University publication series
DOCTORAL DISSERTATIONS 66/2012

On providing reliable and economical intranet connectivity

Antti Mäkelä

Doctoral dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the School of Electrical Engineering for public examination and debate in Auditorium S1 at the Aalto University School of Electrical Engineering (Espoo, Finland) on the 17th of August 2012 at noon (12 o'clock).

Aalto University
School of Electrical Engineering
Department of Communications and Networking (Comnet)

Supervising professor
Professor Jukka Manner

Thesis advisor
Professor Jukka Manner

Preliminary examiners
Professor Hannu Kari, National Defence University, Finland
Professor Olli Martikainen, The Research Institute of the Finnish Economy, Finland

Opponent
Professor Henning Schulzrinne, Columbia University, NY, United States

Aalto University publication series
DOCTORAL DISSERTATIONS 66/2012

ISBN 978-952-60-4634-1 (printed)
ISBN 978-952-60-4635-8 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)
http://urn.fi/URN:ISBN:978-952-60-4635-8

Unigrafia Oy
Helsinki 2012

Finland

441 697
Printed matter

Aalto University

Abstract

Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

Author
Antti Mäkelä

Name of the doctoral dissertation
On providing reliable and economical intranet connectivity

Publisher School of Electrical Engineering

Unit Department of Communications and Networking (Comnet)

Series Aalto University publication series DOCTORAL DISSERTATIONS 66/2012

Field of research Networking Technology

Manuscript submitted 24 October 2011 **Date of the defence** 17 August 2012

Manuscript revised 5 March 2012 **Language** English

☐ **Monograph** ☒ **Article dissertation (summary + original articles)**

Abstract

In recent years, the Internet and data networks in general have been a huge driver in how communications are conducted. Internet-based communication technologies have transformed the way data is handled and transferred both around the world and within organizations. As the modern society and especially businesses have become completely dependent on reliable network connectivity in their operations, we need to have proper measures to ensure fault tolerance.

Telecommunication systems can be designed in a reliable fashion by using redundant components throughout the network, including items such as multiple different, independent links, ports, and nodes. Additionally, environmental factors, such as the stability of power supply have to be accounted for. Obtaining contractually binding reliability guarantees for a flexible network connection has traditionally been very expensive compared to connectivity without such guarantees. This is because of both implementation costs and low competition among solutions and providers.

This thesis is focused on creating a new approach to network reliability that would have equal performance with traditional approaches, act independently of specific service providers and their chosen access technologies, and be more cost-effective. The thesis analyzes whether such an approach is technically feasible, preserves network usability, and is economically sound. The proposed new approach is called 'RAIIC' (Redundant Array of Independent Internet Connections). RAIIC is based on the assumption that while a cheap, unguaranteed network connectivity may experience outages, several such connections can be bundled together. Such a bundle can then provide equivalent service to the traditional reliability approaches.

The thesis conducts a quantitative analysis on the economical feasibility of the concept, as well as evaluates the feasibility of implementing RAIIC based on Mobile IP technology. The work is concerned primarily in the end-user perception of the networking service.

After considering all the studied factors, including results from simulations, implementation work and possibilities to conduct operations over multiple paths concurrently, the conclusions can be drawn that RAIIC fulfills the design goals perfectly. Even the most demanding of common applications, do not overtly degrade in performance during an outage. The technology has been adopted by the Internet Engineering Task Force as an experimental standard.

Keywords Computer networks, reliability, RAIIC, Mobile IP, intranet, business model

ISBN (printed) 978-952-60-4634-1 **ISBN (pdf)** 978-952-60-4635-8

ISSN-L 1799-4934 **ISSN (printed)** 1799-4934 **ISSN (pdf)** 1799-4942

Location of publisher Espoo **Location of printing** Helsinki **Year** 2012

Pages 255 **urn** http://urn.fi/URN:ISBN:978-952-60-4635-8

Aalto-yliopisto

Tiivistelmä

Aalto-yliopisto, PL 11000, 00076 Aalto www.aalto.fi

Tekijä
Antti Mäkelä

Väitöskirjan nimi
Toimintavarmojen ja taloudellisten intranet-yhteyksien tuottamisesta

Julkaisija Sähkötekniikan korkeakoulu

Yksikkö Tietoliikenne- ja tietoverkkotekniikan laitos (Comnet)

Sarja Aalto University publication series DOCTORAL DISSERTATIONS 66/2012

Tutkimusala Tietoverkkotekniikka

Käsikirjoituksen pvm 24.10.2011 **Väitöspäivä** 17.08.2012

Korjatun käsikirjoituksen pvm 05.03.2012 **Kieli** Englanti

☐ **Monografia** ☒ **Yhdistelmäväitöskirja (yhteenveto-osa + erillisartikkelit)**

Tiivistelmä

Viime vuosina tietoverkkojen ja Internetin yleistyminen ovat muuttaneet ihmisten tapoja kommunikoida. Internet-pohjaiset viestintätekniikat ovat mullistaneet tavat, joilla tietoa käsitellään ja siirretään sekä ympäri maailmaa että organisaatioiden sisällä. Koska moderni yhteiskunta ja varsinkin yritykset ovat tulleet täysin riippuvaisiksi luotettavasti toimivista verkkoyhteyksistä, luotettavuuden varmistamiseen on tärkeää löytää tehokkaita ratkaisuja.

Tietoliikennejärjestelmät voidaan suunnitella luotettaviksi käyttämällä kauttaaltaan varmennettuja komponentteja, esimerkiksi useita erilaisia ja toisistaan riippumattomia yhteyksiä ja laitteita. Lisäksi toimintaympäristöön vaikuttavista seikoista, kuten sähkönsaannin vakaudesta, on huolehdittava. Sopimusoikeudellisesti pätevien luotettavuustakeiden saaminen joustavalle verkkoyhteydelle on perinteisesti ollut erittäin kallista, verrattuna verkkoyhteyksiin ilman takeita luotettavuudesta. Tämä johtuu sekä toteutuskustannuksista että vähäisestä kilpailusta.

Tämä väitöskirja keskittyy uudenlaiseen lähestymistapaan luotettavien verkkojen rakentamiseksi, joka olisi yhtä toimintavarma kuin perinteiset tavat, mutta toimisi riippumatta tietyistä palveluntarjoajista ja heidän teknologiaratkaisuistaan, ja olisi lisäksi taloudellisempi. Väitöskirja tutkii, voisiko tällainen lähestymistapa olla teknisesti toteutettavissa, säilyttää verkon käytettävyyden ja olla kaupallisesti kestävällä pohjalla. Uutta lähestymistapaa kutsutaan lyhenteellä RAIIC, joka tarkoittaa toisistaan riippumattomien internet-yhteyksien joukkoa. RAIIC perustuu oletukseen, että vaikka halpa, ilman takeita toimitettu verkkoyhteys voikin kärsiä katkoksista, niputtamalla yhteen useita tällaisia yhteyksiä, voidaan tuottaa ja tilastollisesti osoittaa erittäin korkeaa suorituskykyä.

Väitöskirja tekee kvantitatiivisen analyysin kaupallisille lähtökohdille ja tutkii mahdollisuutta toteuttaa RAIIC Mobile IP-teknologialla. Työssä painotetaan erityisesti loppukäyttäjän näkökulmaa.

Kun otetaan huomioon kaikki tutkitut asiat, mukaan lukien tulokset simulaatioista ja reaalimaailman toteutuksista useamman polun käytöstä yhtä aikaa, päätelmänä voidaan pitää, että RAIIC täyttää vaatimukset täydellisesti. Vaativimpienkin perussovellusten suorituskyky ei merkittävästi heikkene katkon sattuessa. Internetin standardointiorganisaatio IETF on standardoinut väitöskirjassa esitetyn teknologian.

Avainsanat Tietoverkot, luotettavuus, RAIIC, Mobile IP, intranet, liiketoimintamalli

ISBN (painettu) 978-952-60-4634-1 **ISBN (pdf)** 978-952-60-4635-8

ISSN-L 1799-4934 **ISSN (painettu)** 1799-4934 **ISSN (pdf)** 1799-4942

Julkaisupaikka Espoo **Painopaikka** Helsinki **Vuosi** 2012

Sivumäärä 255 **urn** http://urn.fi/URN:ISBN:978-952-60-4635-8

Preface

Once upon a time, in a certain major Internet service provider I happened to work for, one of the regular-as-pendulum quarterly reorganizations occurred. This time, it was one of those larger ones, causing corporate wide fluctuations as Peter Principle [49] was vigorously applied and followed. In this case, I ended up in a place that was responsible for providing services to enterprise customers – only our new silo didn't have any infrastructure for doing so. Everything had to be subcontracted, including broadband offerings delivered by other departments. At that point, the director gave indication that such sourcing does not need to be limited to internal offerings and that there is slight interest in somehow going for infrastructure-provider independent organization. Thus, certain seeds were planted, and how they grew can be seen in this book.

Of course, all this rhetoric was forgotten at next shuffle a few months later, but by that time the damage was already done: I had attached myself to a research project funded by TEKES (the Finnish Funding Agency for Technology and Innovation) and had been yanked away on my first few trips to the IETF meetings. Those were the first firm steps on a path that I had occasionally danced on in the preceding years, but had never really been entrenched in. What has followed has been an exciting journey through various locales, situations and emotional states. I've experienced many marvelous things, and most importantly met a lot of wonderful people, including the most significant of all: the lady who became my wife, Satu. Thank you for your support and perseverance throughout our entire time together.

I wish to thank Dr. Jouni Korhonen for dragging me to the aforementioned projects and reigniting my desire for completing my post-graduate studies, although our paths later lead to different organizations just as I had started picking up steam in conducting research.

I'm very grateful that my supervisor, professor Jukka Manner, gave me the opportunity, support and facilities for completing this work. Besides Jukka himself, I wish to thank rest of his research team, especially Nuutti Varis for his invaluable views on implementation details and Sebastian Siikavirta for his feedback and collaboration. I also had the opportunity to work with two very talented undergraduate students, Juho Paaso and Aurelien Decros, whose efforts are very much appreciated. Finally, I wish to extend my gratitude towards my pre-examiners, professors Hannu H. Kari and Olli Martikainen, whose comments guided me to make this thesis much better.

To my dearest friend (and best man), Emma Herranen, although you never directly contributed to this thesis, I could not have managed it without your continued strength and support throughout this entire millennium. I also wish to thank everybody else of the "Wednesday group" – you know who you are – for keeping me mostly sane throughout both good and bad times. And of course, my parents, who have managed to cope with all the overhauls in my life that have occurred.

I also wish to express my thanks to my co-authors that have not been mentioned yet: Kalevi Kilkki and Henna Warma.

The funding for the research covered in this thesis has come from Aalto University and the TEKES-funded Mercone project and the Finnish Future Internet research program of TIVIT. In addition, I have received a one-time grant from Research and Education foundation of Teliasonera Finland.

Tampere, April 30, 2012,

Antti Mäkelä

Contents

List of Publications

This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.

I Antti Mäkelä. Concept for providing guaranteed service level over an array of unguaranteed commodity connections. In *The 25th Symposium On Applied Computing (ACM SAC 2010)*, Sierre, Switzerland, March 2010.

II Antti Mäkelä, Jouni Korhonen. Home Agent assisted Route Optimization between Mobile IPv4 Networks. *Internet Engineering Task Force, RFC 6521*, February 2012.

III Antti Mäkelä, Jouni Korhonen. Space-efficient signaling scheme for Home Agent Assisted Route Optimization for use in Virtual Networks. In *The 10th International Conference on Telecommunications (ConTEL 2009)*, Zagreb, Croatia, June 2010.

IV Antti Mäkelä, Jouni Korhonen. Space-efficient signaling scheme for IP prefix and realm information in Virtual Networks. *Infocommunications Journal*, Volume III/2010, Special Issue on Novel Solutions for Next Generation Services, pages 34-45, August 2010.

V Antti Mäkelä, Jukka Manner. Bundling consumer connections - a performance analysis. In *The 1st International Workshop on Protocols and Applications with Multi-Homing Support (PAMS 2011)*, Singapore, March 2011.

VI Antti Mäkelä, Jukka Manner. Performance of an economical, redundant system for intranet connectivity. In *The 16th IEEE symposium on Computers and Communications (ISCC'11)*, Corfu, Greece, June 2011.

VII Antti Mäkelä, Henna Warma, Aurelien Decros, Jukka Manner, Kalevi Kilkki. Economic feasibility analysis of seamless multi-homing WAN solution. In *7th EURO-NF conference on next generation Internet (NGI 2011)*, Kaiserslautern, Germany, June 2011.

VIII Antti Mäkelä, Sebastian Siikavirta, Jukka Manner. Comparison of load balancing approaches for multipath connectivity. *Computer Networks*, Volume 56, Issue 8, pages 2179-2195, May 2012.

Author's Contribution

Publication I: "Concept for providing guaranteed service level over an array of unguaranteed commodity connections"

A single-author paper presenting the concept. Although feedback was sought from multitude of sources, almost all of the work was conducted by the Author.

Publication II: "Home Agent assisted Route Optimization between Mobile IPv4 Networks"

Published as an IETF standard RFC 6521. Highly collaborative effort. The Author's main focus was protocol operation, such as design of protocol fields and operation logic.

Publication III: "Space-efficient signaling scheme for Home Agent Assisted Route Optimization for use in Virtual Networks"

The Author wrote the prefix compression algorithms and conducted experiments related to prefix compression.

Publication IV: "Space-efficient signaling scheme for IP prefix and realm information in Virtual Networks"

This publication extends Publication III. Additional work was mostly a collaborative effort.

Publication V: "Bundling consumer connections - a performance analysis"

The Author designed and implemented the simulation and conducted the experiments. Simulation scenarios were designed collaboratively with the co-author.

Publication VI: "Performance of an economical, redundant system for intranet connectivity"

The Author designed the implementation and experimentation setup.

Publication VII: "Economic feasibility analysis of seamless multi-homing WAN solution"

Much of the work was done collaboratively between the authors. Author provided most of the technical perspective on the paper on current WAN technologies as well as the operating models of service providers.

Publication VIII: "Comparison of load balancing approaches for multipath connectivity"

The Author designed flow-based load balancing algorithms, implemented all load-balancing components, and conducted measurements.

List of acronyms

ADSL	Asymmetric Digital Subscriber Line
AR-TCP	Adaptive Rate TCP
AUP	Acceptable use Policy
BGP	Border Gateway Protocol
CN	Correspondent Node
CoA	Care-of Address
CoT	Care-of-Test
CoTI	Care-of-Test Init
CQ	Call quality
DCCP	Datagram Congestion Control Protocol
DNS	Domain Name System
DSL	Digital Subscriber Line
GNU	GNU's Not Unix
HA	Home Agent
HAaRO	Home Agent assisted Route Optimization
HFSC	Hierarchical Fair Service Curve
HN	Home Network
HoA	Home Address
HoT	Home-Test
HoTI	Home Test Init

HSRP	Hot Standby Router Protocol
HTTP	Hypertext Transfer Protocol
ICMP	Internet Control Message Protocol
IETF	Internet Engineering Task Force
IKE	Internet Key Exchange
IP	Internet Protocol
ISP	Internet Service Provider
IT	Information Technology
LAN	Local Area Network
LQ	Listen quality
MIB	Management Information Base
MIP	Mobile IP
MN	Mobile Node
MOS	Mean Opinion Score
MPLS	Multi-Protocol Label Switching
MPTCP	Multi-path TCP
MR	Mobile Router
NAT	Network Address Translation
NEMO	Network Mobility
PESQ	Perceptual Evaluation of Speech Quality
PSTN	Public Switched Telephone Network
RAID	Redundant Array of Independent Disks
RAIIC	Redundant Array of Independent Internet Connections
RED	Random early detection
REP	Resilient Ethernet Protocol
RFC	Request for Comment

RO	Route Optimization
RR	Return Routability
RTO	Retransmission Timeout
RTP	Real-time Transport Protocol
RTSP	Real Time Streaming Protocol
RTT	Round-trip Time
SDP	Session Description Protocol
SIP	Session Initiation Protocol
SLA	Service Level Agreement
SNMP	Simple Network Management Protocol
SP	Service Provider
TCP	Transmission Control Protocol
TOS	Terms of Service
UDP	User Datagram Protocol
VoIP	Voice over IP
VPN	Virtual Private Network
VRRP	Virtual Router Redundancy Protocol
VSP	Virtual Service Provider
WAN	Wide Area Network

1. Introduction

Telecommunication has always been a critically important component of a well functioning society. As technology has progressed and new communication methods have been adopted, they have quickly become of vast importance for the operations of societies, governments, corporations and individuals. In recent years, the Internet and data networks in general has been a huge driver in the latest shift to the way communications are conducted. Internet-based communication technologies have transformed the way data is handled and transferred both around the world and within organizations.

When implementing components that are considered critical for operation, reliability is extremely important. Reliability considerations can be construed being a part of the design in any system, such as power distribution, logistics and machinery. Likewise, telecommunications systems can be designed to operate reliably. There are usually multiple approaches in implementing reliability for a given system, all with different cost factors and different sets of pros and cons. From economics standpoint, the added extra cost of reliable implementation can be considered an insurance premium against the losses that would be caused due to an outage. Naturally, organizations would prefer to obtain a reliable telecommunication system with as low costs as possible.

Despite the advances in technology, very few solutions on reliable IP networking exist that would be attractive to small entities, such as small and medium-sized businesses. The reliability solutions implemented in traditional fashion, that are deployed by *service providers*(SPs), come with extremely high costs to the customer. In return, the customer receives network connectivity with contractually binding reliability characteristics, with the contract terms being enforced by sanction fees. As such, the implementation costs to the service provider are

considerable as well, since the approach usually means maintaining redundant equipment, physically separated infrastructure and on-call staff to deal with any outages. However, the profit margins have generally been high and pricing structure has not significantly changed over the last few years or even decades.

In contrast, more economical reliability approaches do exist. However, such approaches are generally lacking in flexibility, by placing restrictions on what specific types of network traffic scenarios and types of traffic are possible. Furthermore, the reliability obtained may not reach the levels of the traditional approaches, and are rarely contractually binding. Due to this situation, many smaller entities have chosen to disregard obtaining reliable networking altogether. Network outages even at critical junctures are considered a cost of doing business, despite the fact that a network outage could mean all personnel spending their working hours without achieving any productivity.

This thesis is concerned with the current state, in which the stagnated pricing model of traditional reliability approaches and lack of flexibility of the more economical solutions both contribute to the lack of attractiveness of reliability solutions. It should be considered unacceptable that data networking commonly suffers from outages, in contrast to e.g. traditional telephone network which has seldom experienced failures. Furthermore, most of the existing reliability implementation approaches are highly dependent on specific service providers and their offerings. This work is focused on creating a new approach to network reliability that would have equal performance to the traditional reliability approaches, act independently of specific service providers and their chosen access technologies, and have much lower costs as well. The thesis is attempting to analyze whether such an approach is technically feasible, preserves network usability, and is economically sound.

1.1 Implementing reliability in communication networks

The entire concept of “reliability” in networks is a multifaceted topic, and can concern a number of areas. As such, what constitutes a “network” needs to be defined as well as the reliability itself.

A communications *network* consists of *nodes,* and of *links* connecting such nodes together. When a node wishes to communicate with another node, it sends traffic down a link towards the other node. If there is

no direct link connecting the two nodes, the traffic can be forwarded across multiple nodes, traversing several nodes and links in sequence until reaching the intended destination. This sequence of intermediate nodes and links is known as a *path*. Although these terms apply to several networking technologies, in the context of this thesis they can be considered part of an IP-based network.

A node can be an *end-host*, such as a server, terminal, sensor or similar object that usually does not forward traffic for other nodes. Conversely, the nodes forwarding traffic destined for other nodes form the *infrastructure* of the network. Such infrastructure nodes are routers, switches, proxies, media transformers and the like.

The levels of reliability can vary. *High availability* is a system design approach and associated service implementation that ensures a prearranged level of operational performance will be met during a contractual measurement period, and in the context of networking, is typically associated with the traditional reliability approaches. The prearranged level itself can be negotiated as situation warrants. For more, see Section 2.1. With other, less stringent, approaches, the redundant components may not be immediately available (*cold standby*), or the switchover to backup component may take a considerable amount of time. Furthermore, no contractual obligations are provided for the service itself, although some obligations may exist, such as delivery time of spare parts.

The reliability level required for a node is dependent on the usage of the node. For example, a failure on a workstation only concerns a single user, while a failure on a server node may affect a large number of users. When implementing reliability in a network, the primary objective is usually to provide reliable communications between particular end-host nodes, because the users and services of the network are located at the end-hosts. As such, the nodes themselves and all components on the path between them have to be implemented in a way that serves that objective. Various failure modes and how to prepare for them to facilitate immediate recovery, and as such, high availability, are listed as follows:

1. Environment failure. All nodes are physically located somewhere, for example the servers are typically located in data centers. Even virtual nodes are hosted on physical platforms. An environment failure causes affected nodes to fail due to changes in the operating conditions.

An environment failure can appear in a number of forms, from a relatively benign loss of power to massive natural disasters. To avert the effects of such failure, the local preparations can include items such as uninterruptible power supplies and generators. However, with certain types of environment failures, such as flooding, the only way to achieve reliability is to have a redundant nodes at a physically separate location. In the event of such a disaster, the redundant end-host location would take over. This is known as *site redundancy*.

2. End-host hardware or software failure. A node providing services for the network may experience a malfunction, either due to a hardware issue or a software failure. In this case, the effects may be averted by having one or more separate nodes with identical functionalities available, ready to take over as events warrant. Depending on the exact application, these may be part of a load-balancing cluster or acting only as a standby. This is known as *node redundancy*.

3. Infrastructure node hardware or software failure. This is technically similar to end-host failure, however, since the infrastructure node typically provides services for the network connectivity itself for multiple end-hosts, the ramifications are usually much more serious. Same methods for prevention apply here; multiple sets of equipment to provide redundancy.

4. Link failure. Two nodes that are located adjacent to each other in the network may lose interconnectivity via a specific link. The failure can occur either due to the link itself suffering a breakdown or one of the nodes having an issue specific to the link, such as a port hardware failure. Link failure is typically related to hardware, sometimes caused by external factors, such as construction work breaking down cabling. This type of failure can be mitigated by having multiple links between adjacent nodes. In some cases the redundant links are not necessarily connected to the same nodes, but the nodes have been set up in a ring topology, which is common in e.g. SDH [28]. In such a ring, a failure on a single link still preserves connectivity between all nodes in the ring. This is known as *port* or *link redundancy*.

5. Path failure. A path through a network, across multiple nodes and links, may fail. While technically an aggregation of node and link failures, in this context the significance is that the path is usually operated by a third party. For example, a service provider implementing a path between several sites in the organization may experience failures. From the perspective of the organization, the failure has occurred in the connecting path, but the organization cannot directly affect the restoration of the service or preparedness for outages. To avert such scenarios, multiple paths can be used. A service provider may be contractually obligated to implement such alternate paths, as in the case of traditional reliability solutions, or the organization may choose to use multi-homing via several providers. This is known as *path redundancy*.

In the context of the research, the focus is on the reliability of the network infrastructure. As such, although it is clear that the redundancy for other components has to be implemented, the research focus is not in the reliability of end-host equipment or environment, where a long-established best practices and procedures already exist. Furthermore, even within the infrastructure, preparedness for individual node and link failures are an area that has been well studied and has a long operational history.

The focus of the research lies in the path redundancy. While node and link failures are obviously taken into consideration, the overall purpose of the research was to study effective switching between multiple, unreliable, redundant paths between sites of a single organization and provide a reliable overlay network over such environment.

1.2 Current service provider environment

Networking services are typically planned, designed, implemented and operated by entities known as service providers, also called network operators. Individual service providers are highly diversified, based on such variables as customer segmentation, subcontracting relationships and service portfolios. However, the core product can always be considered to be wide area networking, either private connectivity between physical locations of each customer (intranet), connecting several customer networks together (extranet) or providing access to the Internet.

In addition, a service provider may have products including all kinds of services that are related to the network, such as authentication services, LAN design and IT management. However, these services have more of an ancillary role.

The core product of any service provider, the network connectivity, has a large pricing range depending on the specific features included. Network connectivity products have three rough tiers in terms of pricing and feature sets: basic consumer-grade Internet access, business-grade services, and connection with reliability guarantees.

The first tier, the basic consumer-grade Internet access, can be very low-cost indeed. However, customization options for the customer are typically limited to the desired last-mile connection speed and access technology. The consumer-grade product comes without any contract-mandated reliability guarantees whatsoever on the availability of either the network itself or additional services such as helpdesk. Despite this, the connectivity generally works well, as dissatisfied customers can change their providers if the possibility exists (multiple service providers in a given area). Furthermore, for example in Finland, the consumer regulatory authorities may enforce penalties for prolonged network failures, even though no such stipulations are provided by contracts.

The second tier are business-grade services. Such services come with a higher price point and much more flexible feature sets as well. A business-grade connectivity is typically highly customizable, where additional options increase pricing level. For example, at the low-end the service may consists of nothing more than the basic Internet access, technically equivalent to the consumer version. Even at the low-end however, additional “value-added” services such as dedicated helpdesk functionalities are usually included. In such case, the pricing is not significantly higher than the basic consumer-grade Internet access. In contrast, at the other end of the spectrum, the service provider might link all customer locations together over a WAN in addition to public Internet access, allow for dynamic routing protocols, implement traffic classification, prioritization and shaping, and so on. All these kinds of premium options increase monthly prices.

The highest tier is providing the connectivity itself with specific guarantees on behavior of the network, such as the overall reliability, maximum latencies and minimum throughput. Compared to the previous

tier which may have included some guarantees not related to network operation, such as specific customer service response times and a timely access to a dedicated account manager, providing guaranteed functionality for the connection itself requires extra resources from the service provider. These extra resources drive up both capital and operational costs. The extra resources come in variety of forms. For example, the provider has to maintain at minimum two redundant, physically independent links to customer premises, allocate a stock of spare parts for fast replacements in case of equipment failures and have processes in place to rapidly respond to outages. Consequently, such guarantees may increase pricing by several orders of magnitude, even if contractually binding reliability guarantees could otherwise be considered a single feature tacked on to an otherwise typical business-grade service contract.

A huge gap exists between the second and third tiers of pricing. The research detailed in this thesis attempts to provide a "best of both worlds" solution, allowing for reliable WAN connectivity with more affordable prices. Such solution would create a "tier 2.5" market segment between the regular, unguaranteed connections and the highly expensive reliability solutions.

The proposed new approach in this thesis is a concept called Redundant Array of Inexpensive Internet Connections (RAIIC). RAIIC is named as such due to the several analogies with Redundant Array of Inexpensive Disks (RAID) that has a long operational history in the field of storage. Depending on exact configuration, both technologies can be focused to provide either performance or reliability. Just like in RAID, which creates an overlay of reliable storage architecture on top of several unreliable disks, RAIIC creates an overlay of reliable networking on top of several unreliable links and paths.

The original target application of RAIIC was illustrating how a service provider could implement reliable intranet connectivity for corporations and other entities with lower costs than the traditional high availability approaches. The thesis covers the background and issues in current service provider environment, and studies on the economical impact of RAIIC if deployed. In addition, a possible technical approach for implementing RAIIC with Mobile IP is introduced, analyzed and discussed. The discussion includes aspects on how to maximize utilization of available networking resources.

While the economic intranet connectivity can still be considered primary use-case, the RAIIC approach may also be applicable in other environments, such as data centers. These additional use cases are discussed in Chapter 6.

1.3 Relationship to other work

The research in this thesis concerns three related but distinct technology areas: multi-homing of networks, seamlessly switching between available connectivity, and load balancing across the available paths in such multi-homed networks. These three areas are each, individually, rather long-standing topics. However, the research in this thesis attempts to combine these areas together for the purpose of providing a single, economically feasible service.

Multi-homing has been used for redundancy for a long time. If a specific path fails, traffic can be switched over to any existing backup paths. Protocols such as Virtual Router Redundancy Protocol (VRRP) [42] can be used to facilitate switchovers in case of equipment or link failures when all of the network elements are managed by a single entity. On a network level, the typical method involves multi-homing via several service providers by utilizing dynamic routing protocols. For more details on existing multi-path approaches, see Section 2.1.

Load balancing across multiple paths is also not a new concept. However, most of the approaches covering heterogeneous paths that are flexible within changing network conditions are intended as end-to-end solutions, such as with transport layer protocols like Stream Control Transmission Protocol (SCTP) [64] and multi-path TCP [15]. In contrast, when load balancing is implemented within the network infrastructure, it is typically performed in a much more rigid fashion. On the network layer, a very basic set-up [3] involves using routing policies to divide outbound and inbound traffic between different exit points.

Our work is also related to seamless mobility [5], where multiple approaches exist for conducting vertical mobility between different access technologies and networks. Most of the existing work is limited to utilizing single a path – multiple paths may be concurrently used for short periods to facilitate handovers, but most of the time only a single path is used. Notable extensions include, for example, mobility-aware

middleware approach [4], which attempts to provide seamless multi-path routing on a per-application basis.

1.4 Research contributions

The research originally started approaching the problem completely from the technical standpoint. The first objective was to formalize the objectives and to conduct initial testing for the technical feasibility. If the results would be encouraging, the research could be continued and expanded in scope. Therefore, the RAIIC concept was first formally presented in Publication I, along with initial simulation results based on our chosen technology, Mobile IP. The initial research based on very simplistic model and simulations was encouraging, and the efforts were continued with further simulations, and our proposed additions to the Mobile IP protocol were introduced to the Internet Engineering Task Force (IETF) to begin standardization process. The standardization work in the IETF was conducted as part of MIP4 working group throughout the entire research. Publication II, which is now published as an experimental RFC 6521, was the primary contribution to the IETF and gives protocol specification for Route Optimization with Mobile IPv4, a key component when implementing RAIIC with Mobile IP. In Publication III we studied the effectiveness and performance of signaling compression algorithms presented in Publication II. Publication IV extended the study and applicability of the algorithms further, beyond original target application to IPv6.

In the initial work, the simulation approach was chosen due to the possibility to monitor the behavior of the network and all network components in high detail. Their existence, actions, internal logic and behavior could each be individually tracked and conclusions drawn. Publication V details research utilizing simulations. The primary focus of the simulation work was the usability perspective; how well do applications and transport protocols perform when link conditions are not stable due to infrastructure switching to redundant nodes, and the switchover time being as mandated by the technology.

After the simulation work was completed to satisfaction, we had enough confidence for attempting to implement the technology in a real world environment. Our implementation work allowed testing of the scheme with real-world applications and equipment, both validating

the simulations and revealing additional issues. The implementation work published in Publication VI also allowed us to refine the protocol specification that ended up in Publication II while it was still being worked on in the MIP4 working group of the IETF.

The implementation work validated most of the simulation results, and the results indicated that the signaling components of the technology work satisfactorily. As such, the technical focus could be turned on load balancing. Since there are multiple paths available for use, would it be possible to utilize them concurrently? Publication VIII consists of the work related to multiple concurrently utilized paths. At the same time, since the technical feasibility started to appear solid, the economics had to be taken into consideration. In Publication VII the effects of the technology were studied in detail, mostly through a qualitative analysis of different kinds of business models for a new market stakeholder, the Virtual Service Provider.

As can be seen, we have conducted the research in three distinct areas that complemented each other. The structure of the research and how the publications correspond to each focus area can be seen in Figure 1.1. In addition, the research spawned a couple of Master Thesis works: a more general study on economics of WAN connectivity [10], and the programming of implementation [43] used for the work in Publication VI.

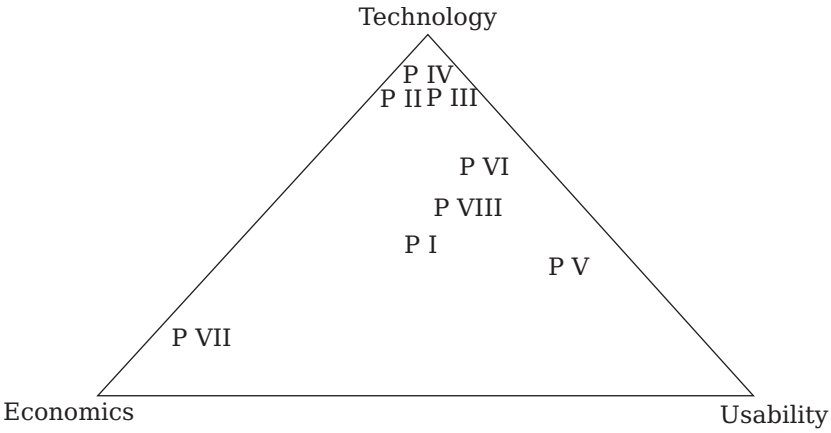

Figure 1.1. Focus areas of the included publications.

1.5 Structure of the thesis

The rest of the thesis is structured as follows: In Chapter 2, the RAIIC concept and how it relates to existing technologies, as well as the economics considerations, are presented. Chapter 3 covers the method for implementing RAIIC utilizing the Mobile IP architecture. In Chapter 4 we evaluate the performance of the Mobile IP-based RAIIC system, first individually in terms of scalability, reliability, responsiveness and effectiveness of load balancing, and then compare the performance to the existing, guaranteed reliability approaches. Chapter 5 discusses additional topics of interest and the ramifications of our work. Finally, Chapter 6 offers the conclusions and finishing statements of the thesis.

2. The RAIIC concept and economic feasibility

As shortly stated in the introduction chapter, a number of issues exist for a customer desiring reliable WAN connectivity. To address these issues, the RAIIC (Redundant Array of Inexpensive Internet Connections) approach attempts to be a feasible method for implementing reliable infrastructure at moderate costs. In this chapter, issues with the current state of affairs are explored further, and the RAIIC concept is introduced in detail as a possible solution. Beside the outline of the concept itself, the economic feasibility of RAIIC is discussed.

2.1 Issues with the current reliability approaches

In a scenario where an entity, such as a company, wishes to obtain WAN connectivity, a number of technical approaches are available depending of the exact use-case. As stated in the introduction, such connectivity is obtained from a *service provider* (SP). Service providers differ highly in operating methods, networking technologies and available products. The low-end products can be as rudimentary as basic Internet access for consumers, and some service providers can be highly focused on such offerings. At the opposite end of the spectrum there are providers that design, implement and operate highly customized solutions for large conglomerates. A service provider may also choose to focus on specific networking technology, or have several approaches that provide services over a number of different access methods, both wireless and wired. A service provider may have complex subcontracting relationships for conducting different kinds of work, but ultimately the SP is responsible to the customer for providing the connectivity.

WAN connectivity can be used for a number of purposes, but the two most prominent use cases are linking customer premises together as

a private network (intranet connectivity) and providing access to the Internet. Sometimes the sites being linked together may belong to different entities, in which case the external networks are known as extranets. From the technical perspective of the service provider, all these uses can be considered identical, and are implemented by creating links from the customer premises to the premises of the service provider, and then using said links to forward traffic to various destinations. The policies on higher levels, such as routing tables or firewall rules, affect the forwarding decisions and create the differentiation between intranet, extranet and Internet connectivity.

A *Service Level Agreement* (SLA) exists between the SP and the customer and is included in the contract when purchasing connectivity, although it may sometimes be implicit. The SLA includes specifications pertaining to the connectivity characteristics, such as availability of technical support, faulty hardware replacement times, and connectivity availability. The SLA terms have a broad range. At low-end, an SLA may simply state that all services are provided "as is", or best effort, and the customer has no recourse for service outages apart from changing providers or possible appeal to regulatory authorities. In contrast, at high-end, an SLA may include very specific acceptable ranges for a number of parameters, such as response and resolution times for helpdesk, maximum downtimes, specifications on network characteristics such as delay and jitter, and so on. If the specified performance is not reached, agreed-upon sanction fees or similar penalties are due for the service provider.

Each connectivity comes with a varying number of services, depending on the level of customization. Most of the services and features included in a connection add linearly to the costs, for example public IPv4 addresses, dynamic routing, traffic classification, and so on. However, the notable exception concerns network reliability. If provisions for reliability are included in the contract, the pricing increases are very steep. The expressions for reliability vary: Such statements as "15 minutes of downtime a year" or percentage indications, such as "99,99% uptime" or "four-nines" reliability, are possible. The higher the reliability guarantee, the more it raises the prices. Gartner postulates [57] that each "9" tacked at the end of reliability percentage increases the price by 30%. The high difference between basic, unguaranteed connectivity and reliable connectivity is illustrated in Figure 2.1.

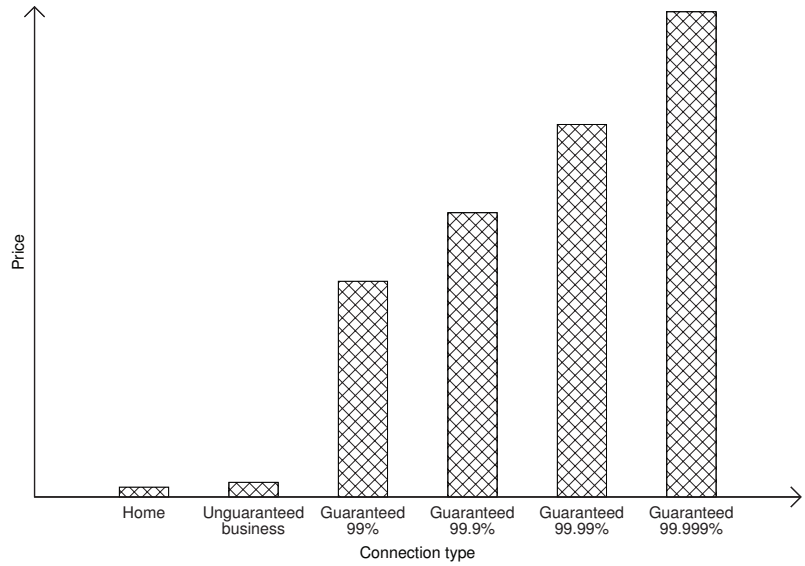

Figure 2.1. Conceptual illustration of pricing differences between different kinds of connectivity types.

One of our interviews in Publication VII illustrates the huge increase in price when reliability (of any level) is included in service. This interviewed customer is operating in southern Finland, obtaining connectivity from a large Scandinavian service provider. The customer is charged 4500 € per month for a guaranteed, reliable 10 Mbps connection to the extranets of their partners. In contrast, the same customer only pays 1500 € per month for their intranet, which consists of an unguaranteed gigabit connection between their two premises. The pricing gap is clearly apparent: A connection with reliability guarantees but only 1% of the bandwidth of the unguaranteed connection costs 3 times more. The pricing structure is observable in historical data as well. The 3-tiered pricing structure has remained relatively constant for a long time, while technology has advanced and access speeds have increased.

This traditional high availability approach obtained from a single service provider is typically implemented with redundant hardware on both provider and customer sides (in provider- and customer edges) and physically separated links. In case one unit or the link between them fails, a backup immediately takes over, with typically sub-second switchover times. To the end-users this kind of redundancy is very transparent, and typically maintained with such protocols as Virtual Router Redundancy Protocol (VRRP) [42]. However, the service provider needs to be able to provide the connectivity to all physical locations of

the customer, which limits the choice of service providers to the ones with most coverage. Certain frameworks such as IP eXchange (IPX) [17] may allow for contracts that span several providers, however, these are relatively rarely used.

As an alternative to the traditional approach with a single provider, the mentioned Gartner report [57] recommends utilizing a less costly approach for reliability: multi-homing via several service providers concurrently by utilizing dynamic routing protocols. In the multi-homing approach, several less expensive connections from multiple providers are obtained, as well as a provider-independent [56] IP address space. This address space is then advertised via a routing protocol, most commonly BGP [55]. Although dynamic BGP routing is one of the services that can typically be obtained with relatively cheap price as part of the connectivity, the increased administrative overhead will increase total cost. The overhead comes in the form of competency and time requirements from the IT administration that are needed for managing an own autonomous system (network with own routing domain). Furthermore, the transparency to end-users is not as seamless as with the traditional solution due to the longer outage detection and response times. In worst cases, the time to completely switch over to backup connection can even take a full hour [66], which is relatively long in contrast to the sub-second time frames of the traditional high availability approaches.

Compared to traditional high availability and multi-homing, more cost-efficient approaches in terms of both administrative and monetary costs exist. Examples include purchasing several basic Internet connections and setting up Virtual Private Networks (VPNs) over the Internet, or setting up multi-homed Network Address Translation (NAT) devices. While these approaches have low requirements from network infrastructure or IT operations, their flexibility and thus applicability are highly limited. As an example of the limitations, the mentioned NAT example works by translating the source address of each outgoing application session and connecting the session via a different service provider. If a connection from a specific service provider fails, it will cease to be used for new sessions. However, reaction time for outages are long, and sessions need to be re-established after an outage. As applications have to re-establish sessions, end-users will notice the breakdowns, sometimes due to an explicit error messages. Although it is possible

to work around some of the limitations, such as handling incoming connections by dynamic DNS updates [69], the overreaching issues still remain.

Another example of limitations in flexibility are visible in the VPN-based approaches. Compared to the NAT method, a benefit of VPN approaches is that they allow flexible data transmission with less restrictions on traffic patterns, as data between internal networks is encapsulated and tunneled. Therefore, with respect to forwarding data, the virtual network can be considered functionally identical to one established with traditional approaches. However, the drawback is that VPN approaches require centralized points (management systems and VPN concentrators) for setting up all the VPN sessions, thus creating bottlenecks and a centralized (sometimes single) point of failure. Another issue is that dynamic reconfiguration in case underlying infrastructure changes is also hard. A relatively common approach is to use an unguaranteed business-grade intranet connectivity backed up by a VPN over the Internet, which can be considered a variant of the aforementioned multi-homing set-up with similar issues.

A summary of the various existing connectivity approaches and their benefits and drawbacks are shown in Table 2.1.

Table 2.1. Pros and cons of different connectivity approaches.

Traditional high-availability solution with SLA guarantees	Pros	Excellent service level Well-established method
	Cons	Highly expensive Limited by coverage of individual service provider infrastructure
Multi-homing via several providers	Pros	More economical than traditional high availability Service provider independence
	Cons	Competence requirements Failover convergence time
Economical approaches: NAT, DynDNS, VPNs	Pros	Economical Simple to deploy
	Cons	Visible to end-users when outages occur Recovery time for inbound traffic

2.2 RAIIC approach

The RAIIC approach, the new possibility proposed in this thesis, attempts to actualize the benefits of all reliability approaches, combining the low costs of economic solutions with the flexibility, transparency, and guarantees of the traditional high availability solutions. Thus, the RAIIC approach attempts to act as a market enabler as well; customers which have not previously considered reliability solutions attractive either due to cost or lack of flexibility might now be convinced to invest in reliable connectivity instead of considering outages as costs of doing business.

The RAIIC approach was comprehensively introduced in Publication I. RAIIC is based on the idea that although unguaranteed broadband connectivity has no *explicit* (and costly) guarantees, they still work adequately most of the time and remain operational, as otherwise dissatisfied customers would flock to other providers. Several such unguaranteed connections are utilized by the RAIIC approach to create a reliable overlay network, or virtual WAN, in an attempt to resemble traditional high availability set-ups. If any one of these connections fail, it ceases to be used and traffic is diverted to an alternate path or paths. This approach of utilizing several inexpensive elements to mimic behavior of a single, more robust one, has been previously seen in for example RAID in hard disks.

Utilizing multiple paths is not a new concept and therefore the RAIIC approach can be considered deceptively simple. However, multipathing is not typically done within the network infrastructure and with end-host transparency, except in a very rudimentary fashion (such as link aggregation with homogeneous links in data centers). More flexible and adaptive multipath transport protocols such as MPTCP [15] and DCCP [34] are implemented inside end-hosts, not as part of the network. However, most end-hosts only support the traditional TCP [52] and UDP [50] protocols, and just like with any other WAN technology, there should be no special requirements or changes to end-hosts beyond basic TCP/IP stack. RAIIC attempts to fulfill a common requirement when deploying new technologies: integration into current business environment with minimal or non-existent changes to current applications, servers and terminal devices. All these issues add to the complexity of implementing and deploying RAIIC.

To avoid the end-host changes, RAIIC has to be implemented as part of the infrastructure. In the RAIIC approach, gateway routers that connect local networks to WAN will generate a virtual, reliable, overlay network over the Internet. The routers need to fulfill certain functional requirements for the overlay creation. The requirements are as follows:

1. Capability for transmitting and receiving arbitrary data, from arbitrary end-hosts in local network to another, arbitrary end-hosts in remote networks.

2. Possibility to establish direct site-to-site connections where possible. As shown earlier, problem with VPN and other centralized, star-topology approaches is creating a single point of congestion. Thus, for any reasonably sized network, at least partial mesh is required. Full mesh may not be necessary if traffic patterns allow it, for example with small sales offices that usually do not communicate directly with each other.

If the Internet connections were *completely* stable, these two requirements alone would be enough to create the overlay. The configuration of the overlay would of course be static and considerations on the exact nature of the underlying Internet connections would need to be made for full-mesh connectivity. However, since the concept is based on using unguaranteed connections, dynamic, automatic reconfiguration must be performed in response to the underlying infrastructure changes. Such reconfiguration-requiring changes do not necessarily need to be outages in the connectivity itself, as other events may warrant it as well, such as a reallocation of dynamic address obtained from the ISP. Thus, two additional functional requirements are needed:

3. Signaling protocol that keeps the overlay informed of the state of the infrastructure, including information needed to obtain direct site-to-site connectivity. In case a router detects an issue with the network, the information must be communicated to all other routers in timely fashion so that they can react appropriately.

4. Timely detection and recovery from outages. At each router, the conditions of all possible paths to peer routers need to be monitored.

If an outage or other problem presents itself, the traffic needs to be quickly diverted to an alternate path. The maximum allowed recovery time that still preserves transparency to the end-users is dependent on the applications being used.

When the above requirements are fulfilled, a reliable, virtual WAN can be implemented over a dynamically changing infrastructure, reacting to outages and other issues. However, connectivity to each site would still be limited by the characteristics of the whatever connection happens to be active. Typically, all links from a single site are operational and outages are rare, in which case the best path can be selected. An additional, optional functionality requirement allows the customer to benefit from the combined resources of all connections at the same time, not just as hot standbys, and can be specified as follows:

5. Scheduling and load balancing algorithm that bundles all available connections together in a fair yet effective manner, making best possible use of all available paths.

In addition, the RAIIC scheme as a whole needs to ensure data integrity, confidentiality and trustworthiness. These and similar basic functionalities on data handling apply to any intranet connectivity requirement, and thus they apply to RAIIC as well.

2.3 Potential network failure modes

There are essentially three possible places for a failure within a typical network service, obtained from a service provider. Failure conditions can be total or partial, where total usually means loss of connectivity and partial is typically an indication of congestion or other interference. Symptoms for such partial failure can for example be a decrease in available bandwidth. The effects of failures on the RAIIC model vary. The failure potential, from smallest to largest effect are access link, core network and peering point failures.

Access link failures occur when a direct link from a gateway router to the network of the service provider fails. In case of a total failure, the failure can be detected immediately with link-layer mechanisms, and recovery can thus be initiated immediately. If the failure is specific to the

single link, all backup paths that connect via the core network of the same service provider remain unaffected.

When the core or distribution network of a service provider experiences failure, all paths traversing the network of the provider may be affected. The effect on the virtual WAN can be avoided by ensuring that the redundant links are provided by independent service providers. The connectivity can then be conducted via the networks of alternate providers.

The failure scenario with the most far-reaching consequences, that affects multiple service providers at once, is a failure at the peering points between providers. Effect of such failure for RAIIC can be mitigated with a strategy where each site obtains connectivity from same group of independent service providers. Although connectivity between different providers might be affected, internal traffic within the service providers' own networks is still forwarded without issues.

As can be seen, the last two failure modes have slightly contradicting mitigation methods. The overall conclusion is that after the service providers have first been selected, at least one of the service providers should be used at every customer site if possible, even if the other providers are site-specific. This customer-wide SP is then used to avoid connectivity issues caused due to any inter-provider peering problems, while the other SPs provide connectivity in all other situations.

2.4 Feasibility of RAIIC as a business model

Assuming RAIIC works as a technology, the approach should have feasibility as a business model to attract development and deployment. An analysis of the current market environment and the roles present is thus necessary. Affecting the analysis is the possibility to set up RAIIC both as an alternative to current reliable WAN approaches or as a complementary offering.

As a starting point for the analysis, certain identified roles in the service provider business area and their producer-consumer relationships are shown in Figure 2.2, where direction of the arrows shows which role provides services for which. Note that a single entity may take part in several roles, such as a large service provider that may act as an ISP for consumers and businesses, provide IT management services, and maintain and host servers. The end-users shown at the bottom of the

diagram eventually use the networking services obtained by the customer. Ultimately, the end-user feedback will affect the customer's policy and procurement decisions on the networking services, which may in turn affect the service provider operations. The "end-user" does not necessarily need to be a natural person, and may be an autonomous process, server or application as well.

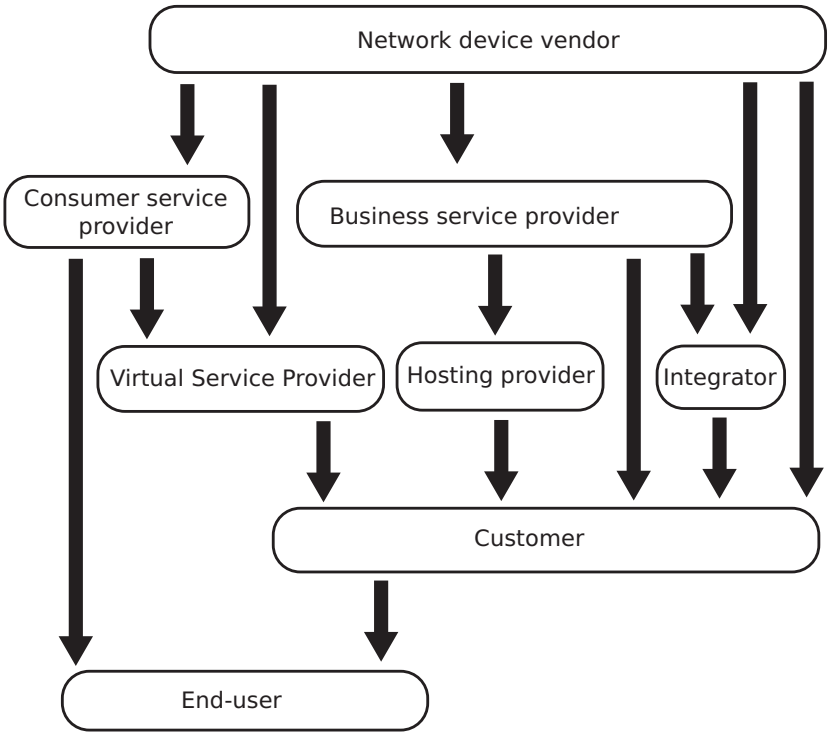

Figure 2.2. The identified market roles. The arrows indicate who is providing services to whom.

As shown in the figure, the customer, such as a corporation, has several possibilities in obtaining network connectivity and other services. The extreme examples are listed below. Typically, the connectivity is obtained as a mix of both approaches.

- Direct approach: the customer purchases equipment directly from a device vendor, the WAN connectivity from a service provider, the IT services from an integrator and the server hosting from a hosting provider.

- Subcontracting approach: the customer purchases the setting up of the connectivity and other related services from an integrator, which obtains

the aforementioned components and services and provides the customer with a "turn-key" solution. The subcontracting can also include several layers, for example the service provider chosen by the integrator may also be responsible for obtaining the network equipment.

Our focus for studying the RAIIC as a business model is the entity shown as "Virtual Service Provider", or VSP. The VSP is the proposed new stakeholder utilizing RAIIC technology. The VSP would purchase cheap Internet connections from the other service providers and implement a reliable WAN service over a virtual network implemented with RAIIC-capable gateway routers. The routers would be obtained from network device vendors.

As previously stated, the customer could also contact the network device vendor directly and then implement RAIIC by themselves. For this reason, the VSP needs to provide added value. This comes in the form of SLA providing explicit, sanction-enforced guarantees for connectivity reliability, just like traditional business service providers.

2.4.1 Offering guaranteed service without owning the infrastructure

As the RAIIC model attempts to provide identical service to traditional high availability approaches, the VSP would be offering an SLA with reliability guarantees to customers. An SLA typically stipulates some sort of sanction fees in case agreed-upon service levels are not met. The question then becomes how can the VSP offer reliable services over infrastructure it does not own, or even control for that matter, and avoid such fees. Furthermore, the parties who own the infrastructure, the existing service providers, may choose to alter their network services to prevent the VSP from operating.

The VSP requires at least some information on reliability levels of basic access service providers in the area as a basis for drafting service contracts. In a simple example, having a basic assumption that each individual link has 2% of downtime, this becomes an uptime of $1 - 0.02^n$ with n links to each site. In case of 3 links, this is already 99.9992% uptime. Of course, uptime is not all that counts, since individual service providers may periodically suffer from other issues such as congested network, which in turn causes jitter and packet loss.

Mitigating issues stemming from utilizing third party infrastructure is possible, even in the cases where the third party also acts as a competitor. However, these mitigation methods may cause extra costs or lost business opportunities for the VSP. As such, resorting to such methods is not desirable, but at least following methods can be considered:

- Offering service only in areas where you can assume smooth operation, for instance metropolitan areas with multiple SPs and relatively quick recovery times.

- Overbooking of redundancy, for example in the previous example of 99.9992% uptime, the VSP can obtain a fourth connection for even more reliability while only guaranteeing the first 99.9992% (or less), thus leaving a higher safety margin for outages.

- Implementation of the service with separate statistics-gathering and normal operation phases. During the statistics-gathering phase the VSP would obtain data on the true behavior of the service provider networks and add or reduce links based on the results. This is typical approach for traditional connectivity as well, in which cases, although the connectivity is operational, the SLA is not considered binding until some time into the operation. This allows the service provider to conduct such operations as reserving a stock of spare parts and verifying the operational capabilities of the connection.

- Having network vendors implement smart heuristics mechanisms on the customer edge routers for making the handover decisions.

As a conclusion, drafting an SLA offering requires a risk analysis depending on the exact situation of each individual customer. Based on such analysis, the feasibility of guaranteed connectivity for each customer can then be estimated, and whether to proceed with the offering and how to tailor it. The risk analysis should consider primarily the factors that affect operations directly, such as environmental conditions and availability of redundant connectivity. However, in some cases additional risks may be related to external factors, for example if the customer is a controversial entity such as a political party or a business that is operating in a contentious field. The specific demeanor of the customer

may act as an invitation for directed denial of service or similar attacks on their networks, which may in turn affect the performance of the connectivity and thus become a liability on the VSP.

2.4.2 Operating environment of a VSP

A Virtual Service Provider in the context of RAIIC is an entity that provides WAN connectivity without owning any of the network infrastructure between their own back-end and the customer sites (hence "virtual"). At the customer sites, the infrastructure of the VSP would consist of the customer edge devices (Gateway routers), while at the back-end the VSP would operate all required support functions.

The VSP would thus implement and operate corporate WAN services to customers. The offerings would consist of planning, design, implementation and operation of high-availability WAN connectivity. The fundamental operation would resemble a consolidation of several traditional roles. To customers, the VSP would appear just like a traditional, business-grade service provider apart from the reduced price in reliability offerings. In this way, the services would be relatively similar. For example, externally, the core product still provides supported connectivity at agreed-upon service levels. Similarly, most of the internal functions, such as helpdesk, customer relations management, and marketing, would be almost identical.

Significant internal differences stemming from SLA requirements demand that the VSP is aware of possible service providers in the desired operating areas and the statistical reliability of their unguaranteed links. The reliability statistics information are to be used for creating guidelines for quality levels in the offerings. In addition, the VSP needs to be able to respond to changing conditions proactively. As an example of such response, if a specific provider shows signs of decreasing reliability, VSP could respond by adding another link in anticipation of outages.

In addition to a highly reduced price compared to traditional reliability solution, additional differentiation that the VSP has over existing providers is flexibility and infrastructure agnosticism. These attributes can be seen in how the VSP does not require the physical network access providers or technologies to be the same at each customer site. Stemming from the flexibility is also the possibility that the RAIIC-based offerings could be even more reliable than the traditional high-availability offerings, especially in areas with multiple providers.

Before a VSP can be established, certain considerations of the operating environment need to be made. The primary requirement is the existence of multiple, independent service providers that operate in the same area as the desired customers, as the entire concept is based on utilizing the infrastructure of several providers and switching paths as necessary. If several providers exist, a number of environmental factors affect the operational costs, and therefore the feasibility of the scheme. Examples of such issues are:

- If a regulation enforcing the sharing of physical infrastructure of service provider (when capacity is available) is in effect, more service providers are typically in operation. More service providers that are available for providing networking infrastructure for RAIIC can be considered beneficial from the perspective of both pricing and capabilities. As a drawback, the VSP needs to be more acutely aware of which specific service providers share infrastructure with each other.

- It is typical for terms of service (TOS) and acceptable usage policies (AUP) of consumer-grade connections to be rather limiting, such as disallowing hosting servers or offering connectivity to third parties (for example outside the immediate family). If such terms are in the service contract, regulations may affect whether these terms are actually contractually binding. Even if they are, the service provider may simply choose not to enforce them. This lack of enforcement can be seen in how consumer-hosted servers are very common even if service contracts have forbidden them.

- Independent access infrastructure between service providers. While a high number of service providers is a benefit for price levels of the infrastructure, it requires more effort to establish truly independent connectivity. The easiest indication is the access technology itself, although a detailed analysis should always be conducted. For example, a cable modem does not share network infrastructure with a xDSL connectivity, and neither of them share infrastructure with a wireless connection. The VSP also has to be aware of shared physical infrastructure, such as different kinds of physical cables located within a shared underground pipe.

- Independent core infrastructure between service providers. Several providers may share a common core of a higher-level network operator. If the core network of the shared network operator becomes congested, all service providers may be affected. Thus, the VSP needs to be aware of such issues.

In harder environments, where there are few providers with strictly enforced TOSes, the issues can be worked around at cost. RAIIC can be implemented over basic, unguaranteed business-grade Internet connections. The cost is higher than with consumer-grade version, but still not exponentially higher as it would be with traditional high availability. If infrastructure sharing is not enforced, the pricing levels are typically higher as well and fewer service providers are available, for example just a single xDSL provider and a single cable modem provider.

2.4.3 Establishing a VSP

As previously mentioned, RAIIC acts as a market enabler by creating a new product at new pricing range: reliability with moderate pricing. The VSP has certain specific expenditures, shown in Figure 2.3. Expenses that are usual to any business, such as office space and support functions are not shown. There are relatively few capital costs. The initial investments consist primarily of setting up the back-end infrastructure. The back-end would also cause the bulk of operational costs. This is due to the need to guarantee continuous operations of the VSP back-end with traditional, expensive high availability methods, although a VSP could of course become a customer for another VSP. Conversely, the costs for statistics-gathering infrastructure for service provider reliability analysis are relatively low. Once the customer base grows, the largest capital costs would stem from obtaining customer access connections. In addition to the initial customer installation, the VSP may also choose to act proactively and install additional connectivity as a precaution. Such precautions could be warranted in cases where the VSP has reasons to expect the reliability of individual connections to decrease, for example due to changes in prevailing weather conditions. In this case, more commonly occurring extreme weather phenomena, such as thunderstorms, would decrease the reliability of individual connections.

Revenue for the VSP would be coming from implementation fees and periodic fees for the virtual connection. The pricing level for the

OPEX (Operational expenditures)
• Back-end infrastructure maintenance • Senior level helpdesk functions • Customer site installation work • Sanction fees
CAPEX (Capital expenditures)
• Purchasing access connections • Gathering statistics on network quality of service providers • Deploying back-end infrastructure

Figure 2.3. Expenditures specific to a Virtual Service Provider.

virtual, bundled connectivity should be distinctly competitive compared to traditional high-availability approaches, but still provide enough revenue to offset the costs. Since the traditional high-availability methods cause such a huge increase in price levels, this is probably achievable. The revenue obviously needs to offset all the costs and offer a satisfactory profit margin.

The most likely approach to adopting the VSP business model will not be a new start-up operating solely as a VSP. A more likely scenario is that an existing stakeholder would like to expand its product portfolio. As such, the three most likely candidates are an existing business service provider, an integrator, or a hosting provider, as they have the best starting points to expand to their respective markets.

A business service provider would simply see the RAIIC approach as a possible way to expand their current reliability offerings to a cheaper price range. Due to their existing role, they already have the processes for service provider operations. Furthermore, it would allow the capability to provide reliability outside the reach of their own network, which could become a competitive advantage. However, RAIIC would essentially be competing with their existing highly priced reliability offerings, and might hurt the overall profitability if not deployed carefully.

An integrator, which is typically an outsourcing partner for the customer, could start operating as a VSP. The RAIIC model could be a lightweight approach to extend into the service provider market. The integrator could obtain basic connectivity from service providers and leverage internal technical expertise to refine it further and provide value-

added, reliable networking service. However, an integrator, especially a smaller entity, might not have capability to transform its processes to continuous and proactive activities required by the VSP model. Integrator processes are typically more oriented towards incident responses.

A hosting provider is already providing high availability services where costs are shared among several customers. Therefore, a hosting provider could avoid the bulk of initial investments by leveraging already existing infrastructure. However, service provider business processes might be too different from existing services and thus difficult to implement effectively.

2.5 Summary

RAIIC works by utilizing multiple unguaranteed, but typically adequately working connections, similar to RAID in hard disks. If certain requirements pertaining to the technical functionality and operating environment are met, the concept appears to be highly feasible from economics standpoint. An entity called Virtual Service Provider could start using RAIIC to set up a virtual infrastructure, and thus start catering to customers desiring an cost-effective yet flexible reliability. A VSP requires expertise in selecting the appropriate service providers for each customer. Considerations, such as expected statistical performance for the individual connections and behavior in each of the possible failure modes, need to be taken into account to allow for establishments of binding SLAs with specific network performance requirements. However, there is a clear indication of a possible new market segment between the highly priced, traditional high availability approaches and completely unguaranteed, basic connectivity or the inflexible reliability approaches.

3. Implementing RAIIC with Mobile IP

To implement RAIIC properly, the functional requirements specified in Section 2.2 have to be fulfilled. The first requirement, transferring arbitrary data between arbitrary end-hosts can be most easily fulfilled with tunnels, as tunnels allow the participating nodes remain relatively unconcerned with the intricacies of the underlying network. In addition, as stated in Section 2.2, the first two requirements alone could be fulfilled with static configuration. However, an unguaranteed, arbitrary network is not a static environment. To fulfill requirements 3 and 4 on fast, dynamic reacting to changing operating conditions, and requirement 5 on load balancing between several paths, an appealing possibility was to use and extend Mobile IP [47]. Since the focus was on deployability in present day Internet, Mobile IPv4 was the primary consideration. Mobile IPv6 [48] has similar features and with sufficient development could be utilized as well.

In this chapter, Mobile IP is introduced and its applicability to RAIIC is covered in detail, including aspects such as signaling, security, scalability, operations and a few alternative approaches. In addition, the functional requirements of RAIIC and how they are fulfilled by MIP are covered in detail.

3.1 Short introduction to Mobile IP

When a node is part of an IP subnet, it is assigned an IP address from that network. Without a valid IP address from the local subnet, the node would not be reachable from other networks due to lack of routing information. Thus, when a node traverses different networks, the address of the node must change, which in turn breaks down existing sessions and

affects usability. Furthermore, the node is not reachable with a static, well-known address.

Mobile IP, in its original form, is a method for nodes to maintain a specific IP address even when connecting via arbitrary networks. Each node has a home network (HN), from which the node operates normally and obtains its address from. In addition, a signaling anchor, known as home agent (HA), is deployed in the HN. The address assigned to the node from the HN is known as the Home Address (HoA).

When a node with MIP capabilities wishes to connect via another network, it is assigned an address from that network. In MIP terms this is known as becoming a mobile node (MN) and connecting via foreign network (FN). However, instead of using the address assigned from the foreign network (Care-of-Address, CoA) for communicating directly, the MN sends a registration request to the HA, stating that the node is now located at the indicated CoA. When the HA accepts the registration request with an acceptance message, the HA begins receiving and forwarding all traffic intended for the HoA of the node via a tunnel. The other endpoint of the tunnel is the CoA, and hence the node. Conversely, the MN now sends all its data via the HA using the same tunnel. If the foreign network hosts an optional service known as the Foreign Agent (FA), the MN may also choose to signal the HA by using the FA as a signaling proxy. The advantage of using a Foreign Agent is that, with an FA present, the MN does not need an IP address from the foreign network at all. In such case, the FA takes care of the registration procedures and the subsequent tunneling with the HA on behalf of the MN.

To external nodes, known as correspondent nodes (CN), all traffic to and from the HoA appears as if the node was still directly connected to the home network. The basic operation of MIP is shown in Figure 3.1. As can be seen, all the communication to and from the MN takes place via the Home Agent.

The home network may also be virtual, that is, the only component in the home network is the HA and no additional nodes ever connect directly to the HN. In such a scenario, all nodes are considered as MNs all the time.

It should be emphasized that Mobile IP, at its core, is only a signaling protocol for establishing virtual connections through a network. The tunnels may use any encapsulation desired: Such encapsulations as GRE [12], IP in IP [45], and UDP [38] are supported. Payload security,

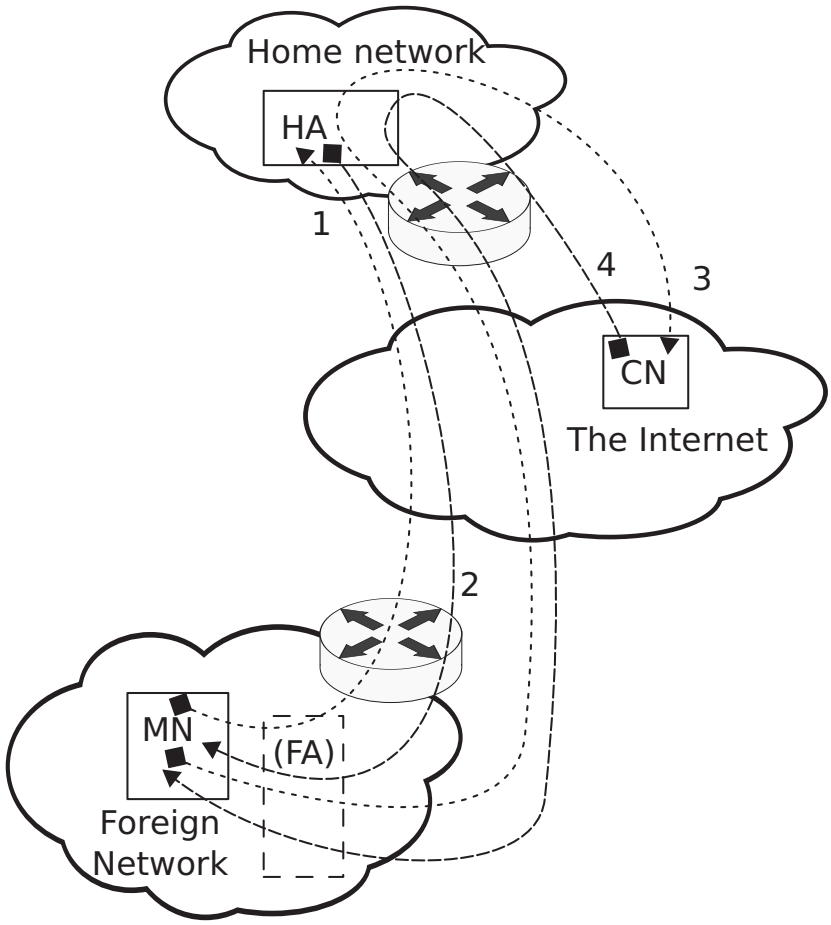

Figure 3.1. Basic Mobile IP operation, consisting of Registration Request (1), Registration Reply (2), and communication to (3) and from (4) a Correspondent Node.

authentication and encryption can be optionally provided by desired security mechanisms, typically using IPSec [33] framework.

3.2 Applying MIP for RAIIC

In its original form, Mobile IP could be considered a lightweight, basic VPN scheme, where client nodes connect to a central server node. The difference to common VPN schemes is that the clients obtain the same address as in their home network. However, the basic MIP protocol was designed to be highly extensible. As a result, numerous extensions were later created. Three of these extensions are of interest when implementing RAIIC with MIP: Network Mobility (NEMO), UDP tunneling, and Home Agent assisted Route Optimization (HAaRO).

Besides the mentioned extensions, there are several others that may be used for additional functionalities, such as identification [31] and integration with IPSec framework with MOBIKE [11]. Furthermore, the optional load balancing requirement may be satisfied with yet another extension [18].

The most rudimentary of RAIIC-related extensions is used for providing mobility to entire networks, not just individual nodes. NEMO [37] extension defines Network Mobility. When a Mobile Node supports NEMO, it can act as a Mobile Router (MR), and represent entire IP subnets to the Home Agent. End-hosts configured to use the MR as their gateway router will not be aware of the mobility taking place.

RFC 3519 [38] defines UDP tunneling. Originally intended for NAT traversal, encapsulation of data into UDP is generally the method that can be best utilized for traversing all kinds of network elements, including firewalls. This is relevant for establishing direct site-to-site connections as efficiently as possible despite obstacles in the underlying infrastructure.

RFC 6521 (Publication II), defines Home Agent assisted Route Optimization for Mobile Networks. The specification was written with RAIIC in mind, allowing for the establishment of full-mesh connectivity in specific cases that apply to RAIIC. Although there is previous work in Mobile IPv6 that included Route Optimization from the outset, the MIPv6 specification only covers RO in the case of individual nodes. The HAaRO specification is designed for facilitating transmission of traffic via optimal paths between entire networks. Furthermore, it addresses some of the end-to-end connectivity issues in the present day Internet.

As a summary, Route Optimization support allows HA to distribute information on networks to MRs as well, so they know which MR is responsible for what network and can connect to them. In a similar fashion to the VPN scheme, HA still exists as a single, centralized point; however, it is now functioning as an anchor point for signaling, not for forwarding data. Thus, congestion issue is averted.

Figure 3.2 shows the full model on how to implement RAIIC with MIP in an example scenario where the customer has 3 sites (A, B and C), and there are three ISPs (1,2 and 3) covering partially overlapping areas. Each site is connected to the Internet via two ISPs. A Mobile Router with an interface to the local network (LAN) and two interfaces for the Internet connections are running as gateway routers on each site. The

Home Agent is operating at back-end of a VSP, which is connected to the Internet with traditional high availability connections (see Section 2.4.2). The nodes implementing the overlay, Mobile Routers and the Home Agent, should also have redundant hardware. In case of hardware failure, existing redundancy methods can be used to assure recovery, such as the previously mentioned VRRP [42].

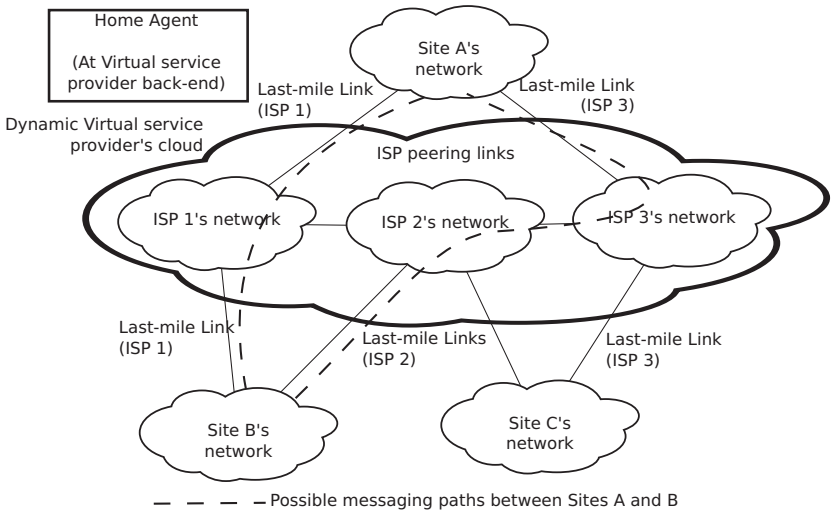

Figure 3.2. Implementing RAIIC with Mobile IP, the discussed example scenario.

3.2.1 Security considerations of MIP-based RAIIC

The lightweight signaling of Mobile IP consists of registration requests sent by the Mobile Routers and responses from the Home Agent and any peer Mobile Routers (also known as Correspondent Routers). Additional, route optimization-related signaling comes in the form of a Return Routability check. The registration requests include information on HoA of the Mobile Router (basic MIP), network address space the MR manages (NEMO) and indication on Route Optimization capability (HAaRO). The responses indicate whether the request has been accepted or rejected, and in the case of Route Optimization, information on networks behind potential CRs.

Since an arbitrary Mobile Router could send out a request claiming to represent arbitrary network, the information has to be verified. To achieve this, the various components have a trust relationship. All Mobile Routers trust information received from the Home Agent, as they have a shared secret. The Home Agent informs the Mobile Routers on which

networks correspond to which Mobile Router HoAs. As such, a Mobile Router can only claim to represent a network that Home Agent has indicated being managed by that Mobile Router.

Return Routability procedure described in HAaRO specification allows MR to establish CoA <=> HoA binding relationship. This leads to the MR being able to send traffic belonging to a certain network to arbitrary address, once it has established that the address is the attachment point for the Correspondent Router. The specification allows usage of pre-shared keys as well and omitting the RR procedure, although the CoA must remain static in such a case. This feature can be used to facilitate Route Optimization between Mobile Routers that are not connected to same logical HA (see Section 5.3).

The end-user data can be protected with existing data security methods, such as IPSec [33], possibly using the aforementioned IKEv2 key negotiation mechanism.

3.2.2 Comparison of RAIIC requirements and MIP

Mobile IP has number of advantages in respect to applicability for RAIIC. The primary advantage is that the signaling is based on a simple request/response approach. While individual request and response messages can become quite large, there is no complicated handshaking or other functionality performed. Thus, signaling message speeds are only limited by path delays.

Mobile IP, when implemented with the NEMO, UDP tunneling and HAaRO extensions, meets the basic functional requirements laid out in Section 2.2.

Requirement 1 specifies capabilities for transparent communication. Mobile IP uses tunneling to preserve transparency as part of the basic specification. There are a number of possible tunneling encapsulations, although due to the requirement 3, UDP encapsulation seems most attractive. The encapsulation may cause fragmentation to occur during transport, but the packets can be reassembled at the receiving router.

Requirement 2 specifies capability for establishing direct connections between sites. With HAaRO extension, this becomes possible. The extension includes provisions for tunnel establishment direction and similar concerns.

Requirement 3 specifies a signaling protocol that allows the network infrastructure to maintain state information. Part of this information

are networks and their reachability. NEMO [37] is an extension to Mobile IP that allows transmission of network information as well. HAaRO specification (Publication II) includes mechanisms for Mobile Routers to receive signaling instead of acting only as initiators, as well as the protocol extensions required to maintain information on point of attachment for each network.

Requirement 4 specifies fast detection and response time to outages. Mobile IP has built-in outage detection in the form of NAT keepalive messages specified in UDP tunneling extension [38]. The specification does not enforce minimum delay between keepalive checks, and the keepalive messages can be sent as often as desired. We reduced the keepalive resending delay to 150 ms at shortest from the default of 110 seconds in our implementation work (Publication VI).

The optional requirement 5 allows for multiple, concurrent paths between customer sites for load balancing purposes. Extension specified in [18] provides this functionality. The specification does not include the exact load balancing mechanism that should be used, leaving this up to the implementation.

3.3 MIP operation when used with RAIIC

The basic operation of MIP-based RAIIC consists of the following functionalities:

1. Distributing information on customer Mobile Routers and the networks each MR manages to potential peer Mobile Routers

2. Setting up tunnels via optimal paths between the Mobile Routers

3. Forwarding end-user traffic to their destinations on the optimal paths

4. Maintaining and reacting to changes in the network based on indications received

The first functionality, information distribution, is relying on the fact that the Mobile Routers are considering information delivered by the HA trustworthy and normative. As an example, if RAIIC approach is used to connect three customer networks, deployed in similar fashion to the

Figure 3.2, the dissemination of information during the initial startup is shown in Figure 3.3.

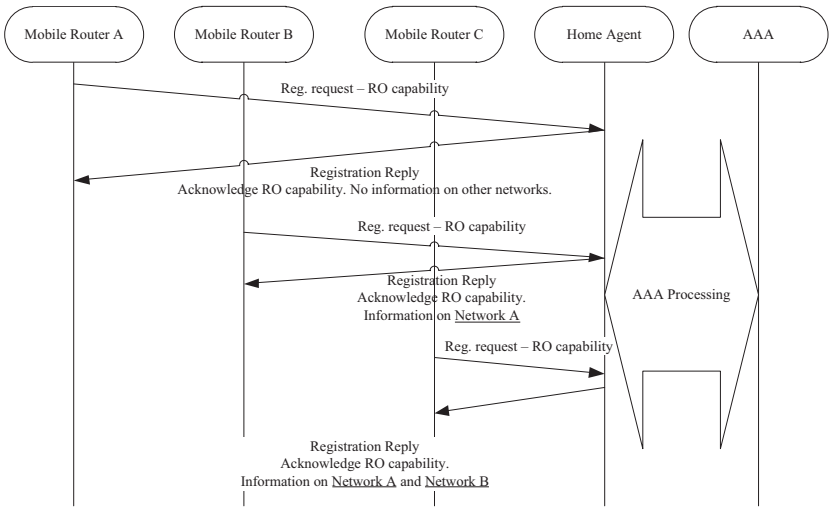

Figure 3.3. Mobile IP signaling during the initial set-up.

In this example case, Mobile Routers are started up in order from A to C and have neither the Mobile Routers or the (continuously operational) HA have any existing state information. The process begins when MR A sends a registration request to the Home Agent via the path it chooses to be most feasible. The registration request message also informs the HA of the network A being located behind MR A. As the HA does not know about any other networks at this point, the registration is accepted, but no additional information is returned to MR A in the following registration reply.

When Mobile Router B subsequently registers, it now receives information on network A being reachable via MR A, or more specifically, via the Home Address of MR A. Finally, when MR C registers, it receives information on both networks A and B and how they are reachable via MRs A and B respectively.

If changes occur in the network causing a Mobile Router to change paths, the MR will re-register to the Home Agent via this new path. Thus, the Home Agent always maintains up-to-date information on network conditions. Additionally, each Mobile Router has to re-register to the Home Agent periodically, by default every 30 minutes, even if no changes occur.

The aforementioned signaling only provides Mobile Routers with information on which network (IP subnet) corresponds to which Home

Address of peer Mobile Routers. This network-HoA binding information is considered something that changes relatively rarely, thus the long delays in information dissemination are acceptable. Furthermore, in the worst case the data will be sent via non-optimal path, however, no data loss occurs.

The second functionality, and in many ways the central piece of RAIIC-enabling MIP is the Route Optimization. Route Optimization means transmitting end-user traffic via a direct path between two networks instead of via the Home Agent, avoiding a single, centralized point of congestion. How such direct path is established with MIP is shown in Figure 3.4. The figure also shows why information on networks does not need to exist on participating Mobile Routers and no mass updates to all MRs are necessary when a new MR joins the overlay. Data that is sent via the Home Agent is marked in red, while data sent via the direct tunnel between Mobile Routers is shown in blue. Unencapsulated data is shown in black.

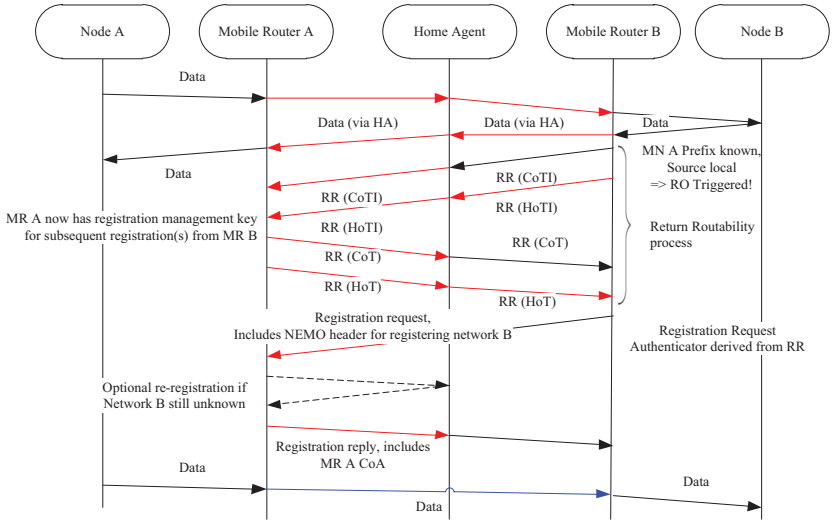

Figure 3.4. Connection establishment in HAaRO using Mobile IP.

The scenario begins right after the signaling in Figure 3.3 has been completed. Right after completion, MR A has not yet conducted a re-registration and thus learned of MR B maintaining network B. In the scenario, node A located in network A wishes to communicate with node B located in network B.

At first, Node A sends a packet (IP datagram) where Node B is marked as the destination, via its gateway router. In this case the gateway is Mobile Router A, which has no specific information on network B. As a

consequence the packet is encapsulated and then forwarded to the Home Agent. The HA knows where network B is connected and forwards the packet onwards to MR B, which decapsulates the packet and forwards it to the final destination.

Upon receiving the packet, node B then sends a reply packet addressed to Node A via its respective gateway router. In this case the gateway router is MR B. Upon receiving the packet, MR B detects that it has information that Network A is located behind MR A. This triggers an attempt in establishment of a direct communication to network A via MR A. To facilitate the direct path, MR B starts signaling for the path establishment. In the meantime, the packet from node B packet is forwarded to Node A via Home Agent. Any subsequent packets will also use the path via HA until a direct path has been established.

Until now, MR B has known that MR A maintains network A. However, MR B is not aware of the Care-of Address(point of attachment) of MR A. To gain such missing information, the Return Routability(RR) procedure is conducted, consisting of CoTI, HoTI, CoT and HoT messages. These messages allow for MR B to credibly establish Care-of Address of MR A and an authentication key to use in next step, which is direct registration request to MR A. The HA does not directly inform the Mobile Routers of the Care-of Addresses due to the possibility of multiple, periodically available links. For faster processing, the Mobile Routers decide on which Care-of Addresses to use independently without involving the HA in the decision-making process.

After the Return Routability procedure has been completed, MR B sends a registration request to MR A, signed with the authentication key established during the RR procedure. This request essentially asks for MR A to send all traffic towards network B directly to MR B instead of via the Home Agent.

Since MR A does not yet know whether MR B truly is maintaining network B, it may now conduct a re-registration to Home Agent to gain up-to-date information on network assignments. After this verification is successfully completed, the MR A can accept the registration. After registration has been accepted, the data will now flow directly from MR A to MR B, without traversing the Home Agent. The same procedure may then be repeated in the opposite direction, or the MR B may choose to forward all traffic towards network A via the direct tunnel even without an explicit request from MR A.

The third functionality concerns transmission of end-user data via appropriate paths. After the paths have been established with MIP signaling, regular tunneling mechanisms and routing table updates allow for packets to be forwarded over the correct tunnels. The path establishment process itself is relatively flexible, as provisions exist for tunnel establishment direction in cases where NAT or firewall may affect tunnel set-up.

The fourth functionality facilitates reacting to network conditions based on indications received. Of such indications, the most important ones concern detection of outages on specific paths. An outage on a path is detected by using keepalive messages. A single keepalive message is a regular ICMP [51] Echo Request sent by an MR to a peer MR or the HA via a tunnel. If no reply is received for three consecutive messages, the path is considered non-operational. As a consequence, routing is reconfigured accordingly, typically by reverting to routing via HA or other direct paths. The failed path may then be re-established using the previously shown procedure.

Additional functionalities can be used to further improve the behavior, such as low-latency hand-offs [39]. Alternate peer authentication mechanisms besides RR procedure are also possible.

3.4 Summary

Mobile IP is a lightweight approach for implementing RAIIC, consisting of signaling protocol intended for network layer operations and tunneling with outage detection mechanisms. As such, it is highly suitable for implementing RAIIC, fulfilling all of the specified functional requirements adequately. Alternatives to MIP exist, however, they appear to have several drawbacks reducing their feasibility. For more on the alternatives, see Section 5.6.

4. Evaluation of MIP-based approach

To validate the technical feasibility of the Mobile IP for implementing RAIIC, we have to consider not only if the functional requirements are fulfilled. The requirements only specify the bare minimum for RAIIC to work at all; performance in different operating conditions has to be evaluated as well. In this chapter, the various issues that affect performance and how to mitigate their effects are discussed. The findings presented in the chapter are the most prominent results of the conducted research. For more detailed results and in-depth analysis, refer to the publications II, III, IV, VI, and VIII.

4.1 Evaluation criteria and approach

When considering how to evaluate the performance, there are several issues that have to be considered:

- Signaling scalability as the size of the network grows
- Effects of the stability of the underlying infrastructure
- Effects on different types of end-user traffic or applications
- Utilization of available resources
- Fairness to users, traffic flows and applications.

For conducting experiments for evaluating the performance of Mobile IP-based RAIIC, there are multiple possible scenarios and network configurations that could be used for measurements. However, most interesting results can be obtained in scenarios with relatively simple network structures. Our base scenario topology includes 3 customer sites, with up to 3 Internet connections each. The small topology makes most of the possible effects on performance observable, in some cases even

more profoundly than a larger one would. The only exception is signaling scalability, which of course requires either a large networking setup or simulations to make relevant observations.

Our experimentations consisted of two distinct methodologies. First of all are the simulations, where every node existing in the network can be strictly controlled and their internal logic easily tracked when needed. Furthermore, simulations allow long timescales to occur in much shorter amount of real time, allowing for adjusting the scenario and parameters in a rapid fashion as observations warrant. The second methodology consisted of conducting experiments with real-world implementations. The implementations consisted of RAIIC-participating infrastructure components such as Mobile Routers. The components were implemented in C programming language and executed on Alix [44] Geode [1] system boards manufactured by PC Engines, each running Debian GNU/Linux operating system.

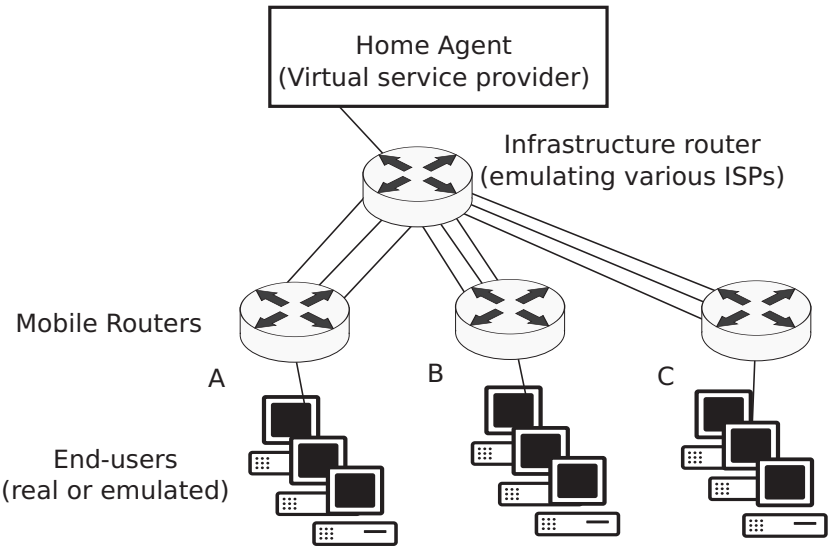

Figure 4.1. The experimentation setup (with the Ethernet switch omitted).

In the case of real-world implementations, the components that are not directly part of the virtual service provider infrastructure can be either real or emulated. Emulation can be used for measuring experiences of a high number of network users. Our chosen method of end-user emulation was Spirent Testcenter [63], an industry-standard network testing platform, which was connected to the infrastructure. For emulating the underlying (non-virtual) infrastructure, a Linux-based infrastructure emulation node was set up. The capabilities of the node include emulating

various link characteristics, such as delay, jitter and packet loss, using netem [22] and emulating different bandwidths with Hierarchical Fair Service Curve (HFSC) [54] queuing disciple. The physical infrastructure is shown in Figure 4.1. An Ethernet switch acting as a central point for all physical connections is omitted from the figure.

4.2 End-user perspective

The basic concept of RAIIC attempts to reach usability levels of traditional, expensive, high-availability approaches. As such, end-user experience and usability considerations are in order. The end-user experience with previously existing reliability approaches were already discussed in Section 2.1.

The usability experience depends on the expectations of the end-user, which in turn vary depending on the application being used. As mentioned before, the end-user does not necessarily have to be a natural person; even automated processes could be designed according to certain expectations, and if these are not met the processes could fail. There are three rough classes of applications and the requirements for data they send over the network:

- Bulk – Applications that do not require interaction or timely responses, such as file transfers or e-mail. Applications use the network facilities in the background; when a user clicks “Send” on an e-mail client, the user does not necessarily know or care when the sent mail is actually underway. For applications sending bulk data, any failures only become apparent during extended breakdowns, that are in the order of several minutes or even several hours.

- Interactive – Responses to activity is required within seconds, such as in the cases of web-based applications or services and instant messaging. End-users do not strictly expect immediate responses, so short breakdowns or transient periods of lower quality are tolerated. Extended breakdowns, in the order of several seconds are apparent and annoying. Shorter breakdowns may also become annoying if they are repetitive.

- Real time – End-user experiences follow network quality very closely, for example with VoIP. Any network quality issues are very quickly reflected in the behavior of the applications, such as hearing noise or other artifacts during a phone call. The quality issues can be somewhat mitigated by various methods, such as different codecs, jitter buffers and similar, allowing for short breakdowns (in the order of hundreds of milliseconds). Despite the mitigation possibilities, the constant quality requirements are generally high.

Besides these three basic classes, applications that do not fit into any of these categories exist as well. For example, certain applications and protocols may have been originally designed for use in LAN environment, and as such, require low network latencies to work properly, even though the end-user might not directly witness any such behavior.

Although there has been considerable research on usability issues concerning different applications, the end-user experience levels vary highly between studies. Furthermore, most of the available research is concerned with stable conditions where experience is relatively constant over time. This model does not work well with the basic concept of RAIIC, which includes a bundle of mostly reliable (although not guaranteed) connections, and rarely occurring outages that are recovered from quickly. Due to the characteristics of RAIIC, there are fewer applicable usability studies, as studies rarely consider the effects of transient, short, usability losses. As an example, a single, short quality degradation during a voice conversation does not necessarily even cause annoyance, as people are used to occasional errors. However, repeated glitches will soon start degrading the overall end-user experience.

Our primary applications of interest are web browsing and VoIP, due to their ubiquitous nature and widespread deployment.

In case of web browsing, a literary survey of research [16] shows that depending on the exact conditions, tolerable pauses may range from 2 to even tens of seconds, depending on the circumstances. However, research on transient delays or failures is hard to come by.

For voice applications, the most commonly used end-user experience metrics are Perceptual Evaluation of Speech Quality [29] and G.107 [26]. PESQ is used to synthetically calculate Mean Opinion Score (MOS), ranging from 5 (Excellent) to 1 (Bad), while G.107 gives an R-Factor ranging typically from 90 (excellent) to 50 (bad). Both measurements

come in CQ (Call quality) and LQ (Listening quality) variants, where call quality gives slightly lower scores as two-way conversations have higher requirements. However, both models have been calibrated for sample sizes of 8-12 seconds, and are typically measured for over the entire duration of the call. When a phone call has perfect quality most of the time and a small pause where quality essentially drops to non-existent, the overall effect on the score is rather low and does not tell the whole truth.

In any case, it can be assumed that transient failures that are quickly recovered from are more tolerable than long, complete failures. Furthermore, the existing user experience studies can thus be used to set an upper limit for failure tolerances since they are based on constant conditions; in RAIIC case the network would only cause occasional issues. The fact that the metrics reflect *Mean* opinion should be emphasized: if such pauses occur once per phone call, the overall score does not decrease significantly. However, if every other syllable is lost due to lost packets, the score drops quickly.

4.3 Signaling scalability

Signaling scalability, alongside end-user experience, is affected by the degree of stability of the underlying infrastructure. If the environment were completely stable, no dynamic detection or responses would be needed at all, as mentioned in the functional requirements in Section 2.2. However, since real-world environments are not completely stable, Mobile IP (and RAIIC in general) attempts to hide the stability issues by directing end-user traffic to paths that are fully operational at each given moment. However, if the operational paths change frequently, the amount of required signaling traffic increases, which in turn affects end-user experience. Such frequent path changes are not considered in our basic use-case, which is tailored towards relatively (but not completely) stable underlying infrastructure.

During special conditions, such as when adding a new node or site to an existing network, the load caused by signaling can be considerable as initialization messaging occurs. However, such conditions should be considered rare events. In contrast, the scalability during nominal, “steady-state”, operating conditions is of paramount importance. The

following considerations are focusing specifically on the signaling during nominal operation.

Any signaling scalability issues may have an effect on end-user experience when re-establishing communication paths during outages. If re-establishment is not achieved quickly enough, the end-user traffic could be transmitted via a non-optimal path, or even worse, dropped altogether. The issues can be mitigated somewhat by establishing multiple paths concurrently. As long as these alternate paths exist, the traffic can immediately be diverted to them without waiting for the signaling process to be completed. In such a model, the signaling to re-establish broken down paths can be conducted in a more leisurely fashion in the background.

The signaling of Mobile IP is based on a simple request-response scheme. A Mobile Router generates a registration request message which is sent either to the Home Agent or another Mobile Router. The recipient, the HA or an MR, then processes the message and responds with registration response message.

As the number of customer sites, and therefore Mobile Routers, grow, the amount of signaling traffic increases. In the case of MIP, there are two categories of signaling messages being used: path monitoring and path maintenance. Path monitoring consists solely of keepalive messages, that are implemented as ICMP echo requests and their respective replies, and are sent when there is no end-user traffic. If no echo replies are received, an outage is detected and recovery actions conducted with path maintenance messages are triggered. Path maintenance messages consist of all other signaling apart from monitoring.

The load caused by an increase in signaling traffic is affected by three factors:

- Signaling message sizes
- Message bandwidth consumption
- Message processing delays

The maximum message size is the maximum size of an UDP datagram, or 65535 bytes, although sending such large messages results in fragmentation. To avoid fragmentation, it would be best to fit the message within a single IP packet. A single registration reply message could grow quite large as the number of Mobile Routers in the network

increase, and the entire MIP-based RAIIC scheme is relying on the Home Agent informing Mobile Routers of their peer Mobile Routers and their networks.

To mitigate the message size growth issue, the HAaRO specification (Publication II) includes a compression algorithm for IP prefix information, with low memory and computation footprints as design goals. The compression algorithm is tailored for the specific use-case of RAIIC, where all different sites belong to the same entity and thus IP addressing of each site is most likely to be continuous. With the awareness of the semi-continuous nature of the IP prefix information, a compact compression algorithm can be designed. Using such specific algorithms avoids resource consumption issues of more generic algorithms. An example of the compression operation is shown in Table 4.1, where information on four networks relatively close to each other is reduced from 20 octets to 10. The first prefix is stored as-is apart from omitting the least significant bits, consisting of zeros. For subsequent prefixes, only a single octet and the prefix length are required. In typical case with larger data sets, the algorithm succeeds in reducing the amount of data, on average, to less than one quarter of the original.

Besides compression of the prefix information, the specification also includes an algorithm for compressing optional realm information, which is a derivative of the basic domain name compression algorithm specified in RFC 1035 [40], which is based on filling a *dictionary* with *labels*: if similar structures exist in multiple realm names, a much shorter reference can be used instead. Inclusion of the realm information in messages is optional, but allows for differentiating between networks belonging to different entities, and are typically in the form of domain names, such as *company.com*. The full efficiency studies for both algorithms are in Publication III, and additional work, such as extension to IPv6 prefixes, are in Publication IV.

Another factor when number of paths and sites grow, is that the number of messages (of all sizes) in the network grow as well. With small number of messages that are generated in small networks, establishing and maintaining even a full-mesh network does not overtly consume bandwidth. However, as the size of the network grows, there are still several mitigating factors to limit the amount of messages as part of the protocol design:

Table 4.1. Examples of prefix compression with four near-consecutive IP subnets.

Subnet	Compressed data	Size, in bytes	
		Uncompressed	Compressed
192.168.1.0/24	“192,168,1,24”	5	4
192.168.3.0/25	“6,25”	5	2
192.168.3.128/25	“7,25”	5	2
192.168.9.0/24	“9,24”	5	2
Total		20	10

- Path monitoring messages for a path are only sent when no end-user data is being transmitted. Thus, the monitoring messages do not consume bandwidth or processing power if end-users are actively utilizing the path

- Path maintenance messages are only sent when topology changes

- The inter-site connectivity may be performed on-demand only

Thus, the primary factor that most affects the amount of signaling is the stability of the underlying infrastructure, as infrastructure state changes need to be communicated. The stability of underlying infrastructure in the intended use-case of RAIIC is still supposed to be relatively good, meaning relatively few changes. Combined with using broadband connections where at least some statistical information of the overall reliability is available (see Section 2.4), certain measures to limit signaling, such as avoiding path flapping, can also be taken. As a result, signaling due to state changes should occur relatively rarely.

In addition, the specification includes provisions for on-demand path establishment. For example, if most of the communications from small offices is to and from a few larger, central sites, and not directly between the smaller sites, partial mesh connectivity is enough for most purposes. To complement partial mesh, it is possible to establish direct paths on-demand as traffic warrants. This provision is also present to prevent a cascading failure in a case where a central site needs to re-establish paths after a breakdown, as not all paths need to be re-established at once. Thresholds and guidelines for when and how to establish the mesh are not specified and are considered to be part of the implementation design. For example, a similar system to PIM-SM [13] could be used, where a

configurable threshold of data can trigger switching from shared tree to shortest-path tree. Similarly, a Mobile Router could choose to initiate direct connection once data from a peer network passes certain threshold, and drop direct connectivity to peers without traffic.

Ultimately, despite all the compression methods and techniques to reduce signaling traffic, a practical upper limit will be reached to the number of concurrent tunnels. While path maintenance messages, regarding handovers and re-registrations, can appear at high rate during system initialization, they are relatively rare during steady-state conditions. As such, the monitoring traffic causes the bulk of signaling load during normal operation. The monitoring traffic consists of keepalive messages, which are ICMP echo requests and their respective replies. As such, the load caused by path monitoring can be calculated relatively easily, as follows:

- The HAaRO specification suggests a maximum of five keepalive messages per second (although this limit was surpassed for testing purposes, see Section 4.5)

- While the length of the ICMP messages are not specified since implementations can vary, the ICMP packet size is typically 64 bytes

- Counting the encapsulation into IP headers, that add additional 20 bytes, each message is thus 84 bytes in size

- This results in $84 \times 5 = 420$ bytes per second

In addition, the link layer headers consume bandwidth as well. Slight variations in true load may manifest due to such issues as buffering, Ethernet preambles or ADSL interleaving. However, in a typical case, the total bidirectional load can be approximated to roughly 5 kbps for each maintained tunnel. Since the tunneling peers, apart from the Home Agent, are also conducting similar monitoring processes, the load for each tunnel is doubled to approximately 10 kbps. Depending on the amount of bandwidth that is allocated for signaling, the number of direct connections may vary. If a maximum of 10% of available bandwidth is allocated for path maintenance, what follows is that a single 1 Mbps link could simultaneously maintain ten tunnels. The value could be slightly

higher, as end-user traffic causes the keepalive messaging to cease. The bandwidth increase is completely linear, as shown in Figure 4.2.

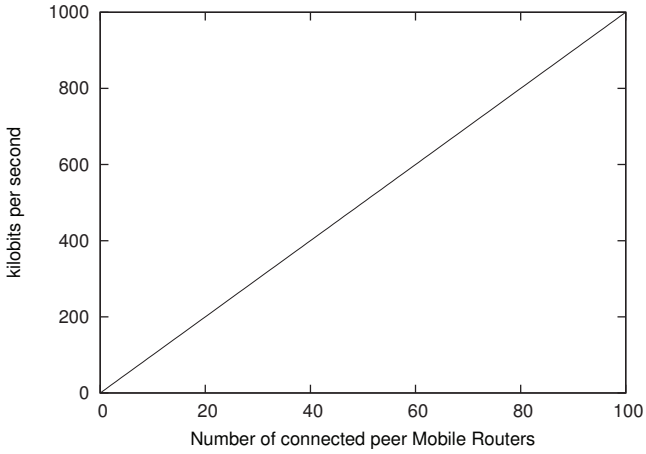

Figure 4.2. The bandwidth consumed by path monitoring messages when the number of concurrent connections increases.

The possibility of ten practical direct connections in case of 1 Mbps link, or more general case, $bandwidth\,in\,kbps/100$ connections is reasonably satisfactory. This guideline can also be considered a threshold when to start establishing connectivity with on-demand basis instead of a full mesh. In a typical case, the services provided by the organizations network are typically concentrated in relatively few locations. With such structure, a hybrid model is feasible, where direct connectivity for these sites is always preserved, but additional connections are established and torn down as events warrant.

As messages sizes grow, processing delays grow as well, as each node needs to parse through larger messages. To address large messages, the specification explicitly states that a node may stop processing HAaRO-specific fields in at any time due to resource or other constraints, allowing for a possibility to gracefully limit the processing load. In our research with a very basic implementation that was not optimized for speed, the effect on message size on processing delay was negligible. The bulk of the total delay was caused by transfer of the messages themselves and the processing delay itself was relatively constant regardless of message size. The longest processing delay caused by a single message was the very first, initial registration to the Home Agent by a Mobile Router, which lasted 60 ms when using the Alix platform. All the other maintenance messages, such as processing of a registration request from

a peer Mobile Router, were considerably faster, with processing delays of under 10 ms. This was due to the Mobile Routers conducing most of the heavy operations, such as allocating information structures for peer networks, during the initial registration procedure to the Home Agent. Thus, the processing delay of messages to and from HA is bound to grow as the size of network increases due to more data being processed. However, the inter-Mobile Router registration processing delay should stay constant.

4.4 Effects on different types of applications

As previously stated in Section 4.2, the expected performance requirements of the network vary depending on the application. The most interesting cases are behavior of interactive and real time applications when a short outage occurs in the network; we did not study bulk transfers at all as the effect of short outages on them is negligible. Additionally, make-before-break cases where the Mobile Router has an advance warning of impending path failure and can pre-emptively conduct a handover to operational path are always transparent to the end-user.

The perceived degradation of experience for the end-user can vary greatly when network is not operating correctly. RAIIC attempts to avert such degradation, to a varying degree of success between applications. In the case of interactive applications, such as web browsing, the aim is to conduct all operations pertaining to RAIIC infrastructure completely transparently. Such transparency would mean that recovery from network outage should be fast enough to be completely unnoticeable. Conversely, in the case of real time applications, the most prominent example being VoIP, any problems in the infrastructure affect the end-user. However, the effect should still be as small as possible.

Most of the research about application performance are related to web browsing and VoIP, and those two applications are focus of the RAIIC-related research as well.

On a network operations level, VoIP can be considered the simpler of the two applications. A VoIP conversation is usually a constant-bit-rate stream of RTP [60] packets over UDP. Such VoIP implementation has no immediate flow control, although a feedback mechanism may exist, for example in the form of RTSP [61]. The application may implement a small

buffer to prevent jitter (variation of delay) affecting quality, although this is not common as it also increases the delay. Other error correction mechanisms for lost packets may be present in the used codec, but packets are not retransmitted. One of the most common codecs, G.711 [27], which is also used in PSTN, does not include any such functionality; any dropped packets are immediately heard as periods of silence. A single packet containing G.711 data typically contains 10 milliseconds of speech.

The other interesting common application, web browsing, is a more complex entity, although for evaluation considerations the mechanisms can be considered simpler. From network evaluation perspective, user browsing web establishes and concludes a number of TCP sessions of varying sizes. Although the TCP connections can be directed to multitude of servers and can be used for retrieving vastly different kinds of data, such as text, images and even video, from RAIIC perspective the contents are not of concern as such. A specific TCP flow works similarly to other TCP flows regardless of the payload. In some cases such as video streaming the effect of outages in the network can be completely transparent to end user, due to buffering hiding short breakdowns.

Based on this information, the effects that primarily concerns end-user experience are the reactions to outages on a transport level. The two transport level issues are how do flow-control mechanisms in TCP react to short connectivity breakdowns, and what is effect of missing data from UDP to VoIP. Besides outages, secondary interest is for situations where the connectivity is operational, but degraded to the point where it affects usability due to such issues as congested network core.

There is a multitude of variants of TCP. However, most of the differences between TCP variants stem from details in calculating transmission window and moving from slow-start to congestion avoidance. On the other hand, the mechanisms for responding to outages or connection breakdowns have been relatively universal: The original retransmission timeout and slow start presented in original TCP specification [52] and Fast recovery/Fast Retransmit mechanisms specified in [2]. Overall, typical viewpoint for almost all TCP variants is that when packets are lost, this should be interpreted as congestion, not as an outage, and transmission speed should be throttled. Exceptions are rare but exist: For example, AR-TCP [53], designed for ad-hoc wireless networks, attempts to differentiate between outage and congestion by observing the rate of throughput changes. If throughout drops faster than a certain threshold,

the cause is interpreted as an outage and RTO timer is not increased exponentially, as it would be with a normal revertion to slow start.

For ordinary TCP, the throughput recovery speeds after an outage are related to the measured round-trip-time (RTT) between the connecting hosts. Figure 4.3 shows the simulated throughput of TCP variant CUBIC, used by default in Linux operating system, when a breakdown for a session with 100 ms RTT occurs at *t=5 s* and the connectivity is restored $N \times RTT$ later. For very short breakdowns (1 or $2 \times RTT$), the effect is negligible, as fast recovery takes care of almost immediate restoration of throughput. For longer breakdowns ($6 \times RTT$ and longer) the sender enters slow-start phase and starts exponential RTO back-off. With slow-start process, the throughput is relatively quickly recovered after connectivity is restored, although the resumption is far from immediate. With breakdowns lasting from 6 to 14 times RTT, the RTO has backed off to roughly 1500 ms (100+200+400+800 ms), and throughput starts to recover at *t=6.5 s*. Such a 1.5 second pause can be considered tolerable for web browsing as it is below the threshold of 2 seconds. Starting with $16 \times RTT$, the RTO growth starts to have significant effect, causing the throughput recovery point to jump to *t=8.5 s* meaning effectively a four second stop in transmission even though throughput could have been recovered much earlier. Yet another RTO expiration would lead to recovery at *t=12 s* (not included in figures).

The figure omits the intermediate value of $4 \times RTT$. During the simulation, the recovery process of TCP failed to work as expected. Data transfer was resumed, but throughput was severely hampered for several seconds. This unexpected observation was later concluded to be invalid, because the behavior not repeated during implementation work. The error was traced to handling of advertised TCP window size. In the simulator, the advertised window of TCP is grown rapidly if the receive buffer is emptied immediately after packet has been received, and huge advertised windows create problems for Fast Recovery mechanism. The behavior did not manifest itself during real-world experimentations, as although the receive buffer is read after packet reception, the operation is not, strictly speaking, immediate. Thus the advertised window size was not increased.

Without external triggering mechanisms, the outage detection mechanisms have a lower limit on how fast the detection and recovery can occur. Apart from the trivial case where link-layer indication on

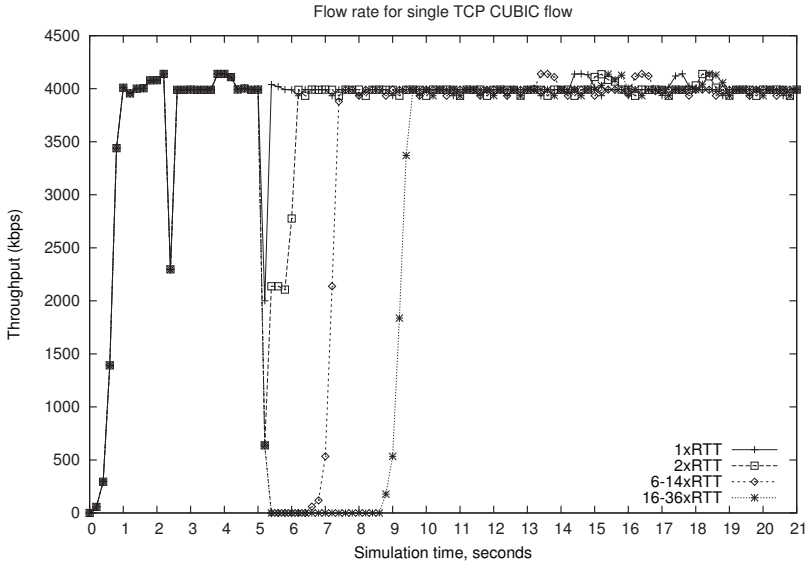

Figure 4.3. Behavior of TCP Cubic with varying length breakdowns.

breakdown can be used for immediate detection, Mobile IP sends periodic keepalive messages if no user traffic has been transmitted for a while. To avoid overloading network with keepalive messages, the specification in Publication II suggest a maximum of five messages per second, and three failed consecutive keepalives should be interpreted as an outage. Although the keepalive message frequency could be slightly increased (see next section), the detection speed is still well within the range of where TCP reverts to slow-start, instead of utilizing fast recovery, unless the delay in the network is considerably high. High delays in this case means in the order of several hundreds of milliseconds, which might occur with technologies such as satellite links. Once in the territory of slow-start, it would be beneficial to recover connectivity as fast as possible, before the RTO timers of TCP sessions start their growth. Since an RTO timer is increased in an exponential fashion, the TCP sessions might not recover their throughput until well after the connectivity has been restored if the timers grows too large.

4.5 Measured end-user metrics

As stated in previous section, when an outage occurs in a break-before-make fashion, the outage has to be detected before recovery processes can be initiated. Once detected, the Mobile Router starts the recovery

processes, and the outage will block end-user traffic until recovery has succeeded. To measure effects of outages and subsequent recovery on end-user applications, an implementation of the Mobile IP-based RAIIC was created (see Section 4.1), consisting of Mobile Router and Home Agent components. End-users were emulated with the Spirent Testcenter, and breakdowns artificially generated with the emulated infrastructure.

In this case, since the interest was solely in the effect of outages, the only emulated behavior introduced into the network infrastructure was the link delay. The link speeds were limited on the link layer to the 10 Mbps Ethernet speeds, which can be comparable to a fast home broadband connection. Overall, the experimentation scenario consisted of having a number of users located in network A accessing a server in network B. The server access would consist of either downloading a web page or conducting VoIP conversation. During the experiment, the Testcenter equipment recorded perceived throughput and VoIP quality metrics of end-users.

As an example, what occurred in the system during one of the generated outages is shown as a timeline in Figure 4.4. At the starting point of the timeline, Mobile Router A has established a path (tunnel) to both the Home Agent and MR B via same network element. The network element fails at t=*116 ms*. At the time of failure, multiple TCP sessions are in established state from the network behind MR A to the network behind MR B, as users are downloading the web pages.

The diagram shows how the outage detection and recovery progresses: after the last TCP packet has been received from network B, three keepalive messages are sent. The keepalive messaging has been constantly active on the path to Home Agent, as it has had no end-user traffic at all. On the other hand, the messages towards MR B start only after the end-user traffic has ceased. After MR A determines that the keepalive messages will not receive a response, alternate paths to the destinations are formed by conducting registration messaging. The registration messages were sent immediately after outage had been detected, unlike presented in Figure 3.4 which shows the Return Routability messages. The RR messaging was not conducted, as it is assumed that in real-world implementations such key exchanges are conducted pre-emptively and therefore authentication keys already exist for such sudden switchovers.

First end-user TCP flow recovers at *t=1069 ms*, so at least one end-user felt the effects of the outage for less than a second. Such a short pause can be considered acceptable for web browsing. However, it should be noted that the time stamp is the time of *first* TCP packet after the outage; as stated previously, TCP slow start process requires more time to increase throughput back to acceptable levels. However, as could be seen in Figure 4.3, the throughput recovery, even from slow-start, is relatively fast. Additional issue, synchronization, where multiple TCP flows are in same state and start up concurrently could be avoided with for example RED algorithm [14] or one of its more advanced variants. Note that in this case, in the 188 ms period between *t=748 ms* and *t=936 ms*, when connectivity via the HA is available but the direct connection between MR A and MR B is missing, no TCP packets were transmitted via the HA, as all the TCP flows were in similar RTO backoff states and did not send any data.

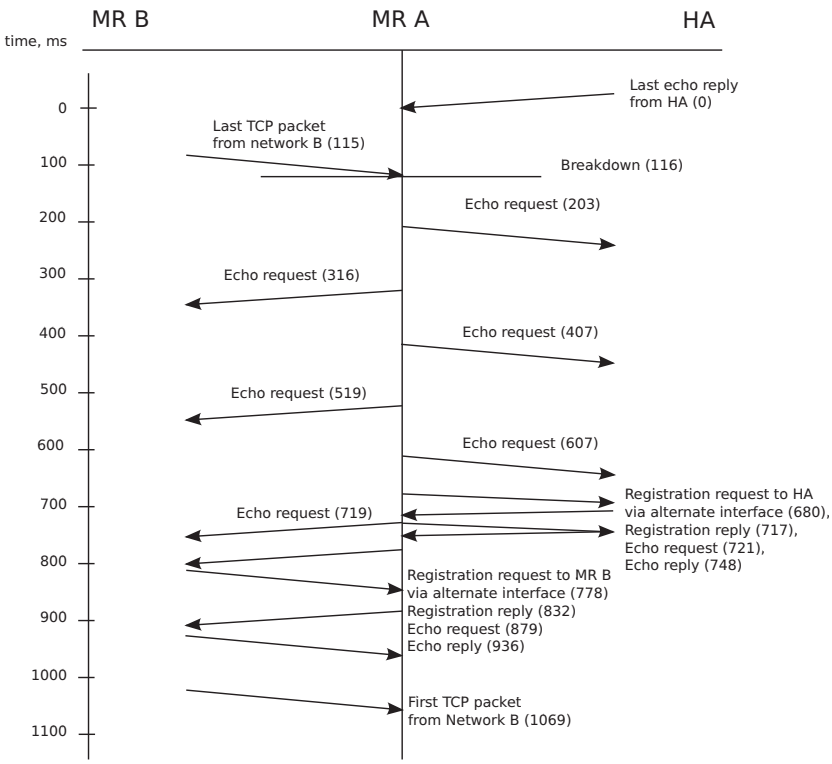

Figure 4.4. Signaling diagram illustrating node behavior during one of the generated outages.

VoIP applications do not typically implement any sort of flow control at all for the RTP session containing the actual voice data. The lack of flow

control causes the sound to continue the moment connectivity has been re-established and packets can be forwarded again. Such characteristics cause the effect on perceived voice quality to be directly proportional to the length of the outage. The observed voice call quality in terms of MOS (PESQ) with varying keepalive interval times are shown in Table 4.2. The table shows the values when calculated over the 3-second or 10-second periods of time where the breakdown occurred. The range is from 5 (Excellent) to 1 (Bad).

The results were obtained by having an emulated user in network A call another in network B. Each conversation lasted 10-30 seconds, consisting of playbacks of male voices reading a passage from Mark Twain's *A Connecticut Yankee in King Arthur's Court* [68] to each other. In our experiments, the paths were dedicated to VoIP traffic. Conversely, in a deployment where other, less real time-dependent applications share the same path with VoIP, voice traffic should be prioritized using methods such as priority queueing. The perceived quality of the calls were equal at both endpoints in all cases.

Table 4.2. The effect of different keepalive message intervals on voice quality during connectivity recovery. "No outage" is the reference case where the connectivity is stable.

Test case	MOS-CQ, over 3 s	MOS-CQ, 10 s
No outage	4.18	4.18
Make-before-break	4.18	4.18
150 ms interval	3.09	3.85
200 ms interval	2.61	3.71

As can be seen, decreasing the keepalive interval allows for faster detection of an outage and recovery procedures. However, even with the default keepalive interval of 200 ms, the opinion score is still at tolerable level, although barely, if the scores are calculated over 3 seconds. However, with the more typical 8-12 second period, the score is much higher, and the conversation quality can be considered good, which is usually interpreted as perceptible but not annoying. One of the conclusions that can be made is that a better metric for evaluating user experience for transient failures should be designed, as the true effect of a single, isolated 600-800 ms pauses to the quality perception in conversation is not readily apparent.

4.6 Utilization and fairness of available resources

As previously mentioned, in RAIIC approach each site is connected via multiple access links. These access links are connectivity resources, which the customer obviously hopes to utilize to the fullest. When striving towards maximum utilization, Mobile IP has three different distinct modes of operation, each requiring more advanced decision-making than the previous one. The most rudimentary of all is the scenario where only single link is used at a time. In such scenario, the only decision the Mobile Router needs to make is choosing the link that gives highest performance. A slightly more advanced version uses multiple links concurrently for per-site load balancing. In per-site load balancing different links can be used to connect to different sites, although all traffic between two specific sites still take a single path. The most advanced option is full-fledged load balancing that uses multiple paths between each site. The different options are shown in Figures 4.5, 4.6 and 4.8. The ISP peering points are not shown in the figures.

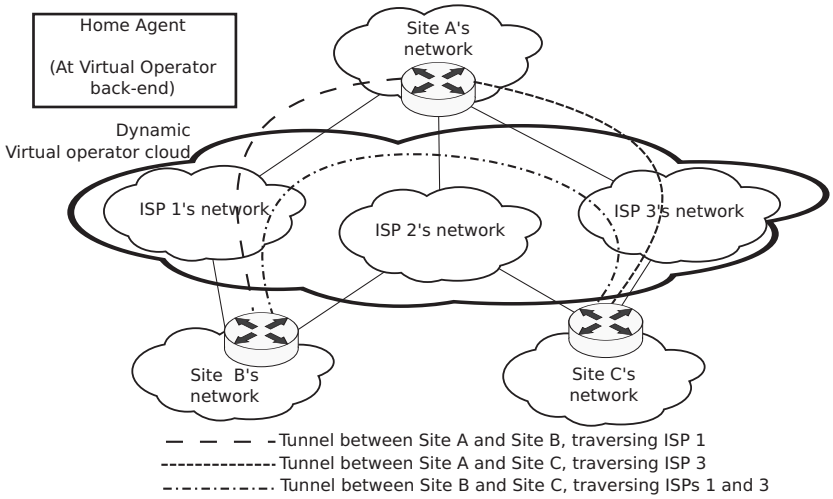

Figure 4.5. No load balancing: Utilizing a single link at each Mobile Router.

4.6.1 Per-site load balancing

The per-site load balancing approach requires no additional specifications for MIP. In per-site load balancing, the Mobile Routers can pick an underutilized link and use it to establish a tunnel to a peer site. From the perspective of a Mobile Router the new tunnel is just an additional, logical interface, and forwarding can be set up with routing tables as desired.

From perspective of applications, there is no difference to the single-path mode. Similarly to the single path mode, the Mobile Router forwards the packets according to the routing table, although routes to different sites may have different paths. For choosing the appropriate path to different sites, different strategies can be used in attempt to maximize both fairness and throughput.

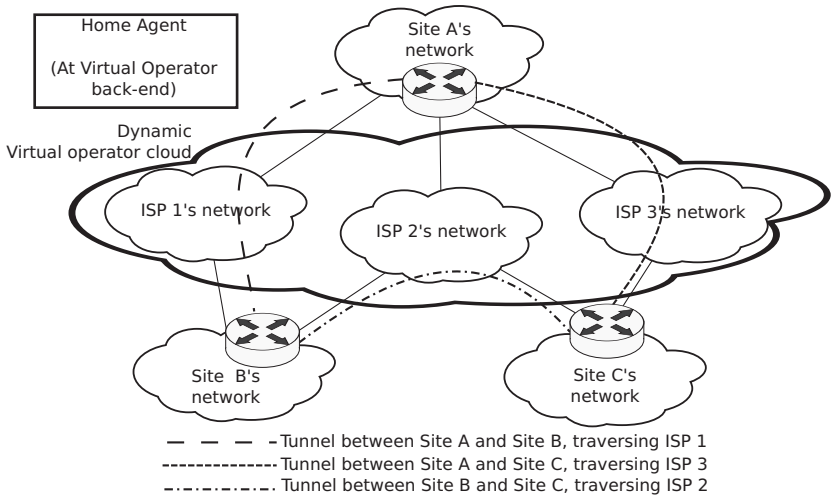

Figure 4.6. Per-site load balancing: Utilizing a single, separate path for each peer Mobile Router.

During the simulation-based study, a simple strategy for the path selection was implemented at each Mobile Router. The MRs attempted to equalize each individual available bandwidth of each flow between different paths. Thus, if a specific peer site had more active flows going to and from it than other peer sites, the path would be allocated to the higher-bandwidth link. A threshold exists to prevent tunnels from flapping back and forth between two equivalent paths.

The effects of this path allocation on per-flow throughput for an example scenario can be seen in Figure 4.7. The network has been set up as previously stated, with all three sites operational. Site A has been set up with 7 Mbps, 4 Mbps and 2 Mbps symmetric links, and thus the 7 Mbps and 4 Mbps links are given precedence over the 2 Mbps link. In Figure 4.7, 20 TCP flows are started towards from site A to site B at $t=0\,s$. At $t=2\,s$, 20 additional TCP flows are started towards site C. When the additional flows are started, the per-flow throughput momentarily drops, and then the MR detects that the new flows are towards a different site, and sets up a new direct tunnel to site C via the 4 Mbps link. Although the individual flows towards site B have higher per-flow bandwidth than

the flows towards site C, the overall throughput is significantly better. The individual flow bandwidths are now 350 kbps (20 flows through the 7 Mbps link) for site B and 200 kbps (20 flows through the 4 Mbps link) for site C.

At *t=10 s*, additional 10 flows are started towards site C. At this point, the MR detects that tunnel towards site C has more flows on it than before, and deduces that the flows would receive better per-flow bandwidth even if they took the same path as flows toward site B, since all 50 flows at 7 Mbps capacity (140 kbps per flow) is higher than 30 flows at 4 Mbps capacity (133 kbps per flow). After this switch has been completed, the same logic follows that tunnel towards site B can now be switched to 4 Mbps link, as it benefits everyone: 30 flows on 7 Mbps link equals 233 kbps per flow, and 20 flows on 4 Mbps link equals 200 kbps per flow.

The full algorithm is discussed in detail in Publication V. The work shows that TCP is demonstrably quite resilient to changes in underlying networking conditions, including sudden bandwidth changes due to path switching. The algorithm itself has safeguards for some basic issues, for example to prevent flapping between two paths if the per-flow bandwidths on different paths are equal or near equal. The algorithm was further developed for full-fledged load balancing (see next Section).

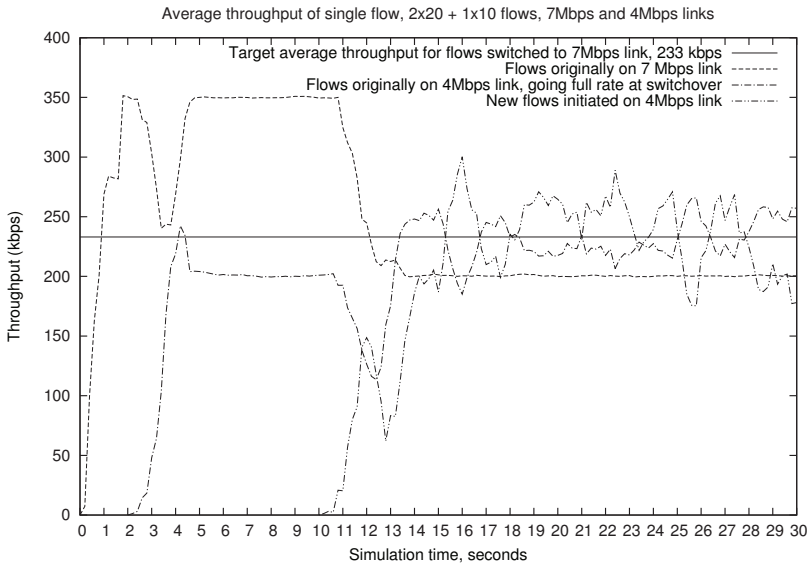

Figure 4.7. Per-site throughput when number of flows change.

4.6.2 Full-fledged load balancing

MIP can be extended [18] to a full-fledged load balancing scenario, where multiple paths exist and may be utilized between two Mobile Routers. To facilitate such utilization, the Mobile Routers need to implement a packet scheduler that decides, based on some metric, which of the several available paths each packet takes. The additional advantage of multiple concurrent paths is faster changing of path in case of an outage, even if no load balancing takes place, as the path set-up with registration requests has already taken place. Furthermore, an intelligent packet scheduler can also better react to less readily apparent failure conditions besides outages, such as a reduction of available bandwidth.

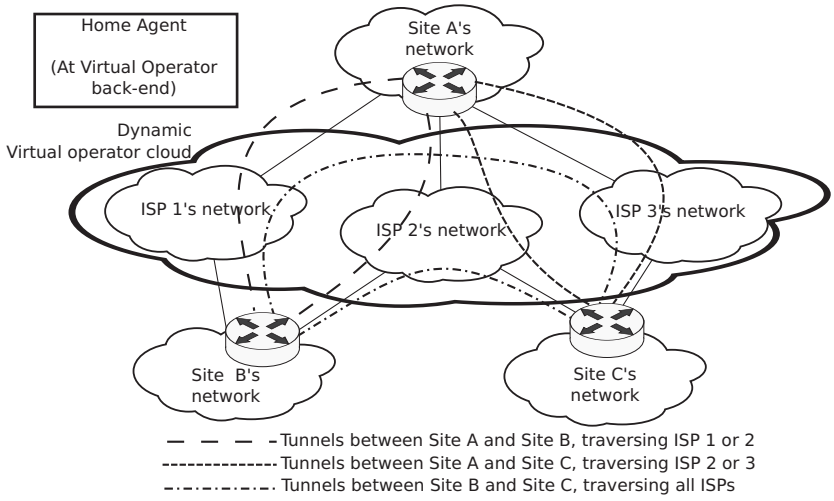

Figure 4.8. Full-fledged load-balancing: multiple, concurrent paths between Mobile Routers.

There are two issues that affect such load balancing schemes. First issue is the path formation, or how to set up the paths (in case of MIP, tunnels) most efficiently to spare resources yet allow for maximum flexibility. Second issue is the scheduling algorithm itself once the paths have been set up and are operational.

4.6.2.1 Path selection

When there are multiple interfaces on each site, multiple potential site-to-site paths exist. In the experimentation scenario, with three sites each having three interfaces with different bandwidths, a full-mesh pathing set-up would already create 27 paths. However, in this case, the optimal topology requires significantly fewer paths: interfaces of each router are

simply connected with interfaces of equal bandwidth on the peer routers, as shown in Figure 4.9. In the example, only three paths are needed between each site, for a nine paths total. Thus, when forming paths, some kind of algorithm is needed make use of all available bandwidth yet use the minimal number of paths to conserve resources and allow for a simpler load balancing logic.

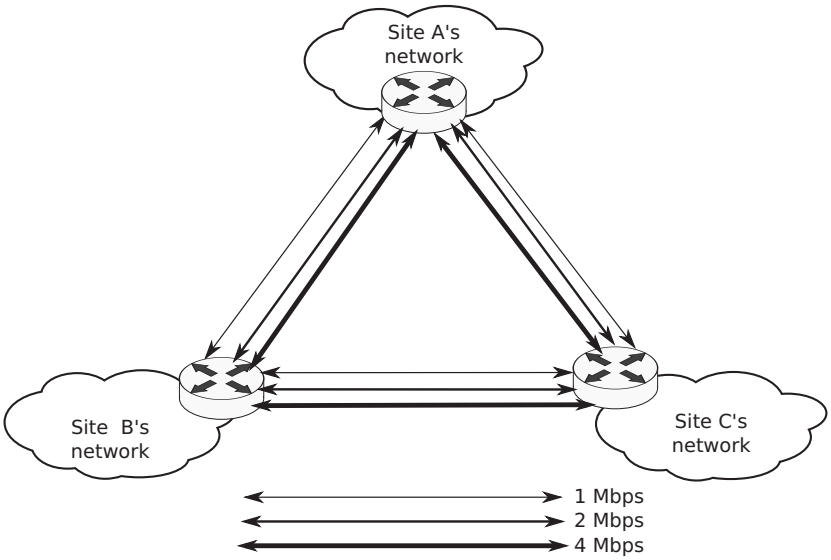

Figure 4.9. Optimal paths between three sites in a simple scenario. Each interface is connected to an equivalent-bandwidth interface on the peer router.

There are several possible approaches in valuation of different characteristics of a potential path. For example, the proprietary EIGRP routing protocol from Cisco [9] uses attributes such as bandwidth, delay, current load, and historical reliability as basis for metric calculations. For experimentation purposes, a relatively simple path forming algorithm was designed and implemented. The algorithm is using only bandwidth as a metric, although other characteristics could be incorporated as well. Furthermore, some sort of threshold could be used to disregard certain paths such as in a case where one of the paths is significantly slower or faster than the others. Conducting load balancing between a high-speed broadband connection and another connection that is an order of magnitude slower would provide little benefit.

The algorithm attempts to facilitate maximum bandwidth, and sets up paths in a fashion that employs as many network interfaces as possible but as few paths as possible. For input data, the upload and download capacities for every interface and for every participating node

is assumed to be well-known. This capacity awareness should not be a problem, as an administrator for any network should know the link bandwidths of their own network. Furthermore, this information could be distributed by Mobile IP signaling as well. Note that this information is only used for path forming, and should therefore be based on the assumption that the connectivity typically behaves close to advertised. If bandwidth slowdowns or similar issues occur during operation, the scheduling algorithm should attempt to react accordingly (see Section 4.6.2.2, below). If, at some point during operation, the routers determine that the path characteristics have changed significantly, the path forming algorithm could be executed again with the new values. However, the path re-forming should be considered a relatively heavy operation, and as such, should be avoided if the scheduling algorithm can correct for minor changes in performance.

The algorithm traverses every node pair in order, starting from the interfaces with highest available bandwidths. Then, the algorithm attempts to form paths from available interfaces to each other until no more upload bandwidth at the source or download bandwidth at the destination is available. As an additional constraint, when an additional path is formed, the algorithm prefers to use a network interface without assigned paths for added redundancy. Without the redundancy requirement, the algorithm would be Pareto efficient, but the redundancy requirement reduces efficiency somewhat, as in the RAIIC use-case reliability is more important factor. More details on the algorithm can be found in Publication VIII, including optimality analysis.

4.6.2.2 Load-balancing approach comparison

Once paths have been established, the scheduling between paths still has to be conducted. There are two primary targets to optimize for: bandwidth utilization and flow fairness.

Bandwidth utilization means that when conducting bandwidth-limited operations (such as file transfers), the utilization (load) on all paths should be at or near 100%. In essence, the customer is able to use all the resources that have been paid for. However, fairness should be assured as well, as without fairness considerations, individual flows might get unfairly large allocations of the bandwidth, while other flows would starve.

Measuring the utilization effectiveness is simple: calculate the combined available bandwidth of all paths together, and measure how much of this combined bandwidth is actually used. On the other hand, for evaluating fairness, Jain's fairness index [30] is considered an objective measure of fairness. The fairness index is defined as

$$f(x_1, x_2, x_3, ..., x_n) = \frac{(\sum_{i=1}^{n} x_i)^2}{n \sum_{i=1}^{n} x_i^2}$$

where n is the total number of flows, and x_i is the throughput of each individual flow. The index ranges from $\frac{1}{n}$ to 1. When the index is 1, each flow gets identical bandwidth (best case). In the worst case of $\frac{1}{n}$, a single flow is receiving all the bandwidth, starving all the other flows. Index values below 1 mean sub-optimal fairness.

Although there are a multitude of algorithms for scheduling packets for load balancing between interfaces, they can be broadly divided into two broad categories: packet-based and flow-based schemes. In packet-based schemes the *contents* of each packet do not affect the path selection decision at all; the packet is routed based solely on meta information, such as arrival time and packet size. In flow-based schemes the *contents* of the packet, typically at least headers, are taken into account. Packet-based schemes attempt to maximize throughput by scheduling packets at maximum rate to each possible path, while the flow-based schemes attempt to prevent issues such as out-of-order packet arrival and variable delays between paths by keeping packets belonging to a single flow on same path.

We have chosen three packet scheduling approaches/algorithms for comparison:

- Weighted round-robin (WRR) is a packet-based approach where packets are distributed according to upload bandwidths of each interface. Our implementation of the approach counts octets instead of packets, and the packet sizes do not matter. Consequently, the actualized weights may vary slightly since the smallest unit of data that can be sent is a single packet.

- Weighted 5-tuple-hash (HF) bases the path selection decision on the contents of the packet. The 5 attributes of an IP packet are combined with a hashing algorithm to a number between zero and the combined weight of all possible paths. The path is then chosen depending on the

weight of each individual path. This ensures that packets belonging to a specific flow will always traverse the same path.

- Intelligent flow distribution (IF), unlike the two previous ones, maintains flow state information in a table, and is based on the simple algorithm used in simulation work. The primary difference compared to the version in the simulation is differentiation between different kinds of flow behaviors. When a new flow is detected, it is allocated a path from available possibilities. The path selection process is relatively simple, using the available data on network state, and details are presented in Publication VIII. However, the key difference to the previous algorithms is explicitly allocating a specific path for each flow and maintaining this information, which is then looked up for each packet to be forwarded. In certain conditions, the algorithm may also reallocate already established flows to different paths. The IF algorithm is included in comparison to observe whether an intelligent decision-making process results in significantly better bandwidth allocation.

The scenario for experiments was set up as in Figure 4.9. During the experiment execution, varying number of emulated HTTP clients located at site A requested web pages from servers at sites B and C. The experiment went through a series of phases, and at each phase, data was gathered for observing throughput and fairness. The throughputs are shown in Figure 4.10.

In the top diagram, 64 TCP flows are active at the same time. In the bottom diagram, a single, larger flow is repeated in succession throughout the experiment. The experiment was also conducted with 4 and 16 concurrent flows, however the difference in utilization is most profound when comparing a single flow to 64 flows. The various distinct phases, each lasting roughly 120 seconds, are clearly visible and covered in detail below.

- From $t=0\,s$ to $t=380\,s$: each individual link is operational one at a time, and throughput can be seen going from 4 to 2 and eventually to 1 Mbps.

- From $t=380\,s$ to $t=500\,s$: right after the single-path 1 Mbps phase, a full 7 Mbps (4+2+1) phase begins. With 64 flows, all algorithms allow for maximum bandwidth. In contrast, with a single flow, the packet-based

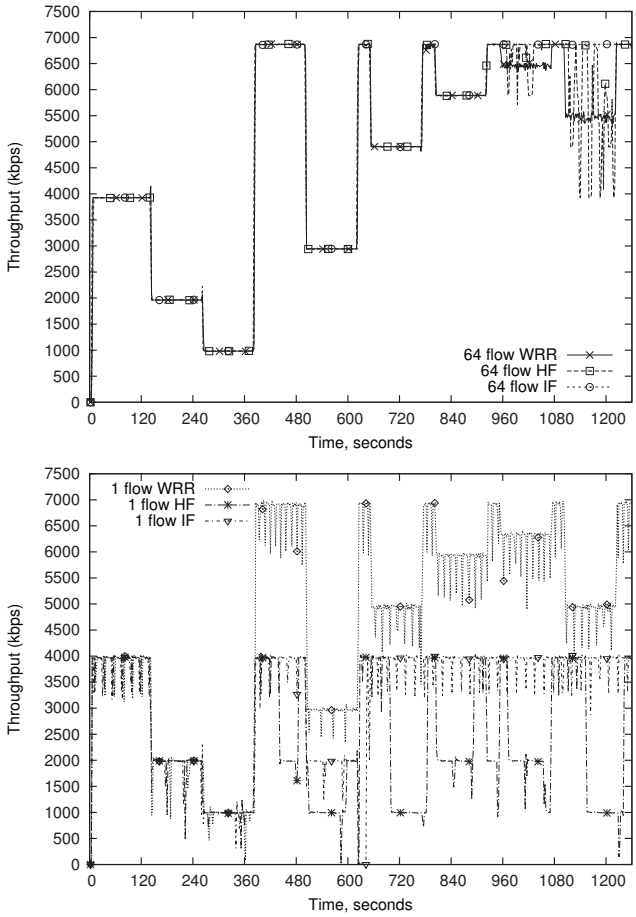

Figure 4.10. Cumulative throughput rates during the experiment: 64 concurrent flows above, single large flows below.

WRR is the only algorithm allowing for maximum utilization, as flow-based algorithms are limited to using a single path since a single flow cannot be split between several paths. The intelligent decision-making of the IF algorithm picks the fastest possible path, with HF the results are quite random and vary with each flow.

- From *t=500 s* to *t=960 s*: alterations on which exact paths are operational for 120 seconds, with short 30 second periods of full operation in between. The exact combinations can be deduced from the utilization graphs with 64 concurrent flows: 2+1, 4+1 and 4+2 Mbps.

- From *t=960 s* to *t=1080 s*: false weights-phase. At this point the algorithms are purposefully given wrong information: the 4 Mbps link

is represented as a 5 Mbps link by giving it a weight of 5. During this phase, the flow-based methods keep providing maximum throughput, and all paths are still utilized, although flows may be allocated in wrong proportions. With packet-based WRR the throughput suffers, since all packets are distributed amongst different paths, and congestion control mechanisms of TCP start detecting congestion and throttle accordingly. The 1 Mbps and 2 Mbps paths become underutilized, since the 4 Mbps path is full and has a configured 20% overweight due to the distribution rations being 5:2:1 instead of the correct 4:2:1. With their combined 3 Mbps of capacity working with 20% underutilization, the end result is throughput of 800+1600+4000 = 6400 kbps.

- From *t=1110 s* to *t=1230 s*: interference phase. In this phase, the traffic from sites B and C are interfering with each other. Different paths have an outage for each peer site: 2 Mbps path is down for site B (total capacity 5 Mbps), and 1 Mbps path is down for site C (total capacity 6 Mbps). The 4 Mbps paths remain connected to both sites. As such, this can also be considered a phase with false weight information, although a more dynamic one, since it depends on the traffic profile. It can be immediately noticed that there is no interference in single flow case, as the only flow is directed from site B to site A, and there is no interfering flow from site C.

The interference phase with multiple flows shows clear differences between the algorithms. Only the IF mechanism keeps providing maximum throughput during the entire interference phase. The WRR suffers from interference and combined throughput drops to about 5.5 Mbps from maximum of 7 Mbps. The drop in throughput is caused by the 4 Mbps path being used by both sites and thus the per-site capacity being roughly 2 Mbps, or half. Since the packet distribution is based completely on weights at the sender, the alternate paths become underutilized by roughly 50%, leading to a total throughput loss of 1.5 Mbps. Similarly, the HF algorithm which relies on hash suffers in performance as well, with the total throughput experiencing high fluctuations from maximum 7Mbps to under 4 Mbps. The fluctuations are caused by flows still being distributed with 4:1 or 4:2 ratio, even though the 4 Mbps path is shared and the second path is not. As such too few flows are sent via the secondary path and too many via the primary. The

more intelligent algorithm detects more accurately when the secondary path should be used and thus provides best performance.

The phases discussed above were studied for fairness as well. The Jain's fairness indexes are shown in Table 4.3. As can be seen, packet-based WRR appears to be the fairest algorithm in all cases, including failure conditions with false weight information and interference. Note that the scale from best to worst is not equal between columns; with 4 flows, the scale goes from 1/4 to 1, while with 64 the scale is from 1/64 to 1.

Table 4.3. The various traffic algorithm and test phase combinations, with respective fairness indexes.

Algo + test phase	4 flows	16 flows	64 flows
4Mbps, single path	0.954	0.893	0.945
7Mbps, WRR	0.969	0.984	0.994
7Mbps, IF	0.935	0.889	0.873
7Mbps, HF	0.835	0.652	0.661
False weights, WRR	0.979	0.990	0.964
False weights, IF	0.940	0.828	0.831
False weights, HF	0.858	0.640	0.585
Interference, WRR	0.978	0.971	0.934
Interference, IF	0.924	0.677	0.641
Interference, HF	0.835	0.742	0.450

Although the flow-based algorithms appear to be rather unfair, the fairness index does not reveal entire truth. The fairness index was originally intended for single-path scenarios. Since this is not a single-path scenario, a more detailed analysis of the fairness distribution was conducted. More illustrative diagrams can be found in Publication VIII, but a good example is shown in Figure 4.11. The diagram illustrates how many of the flows achieve a specific percentage of the average per-flow bandwidth (which is at 100% mark). On the left, a baseline with single path is shown and a spike can be seen near the center, as most flows get near-average bandwidth. On the right, we have the conditions from *t=1150* seconds in 64 flows case of Figure 4.10. As can be seen, even in the baseline the deviations from average are considerable: bulk of the flows achieve 70-140% of the average traffic, with a few outliers. In contrast, with load balancing active and based on wrong information, the deviations with flow-based algorithms are very high and can be observed to provide uneven fairness. However most of the flows are still

concentrated on a relatively narrow area at the lower end of the range. In this case, certain, specific flows are getting much higher bandwidth, but clearly visible concentrations of flows at lower throughputs exist. Although WRR is the fairest algorithm when measured with fairness index, it also distributes error conditions just as fairly and thus affects overall throughput. The flow-based algorithms provide better overall throughput, and most flows are still using nearly same amount of bandwidth, meaning that some fairness is still preserved.

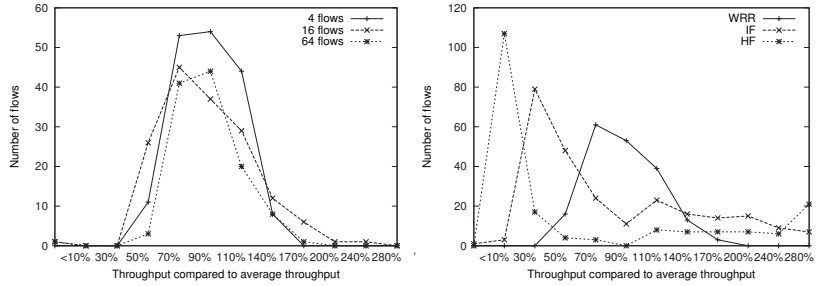

Figure 4.11. Throughput distribution with 64 flows, single path phase on the left, interference phase on the right.

What the behavior and fairness indexes mean in practice are more evident when the basis for the fairness index calculation is shown in more detail in Table 4.4, which shows the minimum, maximum, average and median throughputs per flow for the cases with 64 flows. In addition, the number of completed flows are shown. The values demonstrate the relationship of the fairness index to the spread in per-flow throughputs: Values greater than 0.9 are comparable to the baseline, and the above mentioned 70-140% of average per-flow speeds. When the index decreases below 0.7, the differences between flow start to become very discernible: The worst case of 0.450 (HF algorithm during interference phase) has the highest flow experiencing speeds of nearly ten times the average throughput. The practical effects of poor fairness depend on the application. When transferring a single, large file, consisting of a single flow, the performance variations could mean excellent or very poor performance depending on which path the flow traverses. With web and similar applications, which consist of several flows, the effects are less pronounced, as the differences between individual flows balance each other out somewhat.

In all cases where load balancing was used, there were roughly 200 completed flows, which can be considered an adequate sample size. As

can be seen, the average and median throughputs are always the highest with WRR when compared to flow-based algorithms. This occurs even in the basic load-balancing case with 7 Mbps of combined bandwidth, although the difference is small – it is not even visible on the throughput graphs. The most profound indicator of throughput is the number of completed flows (320 kilobyte TCP sessions). The intelligent, flow-based algorithm appears to provide best overall performance in all conditions and fairness is somewhat preserved: The difference between minimum and maximum throughput is higher than with WRR, but not overtly so. In the interference case, intelligent algorithm is only one which preserves total throughput with nearly 250 completed flows while the others manage to complete less than 200.

Table 4.4. Fairness index and throughputs in bytes per second with 64 flows.

Algo + Test phase	Flows	F-idx	Min	Avg	Median	Max
4Mbps, single path	117	0.945	4889	8108	7679	15413
7Mbps, WRR	227	0.994	10353	12673	12589	15650
7Mbps, IF	246	0.873	6911	14955	13510	44462
7Mbps, HF	233	0.661	5898	22829	14850	124994
False weights, WRR	229	0.964	7807	12804	12806	21800
False weights, IF	244	0.831	6462	15590	13748	46105
False weights, HF	233	0.585	4150	19339	14293	111465
Interference, WRR	185	0.934	6516	11464	10910	22516
Interference, IF	249	0.641	6408	21722	14115	81426
Interference, HF	198	0.450	4103	22183	12274	111794

Based on these measurements, it is possible to make certain recommendations. WRR provides best fairness in all kinds of situations. As such, in relatively stable network conditions where information on the situation is available and accurate, WRR provides both fairness and utilization. Conversely, if accurate information is not available or the information is outright wrong, WRR suffers in utilization. In such situations flow-based methods are more suitable, especially if more intelligent decision-making on the path allocation is performed.

Overall, it is quite feasible to actualize the benefits of all available paths constantly, not simply as backups for outage situations.

4.7 Comparing performance of RAIIC to traditional reliability approaches

As stated in the introduction, the work and our proposed RAIIC approach attempts to deliver equivalent or better performance compared to traditional SLA-guaranteed reliability approaches at significantly lower costs. The economical aspects have been presented in Chapter 2.

From a technical perspective, to compare performance of RAIIC to existing approaches, there are two distinct comparison points: The detection and recovery time compared to traditional high availability methods during an outage, and the actual total availability of service, meaning conformance to the agreed-upon SLA. The actual availability of the service has been covered previously; the whole premise of RAIIC rests on the assumption that some form of connectivity is always available. For more intricate discussion on this, see Chapter 5, especially Sections 5.4 and 5.5. The key points thus becomes the handling of outages and switching to redundant paths, as well as the effectiveness of load balancing.

The traditional reliability approach includes redundant equipment and links on the entire path between two customer sites. However, from the perspective of the customer, it does not matter which specific component fails if the path has an outage. The difference between the traditional approach and RAIIC is that the traditional approach consists of multiple redundant links, managed by a single service provider, that together form a reliable path through the network of the provider. Conversely, the RAIIC approach consists of multiple redundant paths that are established from site to site without considerations on what kind of network is used for transit. If a failure on the path occurs with RAIIC, the entire path is switched to a redundant one, while in traditional approaches only the specific component on the path is changed. These differences are illustrated in Figure 4.12.

With such considerations, the actual key difference between RAIIC and traditional approaches is the behavior at the customer edge, as the edge routers are the components that are shared by both approaches. Furthermore, while the traditional service provider may have a high level of redundancy at the core network, such as the full mesh shown in the Figure 4.12, the RAIIC concept completely ignores such issues and switches the entire path to a one traversing another provider. As such, we

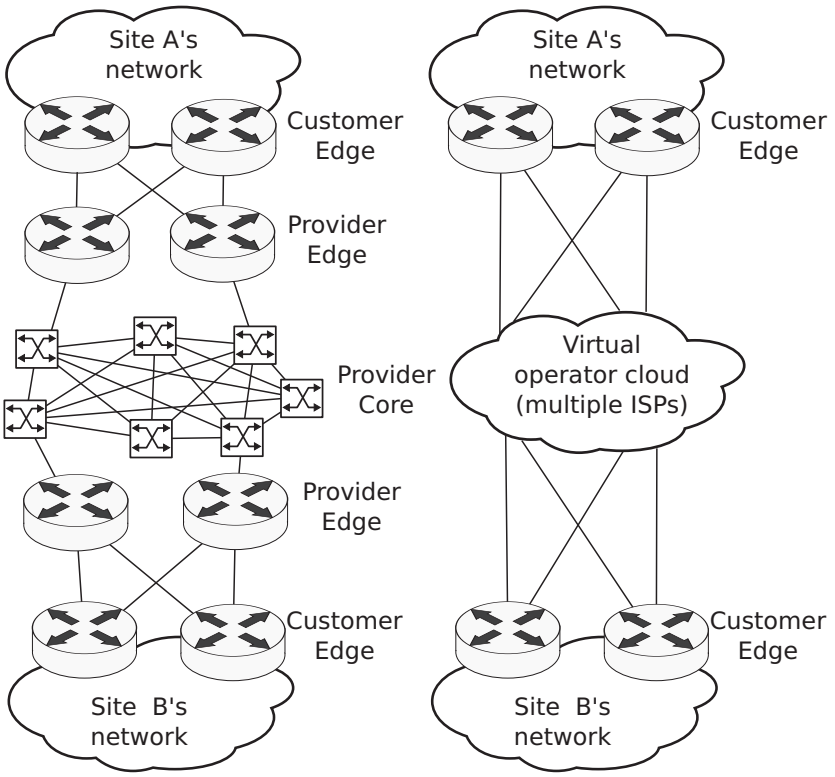

Figure 4.12. Comparison of traditional and RAIIC approaches in connecting two customer sites. The traditional approach (left) has a service operator provide redundancy throughout the path, while RAIIC approach (right) has multiple paths via different providers, and redundancy is implemented at the edges.

can focus the performance comparison to the edge routers. Furthermore, they are the only component that is directly visible to the customer, as the customer edge routers act as gateways towards the rest of the network. Ultimately, the end-to-end performance is the deciding factor.

As stated previously, the reaction to an outage consists of detecting the outage in the first place and responding by switching to redundant components. In some cases, such as physical link failures, both detection and response can be immediate and the traffic is simply forwarded via the alternate path. Both the traditional approach and Mobile IP-based RAIIC (with pre-registrations in place) can execute such responses. For equipment failures, traditional redundancy protocols can be used in both cases with identical performance.

The key difference between technologies then becomes the non-trivial link failures that cannot be immediately detected, and the speed of detection and recovery in such cases. Such failures are typically

conditions where the link is considered operational based on link layer indications, however no data actually passes through. For obtaining accurate values for outage detection and recovery time, we have to study the performance of existing gateway router redundancy implementations that are used in traditional approaches.

The difference between RAIIC-based approach and traditional approaches culminate in the required situational awareness. In MIP-based RAIIC approach, the Mobile Routers need to be aware of the status of the entire path and reachability of the peer Mobile Router, while components in traditional approaches only need to be aware of the reachability of next upstream node. In traditional approach, the nodes have the luxury of assuming that the reliability for the rest of the path is provided independently by other components. Furthermore, the upstream node may provide additional, tailored, services specifically for redundancy.

There are several approaches that can be used to provide detection and recovery. One of the most elementary of all is called a *floating static route*. In this case, the gateway router has two or more routes to same network, via different upstream routers, with different metrics. The router periodically pings a configured primary upstream router. If no responses are received, the link is considered failed and the route via that particular upstream router is dropped, and then all traffic is routed via the secondary upstream router. The behavior is very similar to the keepalive mechanism of MIP-based RAIIC. In this case, the comparison becomes simply of question of thresholds and latencies: Since the RAIIC-based approach requires round-trip for the entire path, the detection time can be slightly higher since the response timeout has to be more tolerant. However, the difference should not prove to be significant in case of regional or national distances, which are typically under 100 ms range.

However, the traditional service providers have additional options beyond the downstream routers actively polling for the status of each link. Since the service provider maintains each node along the entire path, the upstream node can use similar redundancy mechanisms as the gateway router uses for the local network: The previously discussed VRRP and other redundancy protocols. As such, a more thorough analysis of such methods is warranted, as this is the key difference between traditional approaches and RAIIC.

VRRP is the only widely-deployed standards-based solution. VRRP and other similar protocols, such as its proprietary precursor HSRP (Hot Standby Router Protocol), typically work by providing a *virtual IP address*, that is used as a next-hop address by other nodes. The virtual address is assigned to a designated router with the highest configured priority. If the designated router fails, a backup router takes ownership of the virtual address and starts forwarding traffic. The downstream nodes are not aware of the switchover.

VRRP is based on the idea that an election is conducted between redundant equipment to determine which router should be active in handling forwarding. The active router then sends out advertisements periodically to indicate that both the router itself and the link towards the router is still operational. If three consecutive advertisements are not received when expected, the other routers assume that a failure has occurred and a new election process takes place. VRRP has several versions, where the original version had a minimum advertisement interval of one second. As such, the minimum time for a switchover would be three seconds combined with the time for the election process. Subsequent VRRP version 3 and proprietary extensions to earlier versions technically allow for advertisements to be sent more often. However, a huge number of advertisements might overwhelm the network, and in practice such advertisement floods are not configurable in implementations. For example, Juniper's implementation [32] allows for a minimum advertisement interval of 100 ms. In this case, the total time for switching in case of failure would be over 300 ms.

Several proprietary approaches exist as well. For example, ASA firewalls from Cisco [8] implement their own failover system, and require a minimum of five seconds for detecting an outage for a specific case of a failed link (800 ms for failure of an entire unit). The approach of ASA is not based on a virtual address, but instead the failover pair *trade* their addresses when required. Furthermore, state information such as existing flows through the firewall are kept synchronized.

A traditional service provider may also opt to not use routing at all beyond the edge router of the customer, and resort to forwarding traffic solely with link layer technologies. Such an approach is not typical: Usually all customers within a single geographical area are first aggregated at a single distribution router at the provider edge, which then connects to core network of the operator. The core network itself

however, is typically based on various link layer technologies or label switching, such as MPLS [62]. The response times remain comparable in link layer technologies as well. For example, connectivity loss detection in Ethernet spanning trees [24] is based on periodically sent frames, with recovery time in the order of seconds. Faster performance can be achieved with link aggregation using port channels [23], where a failure on single link does not affect performance of adjacent link and does not require spanning tree reconvergence. However, all links belonging to a port channel need to be terminated at same devices. As such, while the approach is suitable for providing link redundancy, failure of a the entire upstream device would still require slower recovery. Proprietary approaches may achieve better performance, however, they have other limitations. For example, the Resilient Ethernet Protocol (REP) from Cisco [7] can achieve 50 ms recovery times in optimal conditions; however, the applications are limited to ring topologies and cannot coexist with other spanning tree technologies.

What follows is that most of the currently available redundancy approaches have similar performance characteristics as RAIIC. The technologies used at the service provider core may provide faster response times than the slower approaches used with routers, but since RAIIC approach omits such considerations completely the path maintenance throughout core is not of real concern. The router-based technologies are slightly faster if tuned to the minimum detection times allowed by implementations, however the difference is not significant, as the detection and recovery times are still starting from a few hundreds of milliseconds in optimal conditions.

4.8 Summary

For evaluating Mobile IP as an approach for implementing RAIIC, the primary consideration is the end-user experience. Traditional end-user experience measurement tools are typically used for determining the overall satisfaction where constant experience over time is assumed. However, in the case of RAIIC, the experience may be perfect for most of the time but suffer from occasional, transient yet total, failures. With Mobile IP, it is possible to reach outage detection and recovery speeds where most interactive applications work throughout the outage without any visible issues. With real time applications, the issues can

be visible to the end user, but their seriousness is hard to quantify. When relying solely on the VoIP MOS scores, the overall quality of a VoIP call does not appear to degrade to overtly poor levels, although MOS as a measurement tool does not completely apply to the situation. Furthermore, when multiple links and load balancing is included in the configuration, MIP provides mechanism to establish and operate multiple, concurrent paths. The actual load balancing strategy between these paths can be chosen arbitrarily, with both flow-based and packet-based methods having merits. Based on the experimentation and comparison to existing approaches, RAIIC appears to be technically equivalent to traditional reliability approaches, although RAIIC requires more stringent preconditions when deployed.

5. Discussion

The topics covered in previous chapters warrant more discussion on their applicability and implications. In this chapter we cover a number of issues, starting from the overall validity of our conclusions, followed by considerations on the effect of load on the RAIIC-based system, extending the system to span multiple organizations, address the remaining fallacies of the MIP-based RAIIC and present possible alternate approaches for implementing similar functionality.

5.1 Validity of conclusions from experiments

The technical portion of the research has mostly been conducted in an incremental fashion. During the process, we have intentionally tried to avoid certain pitfalls, to have as widespread support for our claims on performance of RAIIC as possible. As stated previously, the technical experimentation has had four separate, distinct sub-areas:

- Simulations: Focus on behavior of transport protocols
- Implementation: Focus on signaling performance
- Load balancing: Focus on application throughput and fairness
- Standardization process

To support the claims of our research, we have attempted to create a wide base for our work. The first effort, the simulation work, was implemented from scratch utilizing the ns-3 framework, and the implementation work itself was mostly conducted by the Author. The real-world implementation, that followed the simulation, was based on the Dynamics project [21], a Mobile IP stack originally developed at Helsinki University of Technology. The original Dynamics code was

further improved on by MSc Juho Paaso as part of his MSc thesis. As such, the real-world implementation was independent of the simulation work: Both the code bases and the people engaged in the programming effort were separate. The final work to be implemented, the load balancing experiment, was based on yet another programming effort originally conducted in the EU FP7 Trilogy [67] project, thus yielding a third, separate, independent effort. Furthermore, the standardization process benefited the research with in-depth technical knowledge and feedback to address various intricacies of the system, as the draft of the standard went through several iterations and was modified based on the comments expressed in the MIP4 working group of IETF. In addition, the test cases have been designed to express worst behavior available where possible, such as utilizing relatively elementary codec without any error correction mechanisms for VoIP application testing.

Analytical validation of all the results is not feasible. While this is possible in trivial and simple cases, such as keepalive messaging scalability, validating all possible scenarios with every possible kind of end-user traffic and every kind of path combination would quickly become overwhelming. Such an analysis would have to rely on approximations and would therefore neither validate nor falsify the claims.

5.2 Effect of load on the system

The experiments are based on scenarios where the components of the RAIIC system are operating under much worse conditions than what would be the nominally intended use-case. As has been stated, the overall premise of the work is that while unguaranteed network connectivity may suffer from an outage without any preliminary warning, such events are still relatively rare. Furthermore, some of the signaling can be conducted pre-emptively before the handover even has to take place.

In the case of load balancing experiments, we have been using highly overloaded links and an overtly dynamic environment. In the extreme case, 64 relatively short-lived and greedy flows are concurrently active. Since the flows are short, each flow is going through a lot of different states quickly: TCP slow start, exponential backoff, congestion avoidance, and different kinds of retransmissions, before finally shutting down. A high amount of short flows could translate to hundreds of end-users, as not everyone is constantly accessing services, although peer-to-peer

applications may express similar behavior with lower number of actual end-users. Modern web traffic may consist of larger number of even shorter flows [25]. However, such very short flows never reach the congestion avoidance phase, as during their short lifespans only the slow start is conducted. As such, the effect of any load-balancing process on such flows remains small since they never reach full speed.

Going into the other direction, fewer and longer flows would also be much easier for load balancing components to handle. The loads are much more static, and long-lived flows are in their steady-state congestion avoidance phase for most of the time. As such, the actual performance of the algorithms should be much better when subjected to a network that has been approximately scaled to the intended number of users.

Another consideration besides the user data is the requirents for the VSP back-end. Although the design attempts to avoid directing user traffic via the Home Agent, at some point a single Home Agent may become overloaded due to signaling alone. However, the Home Agent is, in essence, a server responding to requests coming in. The performance of Home Agent can therefore be scaled up by deploying any typical clustering and redundancy approaches. The same applies for most of the other back-end systems, such as the authentication framework. As a result, the Home Agent would still remain a single, logical entity, even when represented via a large number of front-end servers.

5.3 Establishing extranets

When multiple organizations wish to connect their networks directly, this is known as deploying an extranet configuration. Situations where such connectivity is desired vary, and include various subcontracting relationships, joint operations, corporate mergers and the like. There are several existing methods for such connectivity, which are no different from connectivity between sites in same organization. Certain concerns have to be addressed, such as in the case of overlapping IP addresses and the like, and thorough verification that only the desired traffic is passed between organizations. However, none of these concerns are specific to RAIIC.

RAIIC behaves in a similar fashion compared to existing approaches in a sense that reliability guarantees can only be given to connectivity between sites managed by a single operator, virtual or otherwise. If both organizations are customers of same service provider, the connectivity

can be established with reliability guarantees. In the case of RAIIC, any reliability guarantees can only be given if both organizations are customers of same VSP.

Mobile IP-based RAIIC has the additional benefit that Mobile Routers not connected to the same Home Agent, and maintained by different VSPs, can be connected directly. This is possible if the information normally provided by Home Agent, being the the subnet of the partner organization and the Home Address of the Mobile Router, are instead configured statically. The only constraint is that the Home Address of the Mobile Router needs to be an address that is reachable and routable in the Internet, which in a typical deployment is readily satisfied. Even redundancy and reliability can be set up in a similar fashion, although no contractual guarantees can be given. After all, a VSP cannot be responsible for the service levels of another VSP. However, the obtained level of service should still be higher than with the traditional approaches, which in this case is typically a site-to-site VPN over the Internet.

5.4 Fallacies and possible mitigation methods

The RAIIC approach attempts to provide economical, competitive high-availability service by rapidly switching paths upon indication of failure and utilizing all available paths during nominal conditions.

Since the stated objective above all else is redundancy, having load balancing included in the concept appears counterproductive at first. While it increases overall throughput, the alternate paths could potentially be used to send the same data multiple times. If a single path were to fail, no data would be lost. However, in the proposed scenario, this is not feasible, due to the paths not being uniform. The chosen Mobile IP protocol itself supports the concept, and even the original specification from year 1996 [46] support simultaneous bindings. Replicating data at the router would be relatively straightforward.

However, in our scenario, the focus is on site-to-site connectivity with highly diverse connectivity methods and capabilities. Simple data replication would work only with uniform paths. With unequal paths, the throughput would have to be limited to the speeds of the lowest common denominator, or true replication would not be achieved. Even if bandwidths would be equal between paths, the same constraint would apply to all other characteristics of the path, such as latency and MTU.

While simple replication is not feasible in the proposed scenario, it might be more feasible to use some sort of Forward Error Correction (FEC) schema. In such a scenario, the traffic would be divided across all paths, but bandwidths could be unequal. However, even if original packets are encoded and distributed among multiple paths, some characteristics such as latency still affect the overall performance. In fact, some of the research [36] suggest scheduling packets intelligently by delaying transmission differently for each path, causing the arrival to occur at roughly identical delay. Other possible alternatives [70] include having a large receive buffer and reassembling and reordering packets as necessary. Regardless of the approach, the end-result still adds latency to the system, which is problematic for certain applications.

5.5 Mutual dependencies of faults

The overreaching assumption for all the presented work is that with thorough groundwork, a VSP could establish networking service using truly independent connectivity technologies. However, there are some cases where nodes on a specific site may become unreachable. However, the responsibility to the VSP should still remain minimal.

First of all, considering all possible failure modes, VSPs liability limitations need to be taken into account. For example, the VSP is not implementing redundancy for end-hosts, such as servers (See Section 1.1). As such, a scenario where all paths fail due to the customer site suffering from an environmental issue, such as a loss of power, does not have any adverse effects on the operations of the VSP. Furthermore, as stated before, the VSP will provide basic on-site redundancy by installing a minimum of two Mobile Routers as a failover pair.

However, a situation when all available paths are not available can occur. Such a scenario has some sort of common factor which affects connectivity via *all* service providers concurrently. Natural disasters and other extreme conditions typically affect all stakeholders within a specific geographical area. However, for such situations, the SLA should include a *force majeure* clause, limiting liability due to the circumstances. As such, the VSP will not be affected directly, although subsequent effects on customer relationship can vary. However, in such a scenario, the RAIIC-based solution may actually be more resilient than even traditional high

availability connectivity via a single SP, as nothing prevents the VSP from utilizing highly resilient access technologies such as satellite links.

Considering the above, a scenario that affects all service providers and is not covered by such *force majeure* clauses could be caused by e.g. malfunctioning routing [65] or malicious attack on the networks. Ultimately, a scenario where the connectivity fails and the VSP will be held liable can of course be constructed. However, as stated in Chapter 2, the manifestation of such scenarios needs to be taken into account when conducting risk analysis, and the VSP should use any and all mitigation methods available. With all such precautions and limits on liability, the effects of mutual failure of all available paths should remain acceptable.

5.6 Alternative approaches for implementing RAIIC

Mobile IP is one, highly suitable possibility to implement RAIIC, but alternative approaches need to be considered. When considering such alternatives, the previously discussed functional requirements, outlined in Section 2.2, need to be taken into account. The requirements affect both signaling and forwarding of end-user data. However, most of the alternative approaches lack maintenance and monitoring functionalities, in contrast to the payload transfer and tunneling functionalities of Mobile IP. Mobile IP also has readily available mechanisms for implementing both transport and security.

As has been discussed previously, RAIIC has to be implemented within the network infrastructure since no special requirements can be placed on the end-hosts or the upstream network. As such, a network elements implementing RAIIC have certain minimum technical requirements stemming from the functionality, regardless of the approach chosen. The requirements are a signaling protocol for transmitting up-to-date state information and some method to make end-hosts appear behind several network connections at once to the rest of the network. The very basic implementation of a RAIIC-capable node is shown in Figure 5.1, where end-hosts are connected via a “RAIIC element” to multiple, independent service provider networks.

Examples of standardized signaling protocols and protocol suites are covered below, and how they could be used for implementing RAIIC is discussed.

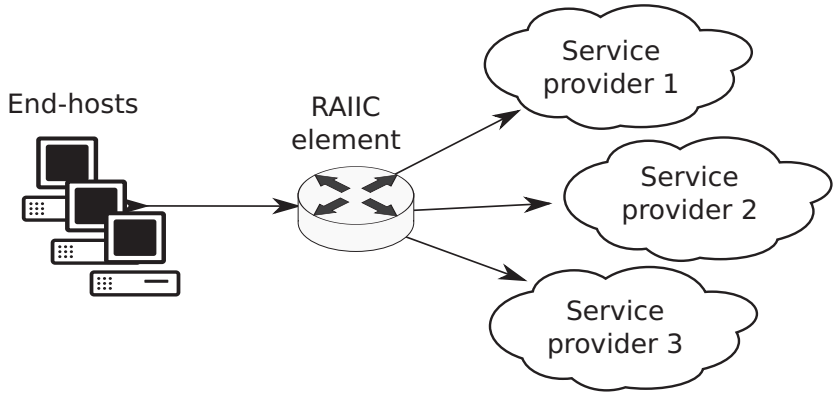

Figure 5.1. The basic conceptual elements of RAIIC for a single site. Peer sites have similar components.

- Simple Network Management Protocol (SNMP) [20] could potentially be used as a signaling protocol for establishing tunnels. However, the resulting approach would have a centralized point for conducting the routing decisions, instead of the distributed model that could be achieved with the other methods. The establishment process could work by having a central coordinator, such as a network management system (NMS) utilizing SNMP SET requests to explicitly command the RAIIC elements. Such commands would require these functionalities to exist in the Management Information Base (MIB). Although there are several MIBs available for requesting information from network elements, additional work would be needed to set up appropriate command signaling format. If conducted using SNMP version 3, bulk requests would allow conducting several operations at once and AuthPriv options could be used to provide security. However, this approach would shift the decision-making to a centralized point, with similar drawbacks as the previously covered dynamic VPN methods.

- Host identity Protocol (HIP) [41] is a basis for an architecture where the locator (IP address) is separated from the identity of a host. HIP-aware nodes communicate using identities, and establish a mapping from identity to locator or locators. The HIP Base Exchange mechanism allows for direct connectivity between HIP nodes, and locating the potential peer nodes is conducted via a special node known as Rendezvous Point (RP). As an analogy to Mobile IP, the identity could be considered Home Address, the locator the Care-of Address, and the RP is Home Agent. A possible advantage of HIP is the provided

alternative for basic NAT or tunneling, as the RAIIC element could represent the identity of end-hosts and instead of NAT, the translation would occur between identities and locators. However, the HIP approach ultimately requires an environment where no restrictions are placed on connectivity, and current Internet is not such an environment. While Mobile IP works around the issue by such techniques as altering tunnel establishment directions, HIP has no comparable mechanisms.

- Session Initiation Protocol, (SIP) [58] is used for establishing communications sessions, typically phone calls in VoIP applications, but can be used for plethora of other purposes. The sessions are described using an encapsulated description Session Description Protocol, SDP [19]. Fundamentally, SIP could be used for similar signaling as Mobile IP conducts. To gain MIP-like functionality, the SDP would need additional specifications to allow for network infrastructure information and tunnel establishment. This additional work could use such standards as IKE over SDP [59] for a basis. One large drawback is that the messaging with SIP is intended for end-user applications, and as typical in application-level protocols, all signaling occurs with text fields. The text-based signaling would affect signaling scalability, which would suffer highly as network sizes increase. Conversely, SIP benefits from an established signal proxying architecture. The proxy functionality could allow for session (tunnel) establishment signaling to take a very complex path without significant additional work.

Comparison of the functional equivalents of different approaches can be seen in Table 5.1.

Table 5.1. Comparison of possible alternative approaches for implementing RAIIC

Approach	Coordinator	Signaling	Security	
			Control	Data*
Mobile IP	HA	Registrations	Return Routability	IPSec
SIP	Registrar	SIP + SDP	TLS, Certificates	TLS, IPSec
SNMP	NMS	SNMP SETs	v3 Priv option	any
HIP	RP	HIP BEX	Diffie-Hellman	IPSec

*Technically, any possible security scheme can be used for securing user data with any of these approaches. However, these are the ones implemented currently.

As noted, there are fallacies with each chosen example, and the examples can be generalized further to cover other similar approaches. The SNMP-based method represents centralized management with any desired signaling protocol, where individual RAIIC elements relegate decision-making to the central system. However, the latencies caused by this relegation would very likely have adverse effects on usability.

The HIP-based method represents various mapping approaches. Such methods can be very effective and relatively lightweight, and HIP itself has been shown to work well in certain applications [35]. However, such protocols place additional requirements on the IP networks and are therefore not suitable for general deployment. The lack of assumptions that can be made on SP networks also rule out traditional routing protocols and, for example, Resource Reservation Protocol (RSVP) [6].

Finally, the SIP example is very close to Mobile IP in terms of functionality. However, compared to Mobile IP, where all the required functionalities are built into the same protocol suite, with SIP the integration is much less focused, and due to the generic nature, does not benefit from the application-specific optimizations that are built into MIP. Mobile IP provides both the control plane (signaling) and data plane (tunneling) together as part of same protocol standards, and consequently is the only widely deployed standard for layer 3 mobility management. While mobility management in general has been developed significantly, the focus appears to be in layer 2 technologies, such as facilitating rapid handovers between wireless networks.

Of course, if the environment has less restrictions on the functionality, the number of possible approaches grow extensively. For example, if the RAIIC element at each site would have static addressing and reachability at each egress interface, the tunnels providing connectivity between sites can be configured statically. In such environments, the focus would shift from maintaining the connectivity towards perfecting the utilization of the network instead, and could benefit particularly from research towards ad-hoc networking.

6. Conclusions

The research was executed to look into the possibility of creating an economic approach to provide reliable network connectivity, equivalent to the traditional “High availability” solutions. During the research it was observed that while the RAIIC approach has its share of caveats, the economic feasibility analysis shows that such an offering could be very interesting for a number of stakeholders. However, establishing the new offering as a credible option might take considerable efforts, as the prerequisites for the environment and personnel competence would be quite high. If deployed successfully, such an offering could result in structural changes to the rather stagnated service market for reliable networking.

As a technology, we have shown that the RAIIC approach is a feasible option in environments where multiple independent service providers operate concurrently. RAIIC can be implemented with Mobile IP, which can be successfully extended to provide a number of features that are required for the scheme to function. We have implemented the required features in our own work, and the technology can be enhanced even further with certain additional extensions. The end-user experience, the most important metric for mass deployment, remains at satisfactory levels with the analyzed applications. From a contractual perspective, if traditional SLA metrics for reliability are used, MIP-based RAIIC could possibly be deployed as a drop-in replacement for traditional high-availability offering. Traditional functionalities such as “five-nines” reliability could be achieved comfortably, even in an environment with relatively frequent outages, such as once or twice a day.

One of the strengths of RAIIC is that from the end-user and customer network viewpoint, it appears to be just like a regular WAN connection. As such, partial and incremental deployment is possible. Furthermore,

traditional QoS tools, such as traffic classification and priority queuing work without changes.

However, many of the existing end-user experience metrics and measurement methods are based on traditional operational environments. Such metrics are tuned for conditions that are relatively static for a prolonged period of time, such as for a duration of an entire phone call or a website visit. The differences apparent in RAIIC approach could require research and development of new, more suitable, end user experience metrics. These new metrics should provide information on how much the experience is actually affected by short and transient outages that are relatively common. As such, more research on the subject should be conducted, and might be applicable in a number of other situations as well. The results may eventually even affect the SLA terms, conditions and requirements.

There are several additional areas that could be studied further. On the technical side, research avenues include methods for reducing the switchover time further and better algorithms for intelligent load balancing and packet scheduling. To reduce the switchover time, two sub-components exist: outage detection and recovery. Although the outage detection typically works only after the outage has occurred, possibilities for triggering the detection earlier, in an anticipatory fashion, due to subtle changes in path characteristics could allow enhancing the performance further. On the other hand, enhancing recovery would consist mostly of relaying the information on the outage to the rest of the infrastructure faster. Shortening the recovery delay further could benefit from the vast latency reducing research conducted on wireless networks.

The more intelligent load balancing and packet scheduling could be achieved by adding more heuristics to the path selection decision. The algorithms designed and implemented during this research are rather rudimentary and work as proof-of-concepts. Creating more refined approaches can result in much better fairness and utilization behavior.

In a more general scope, the research has also demonstrated that load balancing network traffic dynamically among several paths can be conducted as part of the network infrastructure even in wide area scenarios, not only within the confines of a limited area, such as a data center. The strategies that can be used for packet scheduling vary depending on the general network environment, however the

infrastructure could also detect such changes and adjust strategies accordingly.

Continuing on the generalization, the original use-case of RAIIC was virtualizing traditional corporate networks with reliability requirements. However, the same approach where multiple, less expensive connections are bundled together to provide an overall reliable path between nodes could serve other purposes as well. An interesting possible case is data centers. In data centers, multipathing implemented as part of the infrastructure would place no requirements on the end-hosts, unlike application-level multipath approaches. In a data center environment with highly variable loads on individual servers, infrastructure-level implementation of dynamic multipathing appears to be a very attractive option. A RAIIC-like approach could be used to dynamically pick optimal paths between individual end-hosts with the information on the state of the entire network available constantly. Furthermore, the possibility to conduct load balancing between multiple, heterogeneous paths could be very attractive for "Big Data" applications where massive quantities of data are sent between separate data centers. Both the reliability and optimization for throughput could be addressed as part of the same solution.

Of course, the possibilities do not stop there. Even if the RAIIC is not ever deployed in the proposed fashion, the solutions obtained during this research could be applied even to a relatively consumer-oriented application – such as providing reliable and unrestricted Internet connectivity for apartments, where a landlord could use a RAIIC-like scheme to transparently bundle several regular Internet services together. If nothing else, it is my wish that the work in this thesis has shown that obtaining true reliability for your network does not necessarily require either sacrificing flexibility or paying hideously expensive prices.

Bibliography

[1] AMD Geode LX Processor Family, Advanced Micro Devices, Inc, 2012.

[2] M. Allman, V. Paxson, and W. Stevens. TCP Congestion Control. Internet Engineering Task Force, RFC 2581, April 1999.

[3] T. Bates and Y. Rekhter. Scalable Support for Multi-homed Multi-provider Connectivity. Internet Engineering Task Force, RFC 2260, January 1998.

[4] P. Bellavista, A. Corradi, and C. Giannelli. Mobility-aware middleware for self-organizing heterogeneous networks with multihop multipath connectivity. *Wireless Communications, IEEE*, 15(6):22 –30, December 2008.

[5] Vaduvur Bharghavan. Challenges and Solutions to Adaptive Computing and Seamless Mobility over Heterogeneous Wireless Networks. *Wireless Personal Communications*, 4:217–256, 1997.

[6] R. Braden, L. Zhang, S. Berson, S. Herzog, and S. Jamin. Resource ReSerVation Protocol (RSVP) – Version 1 Functional Specification. Internet Engineering Task Force, RFC 2205, September 1997.

[7] Cisco Resilient Ethernet Protocol, White paper, Cisco Systems, 2007.

[8] Cisco ASA 5500 Series Command Reference, 8.4 and 8.5, Cisco Systems, 2012.

[9] Enhanced Interior Gateway Routing Protocol, Cisco Systems, Cisco document id 16406, September 2005.

[10] A. Decros. Business analysis of a high-availability Intranet solution. Master's thesis, Aalto University, School of Electrical Engineering, 2011.

[11] V. Devarapalli and P. Eronen. Secure Connectivity and Mobility Using Mobile IPv4 and IKEv2 Mobility and Multihoming (MOBIKE). Internet Engineering Task Force, RFC 5266, June 2008.

[12] D. Farinacci, T. Li, S. Hanks, D. Meyer, and P. Traina. Generic Routing Encapsulation (GRE). Internet Engineering Task Force, RFC 2784, March 2000.

[13] B. Fenner, M. Handley, H. Holbrook, and I. Kouvelas. Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised). Internet Engineering Task Force, RFC 4601, August 2006.

[14] Sally Floyd and Van Jacobson. Random Early Detection Gateways for Congestion Avoidance. *IEEE/ACM Transactions on Networking*, 1:397–413, 1993.

[15] A. Ford, C. Raiciu, M. Handley, S. Barre, and J. Iyengar. Architectural Guidelines for Multipath TCP Development. Internet Engineering Task Force, RFC 6182, March 2011.

[16] Dennis F. Galletta, Raymond M. Henry, Scott McCoy, and Peter Polak. Web Site Delays: How Tolerant are Users? *Journal of the Association for Information Systems*, 5(1):0–, 2004.

[17] Inter-Service Provider IP Backbone Guidelines, GSM Association official document IR.34, version 5.0, December 2010.

[18] S. Gundavelli, K. Leung, G. Tsirtsis, H. Soliman, and A. Petrescu. Flow Binding Support for Mobile IPv4, IETF draft (work in progress) draft-ietf-mip4-multiple-tunnel-support-03.txt, February 2012.

[19] M. Handley, V. Jacobson, and C. Perkins. SDP: Session Description Protocol. Internet Engineering Task Force, RFC 4566, July 2006.

[20] D. Harrington, R. Presuhn, and B. Wijnen. An Architecture for Describing Simple Network Management Protocol (SNMP) Management Frameworks. Internet Engineering Task Force, RFC 3411, December 2002.

[21] Dynamics Mobile IP – Introduction, Helsinki University of Technology, 2001.

[22] S. Hemminger. Network Emulation with NetEm. In *proceedings of Linux.Conf.Au (LCA)*, April 2005.

[23] IEEE Standard for Local and metropolitan area networks – Link Aggregation, IEEE Standard 802.1AX-2008, 2008.

[24] IEEE Standard for Local and metropolitan area networks – Media Access Control (MAC) Bridges, IEEE Standard 802.1D-2004, 2004.

[25] Sunghwan Ihm and Vivek S. Pai. Towards understanding modern web traffic. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, IMC '11, pages 295–312, 2011.

[26] The E-model, a computational model for use in transmission planning, ITU-T Recommendation G.107 (2009-04), 2009.

[27] Pulse Code Modulation (PCM) of Voice Frequencies, ITU-T Recommendation G.711, November 1988.

[28] Architecture of transport networks based on the synchronous digital hierarchy (SDH), ITU-T Recommendation G.803 (2000-03), 2000.

[29] Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs, ITU-T Recommendation P.862 (02/01), 2001.

[30] Rajendra K. Jain, Dah-Ming W. Chiu, and William R. Hawe. A Quantitative Measure Of Fairness And Discrimination For Resource Allocation In Shared Computer Systems. Technical report, Digital Equipment Corporation, September 1984.

[31] F. Johansson and T. Johansson. Mobile IPv4 Extension for Carrying Network Access Identifiers. Internet Engineering Task Force, RFC 3846, June 2004.

[32] Configuring the Advertisement Interval for the VRRP Master Router, JUNOS 9.5 High Availability Configuration Guide, Juniper Networks, 2010.

[33] S. Kent and K. Seo. Security Architecture for the Internet Protocol. Internet Engineering Task Force, RFC 4301, December 2005.

[34] E. Kohler, M. Handley, and S. Floyd. Datagram Congestion Control Protocol (DCCP). Internet Engineering Task Force, RFC 4340, March 2006.

[35] Jouni Korhonen, Antti Mäkelä, and Teemu Rinta-Aho. HIP Based Network Access Protocol in Operator Network Deployments. In *First Ambient Networks Workshop on Mobility, Multiaccess, and Network Management (M2NM 2007)*, 2007.

[36] M. Kurant. Exploiting the path propagation time differences in multipath transmission with fec. *Selected Areas in Communications, IEEE Journal on*, 29(5):1021 –1031, May 2011.

[37] K. Leung, G. Dommety, V. Narayanan, and A. Petrescu. Network Mobility (NEMO) Extensions for Mobile IPv4. Internet Engineering Task Force, RFC 5177, April 2008.

[38] H. Levkowetz and S. Vaarala. Mobile IP Traversal of Network Address Translation (NAT) Devices. Internet Engineering Task Force, RFC 3519, April 2003.

[39] K. El Malki. Low-Latency Handoffs in Mobile IPv4. Internet Engineering Task Force, RFC 4881, June 2007.

[40] P.V. Mockapetris. Domain names - implementation and specification. Internet Engineering Task Force, RFC 1035, November 1987.

[41] R. Moskowitz, P. Nikander, P. Jokela, and T. Henderson. Host Identity Protocol. Internet Engineering Task Force, RFC 5201, April 2008.

[42] S. Nadas. Virtual Router Redundancy Protocol (VRRP) Version 3 for IPv4 and IPv6. Internet Engineering Task Force, RFC 5798, March 2010.

[43] J. Paaso. Guaranteed access over consumer-level connections. Master's thesis, Aalto University, School of Electrical Engineering, 2011.

[44] ALIX system boards, PC Engines GmbH, 2011.

[45] C. Perkins. IP Encapsulation within IP. Internet Engineering Task Force, RFC 2003, October 1996.

[46] C. Perkins. IP Mobility Support. Internet Engineering Task Force, RFC 2002, October 1996.

[47] C. Perkins. IP Mobility Support for IPv4, Revised. Internet Engineering Task Force, RFC 5944, November 2010.

[48] C. Perkins, D. Johnson, and J. Arkko. Mobility Support in IPv6. Internet Engineering Task Force, RFC 6275, July 2011.

[49] Laurence J. Peter and Raymond Hull. *The Peter principle*. William Morrow and Company, 1969.

[50] J. Postel. User Datagram Protocol. Internet Engineering Task Force, RFC 768, August 1980.

[51] J. Postel. Internet Control Message Protocol. Internet Engineering Task Force, RFC 792, September 1981.

[52] J. Postel. Transmission Control Protocol. Internet Engineering Task Force, RFC 793, September 1981.

[53] B. Venkata Ramana, B. S. Manoj, and C. Siva Ram Murthy. AR-TCP: a loss-aware adaptive rate based TCP for ad hoc wireless networks. *Journal of High Speed Networks*, 15:53–72, January 2006.

[54] Klaus Rechert, Patrick McHardy, and Martin A. Brown. HFSC Scheduling with Linux, http://linux-ip.net/articles/hfsc.en/, Retrieved on 24th of October, 2011, 2006.

[55] Y. Rekhter, T. Li, and S. Hares. A Border Gateway Protocol 4 (BGP-4). Internet Engineering Task Force, RFC 4271, January 2006.

[56] IPv4 Address Allocation and Assignment Policies for the RIPE NCC Service Region, Réseaux Internet Protocol Européens (RIPE), RIPE document id 524, August 2011.

[57] N. Rickard. Cost Cutting by Rightsizing Network Reliability, Gartner research report G00155940, April 2008.

[58] J. Rosenberg, H. Schulzrinne, G. Camarillo, A. Johnston, J. Peterson, R. Sparks, M. Handley, and E. Schooler. SIP: Session Initiation Protocol. Internet Engineering Task Force, RFC 3261, June 2002.

[59] M. Saito, D. Wing, and M. Toyama. Media Description for the Internet Key Exchange Protocol (IKE) in the Session Description Protocol (SDP). Internet Engineering Task Force, RFC 6193, April 2011.

[60] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson. RTP: A Transport Protocol for Real-Time Applications. Internet Engineering Task Force, RFC 3550, July 2003.

[61] H. Schulzrinne, A. Rao, and R. Lanphier. Real Time Streaming Protocol (RTSP). Internet Engineering Task Force, RFC 2326, April 1998.

[62] V. Sharma and F. Hellstrand. Framework for Multi-Protocol Label Switching (MPLS)-based Recovery. Internet Engineering Task Force, RFC 3469, February 2003.

[63] Spirent Avalanche Datasheet, Spirent Communications, Inc, 2011.

[64] R. Stewart. Stream Control Transmission Protocol. Internet Engineering Task Force, RFC 4960, September 2007.

[65] Peter Svensson. Pakistan Causes Worldwide YouTube Outage, Associated Press, February 2008.

[66] Renata Teixeira, Steve Uhlig, and Christophe Diot. BGP route propagation between neighboring domains. In *Proceedings of the 8th international conference on Passive and active network measurement*, PAM'07, pages 11–21, 2007.

[67] Trilogy - Architecting the Future Internet, Trilogy Consortium, 2011.

[68] Mark Twain. *A Connecticut Yankee in King Arthur's Court*. Charles L. Webster and Company, 1889.

[69] P. Vixie, S. Thomson, Y. Rekhter, and J. Bound. Dynamic Updates in the Domain Name System (DNS UPDATE). Internet Engineering Task Force, RFC 2136, April 1997.

[70] Hui Zhang, Wei Jiang, Jin Zhou, Zhen Chen, and Jun Li. M3FEC: Joint Multiple Description Coding and Forward Error Correction for Interactive Multimedia in Multiple Path Transmission. *Tsinghua Science and Technology*, 16(3):320 – 331, 2011.

Errata

Publication I

No erratas

Publication II

No erratas

Publication III

No erratas

Publication IV

No erratas

Publication V

Figure 5 on page 5 does not properly show the messaging paths between HQ and site A. Two paths should be visible; one via ISP 1 and other via ISP 2. Figures 9 and 10 are not referenced; in Section V, Subsection C, Figure 9 is about the studied "full scenario", and Figure 10 about the "potential worst-case scenario". The header in Figures 8, 9 and 10 should be "Average throughput of single flow, 2x20 + 1x10 flows, 7Mbps and 4Mbps links".

Publication VI

No erratas

Publication VII

No erratas

Publication VIII

No erratas

Aalto-DD 66/2012

ISBN 978-952-60-4634-1
ISBN 978-952-60-4635-8 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934
ISSN 1799-4942 (pdf)

Aalto University
School of Electrical Engineering
Department of Communications and Networking
www.aalto.fi

BUSINESS + ECONOMY

ART + DESIGN + ARCHITECTURE

SCIENCE + TECHNOLOGY

CROSSOVER

DOCTORAL DISSERTATIONS