Andres Javier Gonzalez Martinez

Methods For Guaranteeing Contracted Availability In Connection Oriented Networks

Thesis for the degree of Philosophiae Doctor

Trondheim, March 2013

Norwegian University of Science and Technology
Faculty of Information Technology, Mathematics
and Electrical Engineering
Department of Telematics

NTNU – Trondheim
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the degree of Philosophiae Doctor

Faculty of Information Technology, Mathematics and Electrical Engineering
Department of Telematics

ISBN 978-82-471-4303-2 (printed ver.)
ISBN 978-82-471-4304-9 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2013:102

Printed by NTNU-trykk

Abstract

Real telecommunications networks are not failure free. Any single disconnection impacts the network provider reputation and finances, and produces incalculable consequences to the customers through the affected applications. A common policy to handle this issue is the stipulation of the availability to be guaranteed in a business contract known as Service Level Agreement SLA. The stipulated availability must be commercially competitive, and it must fit the customer needs. However, the fulfillment of the SLA may imply huge costs in terms of the resources to be reserved and/or the penalties associated with the violation of the agreement. In addition, the selection of the availability to be stipulated, is a difficult task due to the following challenges: (1) SLAs are defined for a finite time interval which demands the study of the probability distribution of the interval availability and the respective risk that it represents. (2) Having failure and repair data from operational networks is a must, in order to assess accurately the SLA risk. However, this kind of information is limited for a number of reasons, among them that failures are not what operators like to have exposed in a competitive commercial marketplace. (3) The end-to-end interval availability is affected by several stochastic processes. The networks addressed in this thesis are compound systems where Markovian assumptions do not apply. The duration of up and down times are not exponentially distributed, and correlation between failure/repair processes may exist. (4) Designing assignment policies to use efficiently network resources is a classical challenge that becomes harder when SLA availability constraints have to be considered. This thesis addresses the mentioned challenges as follows.

In the first part of the thesis, a theoretical study of the probability distribution of the interval availability is made. A mathematical approximation to evaluate the cumulative downtime distribution of a single-component system that does not have the Markovian properties is proposed, and the evolution of the distribution of the interval availability with the increase of the observation period is studied. This study shows how sensitive is the SLA risk to: i) The distribution of up and down times. ii) The duration of the SLA.

An important part of the thesis is the study of operational failure and repair events obtained from measurements of the UNINETT core network. In this study, up and down times of routers and links are characterized, and the correlation between failure and repair processes is studied. Network components are classified according to their dependability characteristics. The information obtained from the characterization

phase is used as an input in the simulations developed in later parts of the thesis, and in order to justify some of the assumptions made. These analyzes lead to the conclusion that failure and repair processes in a backbone network are not independent and do not have the Markovian properties.

In this thesis, the probability that the availability offered to a compound network connection after the contract duration is less than the availability promised in the SLA is assessed using simulation techniques. First, unprotected, shared protected and dedicated protected connections under non-Markovian failure and repair processes are studied. In addition, two methods to model correlated Weibull, gamma and empirically distributed up and down times are proposed. The first method uses trace driven simulation combined with random circular shifting. The second method uses Monte Carlo techniques. Through these methods, the SLA risk may be assessed, considering real/operational network features.

This thesis discusses how to allocate requested connections in a given network topology under SLA availability constraints. An intelligent sharing mechanism to use the bandwidth efficiently is proposed. In addition, this thesis studies the SLA penalty scheme and it proposes a model to allocate connections, fulfilling SLA requirements, and maximizing the operator profit, through a two-stage stochastic optimization program. The model considers the stochastic behavior of network components, correlation between failure and repair processes, the SLA finite duration, and the flexibility to allocate or reject a connection based on its impact on the provider profit.

Finally, the problem of guaranteeing SLA availability is studied in cloud computing environments. This study, proposes the use of the *SLA-budget* for the implementation of smart policies in: **i)** the assignment of spare servers when virtual machines are restored. **ii)** the dynamic use of fault tolerance licenses. The result is a considerable reduction in the probability of failing the SLA availability requirement by making an efficient use of the cloud resources available. This work is a first step in the design of SLA-aware cloud computing management, and it illustrates how the distribution of the interval availability may be manipulated by mechanisms under the provider control.

To my beautiful grandmother Belen Reina and everything she represents.

Preface

This thesis is submitted in partial fulfillment of the requirements for the degree of philosophiae doctor (PhD) at the Norwegian University of Science and Technology (NTNU). The work was performed at the Centre for Quantifiable Quality of Service in Communication Systems (Q2S), Centre of Excellence (CoE), during 2008-2012, and has been supervised by Professor Bjarne E. Helvik.

Parts of this work were conducted within the European project on the Network of the Future Euro-NF. In addition, this work was strengthened by the help received from the Norwegian Research Network UNINETT, who provided the log of failure and repair events in its backbone network for the period 2001-2009.

The last six months of the PhD were funded by Telenor and the Department of Telematics of the Norwegian University of Science and Technology.

The papers included in this thesis have been subject to minor editorial changes since their publication.

Acknowledgements

Numerous people have directly or indirectly contributed to the work presented in this thesis. First of all, I would like to thank my supervisor Professor Bjarne E. Helvik. This work would never have been completed without his help, support and dedication through the interesting discussions that we had during the last four years. The help received by Jon Kåre Hellan and Olav Kvittem from UNINETT was fundamental, due to it gave me the advantage of working with operational data, and the motivation to develop further work. I would also like to thank my co-authors Pirkko Kuusela for fruitful discussions on data analysis and the thesis preparation. Then, I would like to thank my office mate Addisu Tadesse Eshete for four invaluable years of academic and personal growth. In addition, thanks to all my friends Anne Nevin, Laurent Paquereau, Mark Stegelmann, Kashif Mahmood, Máté J. Csorba, Rune Ødegård, Atef Abdelkefi, Eirik Larsen Følstad, Fazal Wahab, Pern Hui, Peiqing Zhang, Viet Thi Minh, Benedikt Westermann, Ulrich reiter, Jordi Puig, Razib Khan, Fitri Rahayu and all colleagues who have created a very enjoyable atmosphere at Q2S in coffee breaks and lunches. Special thanks to Annikken Skotvoll for her incredible kindness and help, and to Svein Johan Knapskog, Hans Almåsbakk and Mette Veronica Olsen for providing the environment needed to make this thesis possible.

Finally, I would like to thank my family. My parents: Julia Elvira Martinez and Jose Gonzalez who have been my support and inspiration always. With your example and love, you gave me the strength and motivation needed to finish this dissertation. My wife Adriana Maria Sanabria and my daughter Maria Paz Gonzalez for making my life more meaningful, and filling me with love and incredible happiness. My sister Laura Gonzalez, together we took the first step into this four years adventure, and we will support each other in the projects that will come. My grandfather Mardoqueo Martinez and his calls and prays that motivate me incredibly. My nice Manuelita, my uncles, aunts and cousins for cheering me up all the time, and specially to my beautiful grandmother Belen Reina for taking care of me and for guide me with her sublime example.

Contents

Abbreviations

CDF Cumulative Distribution Function

CHF Cumulative Hazard Function

CRM Customer Relationship Management

DPP Dedicated Path Protection

FEC Forward Equivalence Class

FIFO First In, First Out

FT Fault Tolerance

HA High Availability

i.i.d. independent and identically distributed

IaaS Infrastructure as a Service

IETF Internet Engineering Task Force

ILP Integer Linear Programming

IP Internet Protocol

IS-IS Intermediate System To Intermediate System

ISP Internet Service Provider

LB Lower Bound

LSP Label Switched Path

MDT Mean Down Time

MLE Maximum Likelihood Estimation

MPLS Multi-Protocol Label Switching

MTBF Mean Time Between Failures

MTTF Mean Time To Failure

MTTR Mean Time To Repair

MUT Mean Up Time

n.e.d negatively exponentially distributed

NOC Network Operations Center

PDF Probability Density Function

PSG Potential Sharing Group

QoS Quality of Service

Q-Q Quantile-Quantile

RPSG Reduced Potential Sharing Group

RSVP Resource Reservation Protocol

SAA Sample Average Approximation

SAN Storage Area Network

SBPP Shared Backup Path Protection

SLA Service Level Agreement

SNMP Simple Network Management Protocol

SRLG Shared Risk Link Group

TDS Trace Driven Simulation

UB Upper Bound

VM Virtual Machine

WDM Wavelength Division Multiplexing

List of Papers

Publications Included in the Thesis

These papers are included as Part II of this thesis. Note that some of the papers have been subject to minor editorial changes since their publication.

- PAPER A:
  Andres J. Gonzalez and Bjarne E. Helvik. *A Study of the Interval Availability and its Impact on SLAs Risk*. SPRINGER Proceedings of the International Conference on Computer Science, Engineering and Applications ICCSEA. Delhi, India, May, 2012.

- PAPER B:
  Andres J. Gonzalez and Bjarne E. Helvik. *Characterization of Router and Link Failure Processes in UNINETT's IP Backbone Network*. INDERSCIENCE International Journal of Space-Based and Situated Computing (IJSSC). Vol. 2, No 1, pp. 3 -11, 2012.

- PAPER C:
  Andres J. Gonzalez, Bjarne E. Helvik, Jon K. Hellan and Pirkko Kuusela. *Analysis of Dependencies Between Failures in the UNINETT IP Backbone Network*. IEEE Proceedings of the 16th Pacific Rim International Symposium on Dependable Computing PRDC. Tokio, Japan, December, 2010.

- PAPER D:
  Andres J. Gonzalez and Bjarne E. Helvik. *Guaranteeing Service Availability in SLAs; A Study of the Risk Associated with Contract Period and Failure Process*. IEEE Latin-American Conference on Communications LATINCOM. Medellin, Colombia, September, 2009.

- PAPER E:
  Andres J. Gonzalez and Bjarne E. Helvik. *SLA Success Probability Assessment in Networks with Correlated Failures*. ELSEVIER Computer Communications Journal. (ACCEPTED).

- PAPER F:
  Andres J. Gonzalez and Bjarne E. Helvik. *Dynamic Sharing Mechanism for Guaranteed Availability in MPLS Based Networks*. IEEE Proceedings of the

International Communications Quality and Reliability Workshop CQR. Vancouver, Canada, June, 2010.

- PAPER G:
  Andres J. Gonzalez and Bjarne E. Helvik. *Guaranteeing SLA Availability in Telecommunications Networks*. IEEE Proceedings of the International Telecommunications Network Strategy and Planning Symposium. Rome, Italy, October, 2012.

- PAPER H:
  Andres J. Gonzalez and Bjarne E. Helvik. *System Management to Comply with SLA Availability Guarantees in Cloud Computing*. IEEE International Conference on Cloud Computing Technology and Science CloudCom. Taipei, Taiwan, December, 2012.

Other Papers by the Author

These papers were also prepared while working with this thesis.

- Andres J. Gonzalez and Bjarne E. Helvik. *Guaranteeing Service Availability in SLAs; A Study of the Risk Associated with Contract Period and Failure Process*. IEEE Latin America Transactions, Volume 8, Issue 4. 2010.
  – This paper is the Spanish version of Paper D.

- Andres J. Gonzalez and Bjarne E. Helvik. *Analysis of failures characteristics in the UNINETT IP Backbone network*. IEEE Proceedings of the 7th International Symposium on Frontiers in Networking with Applications, FINA. Singapore, March, 2011.
  – This paper is an early version of Paper B.

- Andres J. Gonzalez and Bjarne E. Helvik. *Guaranteeing Service Availability in SLAs on Networks with Non Independent Failures*. IEEE Proceedings of the 8th International Workshop on Design of Reliable Communication Networks DRCN. Krakow, Poland, October, 2011.
  – This paper is an early version of Paper E.

- Andres J. Gonzalez and Bjarne E. Helvik. *Correlation Between Failure Events and Geographical Distance in the UNINETT IP Backbone Network* Euro-NF Workshop on Traffic Engineering and Dependability in the Network of the Future (WTEDNF) 2010, Warsaw, Poland.

- Andres J. Gonzalez and Bjarne E. Helvik. *Stochastic Optimization for Allocate Connections under SLAs Guarantees* Euro-NF Workshop on Traffic Engineering and Dependability in the Network of the Future (WTEDNF) 2011, Trondheim, Norway.

Part I

THESIS INTRODUCTION

Introduction

The importance of telecommunications networks in human life has increased tremendously during the last years. For this reason, guaranteeing their appropriate operation is a matter of prime interest, not only for networks providers, but also for society as a whole.

Network failures are events that may attract remarkable attention, due to the considerable consequences that they may imply. For instance, a failure in one of the largest telecommunications operators in Norway was widely covered by the media, and even politicians had to take part, due to the impact generated on the entire society [Sti11].

Real world networks are not fault free. Every day, several network elements present failures that affect the normal operation. This is a fact that providers and customers have to face. Network providers cannot promise uninterrupted services due to the unavoidable failures events. Customers need to know the percentage of time that the offered service will be operational (availability), in order to plan the use the hired services. In order to tackle this situation, the lowest availability α that may be accepted by the customer without incurring penalties for the provider has to be clearly specified in a Service Level Agreement (SLA), which is an important tool used among others, to define requirements in business relationships.

In this context, a natural question for network providers is: How to define and guarantee the availability requirement? The goal of this thesis is to provide new concepts and methods for solving this question.

Figure 1 offers a general overview of the problem addressed by this thesis. It involves a network provider who owns a network infrastructure that may offer connectivity between different points, and a customer who needs connectivity between two locations. This is a conventional supply/demand situation, where both parties have to negotiate and agree the obligations and limits of the service to be provided. This scenario considers an SLA that includes among others, the price to be paid by the customer, the availability to be fulfilled, the duration of the contract, and the penalty scheme in case of contract violation. On the other hand, there is the operational state of the end-to-end connection, which depends on the stochastic behavior of network components, repair processes, and network management.

Estimating the availability that a network provider may offer over a certain time interval is a challenging task. One can say that to date, network operators do not have appropriate tools in order to assess precisely the risk of failing an SLA, as well

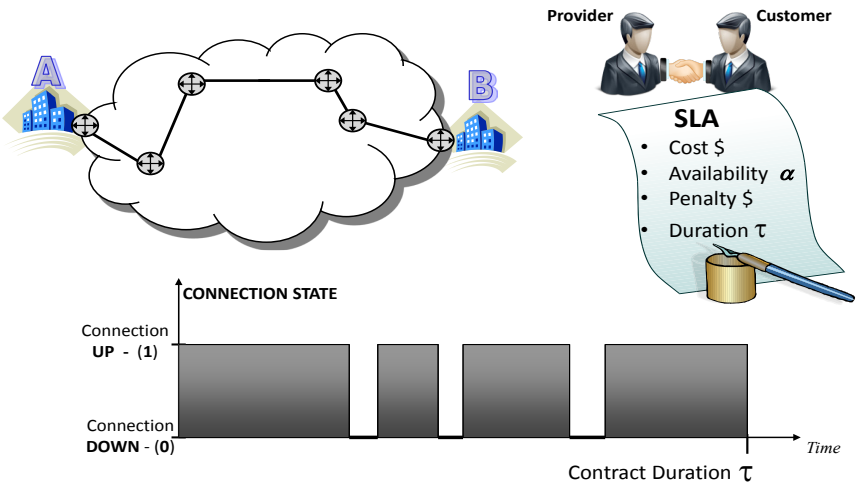


Figure 1. Guaranteeing contracted availability (General Overview).

as means to manage network operations to minimize the probability of failing the contract.

This thesis is composed by two parts. Part I gives an introduction to the thesis. It presents the research problems, the different challenges involved, and it shows how they were solved through research reported in the papers written during the PhD. The main part of this thesis is Part II. It is a collection of eight papers (Papers A-H) analyzing different related topics. The sequence of the papers presented in Part II was chosen in order to have a clearer presentation and easy flow of the concepts addressed, and it does not strictly match the chronological order of the publications.

The introduction of the thesis (Part I) is organized as follows: A presentation of the thesis objective and context is given in Section 1. The general background of the work is presented in Section 2. Then, the state of the art and the open issues on the problem of guaranteeing the SLA availability requirement are presented in Section 3. The research goals of the thesis are presented in Section 4. The research methodology is presented in Section 5. The contributions of the thesis are presented in Section 6. Finally, the thesis conclusions are presented in Section 7.

1. Thesis Starting Point

The starting point of this thesis considers a network provider and a customer that need to establish an SLA, where the availability to be guaranteed during the contract period has to be defined. For this thesis, the most relevant SLA terms are:

- The availability requirement α.
- The price of the service, which may be affected by the availability requirement α (customers are willing to pay more to obtain a higher level of availability).

- The penalty that the provider has to pay when the requirements are not fulfilled.
- The duration of the contract τ.

The main research question addressed by this thesis is: ***How to assess the availability to be stipulated in the SLA, and which resources must be provided in order to guarantee that requirement?***

If this question is addressed properly, at the end of the contract, the customer will receive an stable service to perform successfully all his tasks. In addition, the network operator will obtain a profit that allows him to continue running his business successfully. These two goals can be summarized as a common interest aiming for the success of the SLA, which is in fact the main motivation of this thesis.

The general objective of this thesis is to provide a path that goes form the SLA definition until its successful realization. In order to make such path, the following issues should be solved:

- **How is the probability distribution of the interval availability?** Modeling the availability of a system is usually a complex and demanding task. For this reason, in many studies, it is assessed using expected values. Usually, obtaining the first moment of a random variable is simpler than obtaining the entire probability distribution. In addition, some models assume steady state conditions and Markovian properties, in order to solve the problem easier. However, this thesis considers the fact that SLAs are signed for a finite time interval, which demands the study of the distribution of the interval availability. The fulfillment of the SLA availability requirement means that at the end of the contract, the availability offered lies in a gap between the lowest availability that may be accepted (α) and a maximum possible value of 1 that represent the case of uninterrupted service. Assessing the probability to be inside this gap is a main issue in this thesis. Previous studies addressed this issue by assuming Markovian properties. The analysis made in this thesis showed that real operational networks do not follow such properties. For these reason, one can say that the shape and behavior of the probability distribution of the interval availability is unknown. This thesis will provide information to understand better such distribution in real/operational scenarios.
- **What is the operational behavior of backbone network components?** Backbone networks are composed mainly by routers and links distributed across a geographical area, in order to provide connectivity. Those components are exposed to failures that may arrive at any random time. In addition, the duration of the repair time is also a random variable with huge influence on the availability that the network may offer. For this thesis, having a better knowledge of the stochastic properties that may rule failure and repair processes is very relevant. The better the real operational behavior of network components is modeled, the more accurate the methods developed to assess the risk of failing an SLA. However, the access to this data is limited. Failures of their networks are not what operators like to have exposed in a competitive commercial marketplace.

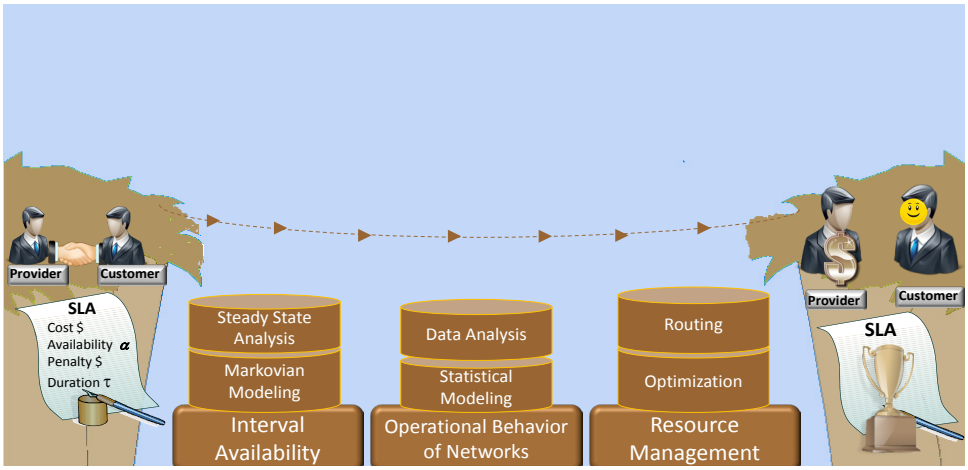


Figure 2. Thesis objective and context.

In this thesis, this information will be measured, analyzed, characterized and published.

- **How should the resources be managed?** This thesis assumes that a customer needs to interconnect locations A and B with a specified capacity, and the network provider is able to provide such connection. The provider may use protection mechanisms, where main and backup resources are reserved in advance, or using restoration mechanism, where there is no pre-planned backup reservation. However, to know exactly which path or set of paths should be assigned to supply this request is a challenging issue that has been studied during many years. It may be solved in many different ways, depending on the specific conditions of the problem. A classical approach is the use of graph theory and routing algorithms, where formal optimization methods, or heuristic techniques may be used. The conditions and requirements of the problem addressed by this thesis have not been considered before. Factors such as the finite duration of the SLA, the influence of the interval availability, and the real operational behavior of network components, make the resource allocation problem hard. In this thesis, new approaches are developed in order to find the networks resources to be assigned.

Figure 2 summarizes the concepts described in this section. It presents the need to build a bridge to interconnect the SLA definition with its successful realization. Based on the issues previously described, it presents an overview of the established concepts that may be used as the initial foundations of the bridge, or the base of the PhD thesis.

Next sections will show how the remaining parts of the bridge presented in Figure 2 are built, using the concepts proposed in this thesis. The finished bridge will be presented in Figure 9 (Section 7.1).

2. Background

This section presents a short background of the problem of guaranteeing contracted availability in connection oriented networks. It describes basic concepts constantly used through the thesis, and hence, a common framework has to be provided. The state of the art will be described in Section 3. Here, the definition and general features of an SLA are presented in Section 2.1. Section 2.2 gives a short overview of issues related to Network Dependability. Then, connection oriented networks which are the kind of networks addressed in this thesis are presented in Section 2.3. Some techniques related to the characterization of stochastic processes of operational networks are described in Section 2.4. Finally, Section 2.5 concludes this part with a discussion on the business issues involved in the addressed SLA context.

2.1 SLA

A Service Level Agreement (SLA) is a business contract that defines the obligations and requirements of providers and customers. When a network provider offers a service, it tries to fulfill the customer needs, being a continuous and uninterrupted operation one of the most common. However, networks are not failure free, and hence having a common understanding of the delivered service is important. When the conditions, responsibilities, guarantees and expectations of a service are clearly specified, uncomfortable confrontations can be avoided, and the utilization of delivered services can be better planned.

The specific details that define an SLA are usually confidential and they may vary depending on the provider, the customers, the service, and the requirements involved. However, the availability to be guaranteed is a common element in every SLA [ITU02a], [ITU00], [ITU02b]. Availability is usually specified in percentage e.g., 99.5%, where the maximum possible value is 100%, representing an uninterrupted service. An important fact in SLAs is that they are defined for a finite time interval (typically specified in months). Therefore, the availability requirement is usually defined in terms of the maximum allowed cumulated downtime during a period.

An important SLA parameter is the penalty that the provider has to pay when the requirements are not fulfilled. The penalty plays a very important roll, due to it encourages the provider to implement the respective adjustments on its network and to put all the needed effort in order to keep the service inside the requirements. At the same time, it gives to the customers the possibility to compensate part of the losses generated by service disconnection.

The SLA penalty may be defined in two different ways [GH12b]. The first possibility is a binary model, where the penalty to be paid is $0 if the delivered availability when the contract finishes is above the availability stipulated in the SLA, or a fixed amount of money if the provider does not fulfill the requirements. However, some business scenarios demand a penalty scheme with a better granularity. Therefore, contracts with increasing penalty may be used by operators in order to fit that demands. In this kind of contracts, after a certain cumulated downtime, the price of the penalty increases. For instance an SLA may specify a penalty free gap for downtimes shorter

that N hours, a refund of 10% of the service price if the cumulated downtime is among N and $N+X$ hours, a refund of 50% of the service price if the cumulated downtime is among $N+X$ and $N+X+Y$ hours, and finally a refund of 150% of the service price if the cumulated downtime is bigger than $N+X+Y$ hours.

Independently of the SLA penalty scheme, the violation of a contract is not desirable. The development of methods that helps to define requirements that can be fulfilled is a fundamental issue.

2.2 Network Dependability

Dependability is an important field in computer science. Initial works in this area started with previous related topics such as fault tolerance and system reliability. A complete dependability framework, including basic concepts and taxonomy can be found in [ALRL04]. They define dependability as "*the ability of a system to deliver service that can be justifiably trusted*". An alternative definition mentions that dependability is "*the ability to avoid service failures that are more frequent and more severe than is acceptable*"

Dependability is a high-level concept composed by five attributes: (1) Availability: Ability of the system to provide a set of services at a given instant of time or at any instant within a given time interval. (2) Reliability: Ability of a system to provide uninterrupted service. (3) Safety: Ability of a system to provide service without the occurrence of catastrophic failures. (4) Integrity: Absence of improper system alterations. (5) Maintainability: Ability to undergo modifications and repairs.

For practical reasons, availability is the most common attribute used when an SLA is defined. ITU-T [ITU94] defines the availability of a system as "*the ability to be in a state to perform a required function at a given instant of time or at any instant of time within a given time interval*".

For the assessment of the availability of a network connection the following concepts have to be defined:

- Downtime: Time period in which a connection is not working due to a network failure. Given that a connection may have several downtimes, the use of the average downtime or Mean Downtime (MDT) is very common.
- Uptime: Time period when a connection is delivering its service properly. As explained for the downtime, the use of Mean Uptime (MUT) is very common.
- Interval Availability ($\hat{A}(\tau)$): Is a dependability measure that represents the fraction of time during which a system delivered successfully a given service over an specified interval τ.
- Asymptotic Availability (A): Probability that the network is able to deliver a service (according to some requirements) at some point in time in the future when the network is in steady state. It can be defined as

$$A = \frac{MUT}{MUT + MDT}$$

- Asymptotic Unavailability (U): This parameter is a complement of the asymptotic availability. It may be defined as $U = 1 - A$.

The resilience of a network may be defined as the ability of a network to automatically react to failures through the use of alternative failure-free paths. Planning redundancies and make use of them when a failure affect the service is one of the keys to provide dependable services. Protection is one of the most common strategies used to meet dependability requirements with a high probability. In this approach the network resources that provide connectivity between two points are planned in advance, when the connection request arrives, and before any failure affect it. For this, the use of predefined backups is a common policy. Depending on the characteristics of the network elements and on the implemented backup scheme, different dependability degrees may be obtained.

According to the protection scheme used, a network connection can be classified as unprotected, dedicated, or shared backup protected [MH08b]. For unprotected connections only a single path is assigned to the connection. The path is formed by a series of interconnected links, where a failure in any of them affects the connectivity of the whole path. This technique can only be used in scenarios where the components of the path are ultra-reliable, or in connections with not very high availability requirements. Dedicated path protection is a scheme where the capacity used by the working and the backup path is reserved exclusively for one connection. A network connection that uses dedicated backup goes down when any of the links of the working path goes down and the backup path is not available due to another failure already affected it. Finally, in Shared Backup Path Protection (SBPP) a connection may share bandwidth on its backup path with other connections. In this case, the offered availability not only depends on the network resources assigned to one connection, but also on the state of other connections. In SBPP, the working and backup paths have to be disjoint and the Shared Risk Link Group (SRLG) rule which states that the connections affected by one failure cannot share any backup resource [RBS+01] has to be considered.

A complete framework on network dependability and the details of the different protection schemes used in telecommunication networks may be found in [Gro04] and [VPD04].

2.3 Connection Oriented Networks

Networks where the end-to-end connections are pre-established before sending any data are known as Connection Oriented Networks. In [ITU01], the definition and main features of this kind of networks are mentioned. In addition, [Tan03] explains that the idea behind the pre-established paths is to avoid having to choose a new route for every packet sent. In this way, the routing processes are simplified and the transmission is faster.

To know in advance the path to be used offers additional advantages when SLAs with availability requirements are specified. For instance, given that the path assigned to a connection is known, the end-to-end availability offered to that connection can be assessed. To do this, the network provider has to characterize the dependability

behavior of the elements that compound the network. If a single path does not fulfill the availability requirement, redundant paths can be added, in order to increase the availability to the desired value. Connection oriented networks allow a controlled planning of the resources to be assigned, the assessment of the availability to be offered, and thereby the definition of SLA requirements that can be met with a high probability.

MPLS networks are nowadays a common example of connection oriented networks. It started as an IP-switching project to combine the routing advantages of the layer 3 of the OSI model, with the speed of switching and layer 2 techniques [Sol03]. The process of sending a packet through the MPLS network is the following: First, the Forwarding Equivalence Class (FEC) is defined. The FEC is a group of packets with similar characteristics (IP address source, IP address destination, port number, etc). Second, a path between the source and the destination is found using routing algorithms. This path is identified on each router through the use of one specific label. Finally, the first router on the MPLS domain labels each incoming packet, and sends it through the predefined path. MPLS has become one of the most preferred technologies for network operators due to its traffic engineering facilities. It allows the visualization of end-to-end paths instead of the hop by hop vision offered by other technologies [EA02], and the implementation of path protection using the fast rerouting mechanism [LBCG02].

2.4 Characterization of Failure and Repair data from Operational Networks

Analysis of real failure processes form operational networks is compulsory in order to get the appropriate information for availability dimensioning, and to deal with the risk of failing the SLA. An important part of the characterization of operational networks is to fit the observed failure and repair processes of each of the network components with well known distributions, in order to have information that may be replicated and used as input in future studies.

A recommended practice to make data characterization is the use of visual tools such as Q-Q plots to compare the sampled data with well known distribution. Through the use of Q-Q plots a fair graphical analysis may be performed. In this technique, the pattern of points in the plot is used to compare two distributions, the one obtained from measurements containing empirical data, and the other representing a theoretical parameterized distribution. An indication that the empirical data fits the estimated distribution may be assumed if there is a similar pattern between the two plotted distributions.

However, alternative techniques are needed in order to get more information from the raw data. The use of the maximum likelihood estimation method is common in order to estimate parameters that may fit a hypothesized theoretical distribution with respective confidence bounds. This technique evaluates if the empirical CDF does not lie beyond the limits of the confidence bounds to verify if the tentative CDF may fit the empirical data. The parameters obtained from the maximum likelihood estimation characterize the hypothesized distribution and may be tested through the

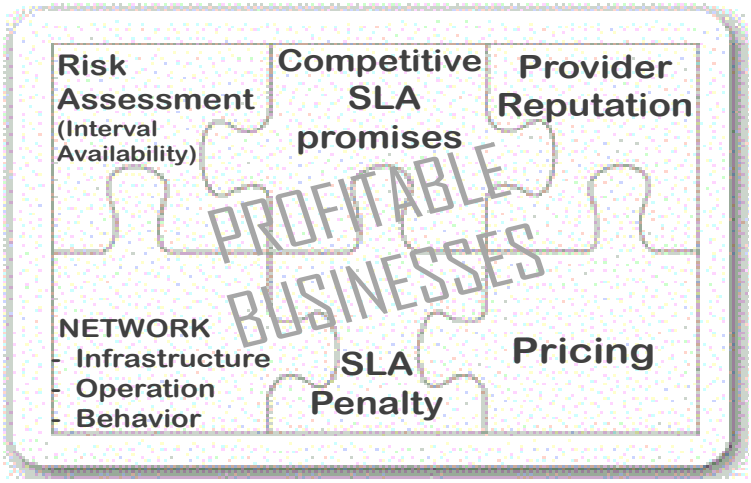


Figure 3. Profitable Network Business in an SLA context.

use of well known goodness-of-fit test i.e. Kolmogorov-Smirnov, Camer-von Mises and Anderson Darling. In [WVMD84] are analyzed the properties of those tests for Gamma distributions. Additionally in [SL99] the same tests are analyzed for Weibull distributions.

Given that this kind of procedures are nowadays widely needed for the scientific community, the NIST in cooperation with other institutions have developed a handbook [NIS11], where all these issues are explained in detail. In addition, they offer software tools that may be used to analyze empirical data.

2.5 SLA in a Business Context

Network operators always try to turn their operations into profitable businesses. Like in every business, there are important variables such as income, investment, revenue, expenses and a final profit.

The work developed in [Med10] presents a global overview of the Internet QoS puzzle, where standards, regulation, SLAs, pricing and marketing considerations are discussed. This section takes inspiration from that work, presenting in Figure 3 the most important aspects that influence the possibility of having a profitable business, focussed on the context of this thesis.

The first variable to be considered is the network itself. It is composed by equipment that is purchased at some cost. The network infrastructure has to be deployed according to potential demands and needs. The higher the customer availability requirements, the more reliable the resources to be assigned, which most likely implies a higher cost [FDL00]. However, this is not the only expense related to network infrastructure. In addition, operational and maintenance costs are part of the everyday life of a provider.

The availability stipulated in the SLA may have a considerable influence on the price that customers pay for the received services, and also in the perception of the provider reputation [Das00]. Telecommunications is a business where competitors try to attract customers offering high quality products, and hence, α has to be competitive. On the other hand, the violation of a contract impacts considerably the reputation of a provider. Therefore, the availability stipulated in the SLA has to be commercially competitive, but at the same time according to the dependability features of the network infrastructure.

In the risk assessment part, the stochastic behaviors of the paths assigned to customers are studied. Through an appropriate risk assessment, one can evaluate if the resources used by the connections are enough in order to fit the requirements signed in the SLA.

Section 2.1 argues that penalties encourage providers to put the necessary effort to deliver the service within the agreed values, and they give to customers the possibility to compensate part of the losses that a potential disconnections may bring. The SLA penalty is triggered by violation of the SLA requirements, and as a basic business rule, it has to be avoided.

As explained before, the availability stipulated in an SLA has to be based on a responsible assessment, due to any violation on the agreement hits directly the provider reputation. In addition, network providers have to understand that any failure before breaking the SLA availability requirement affects also their reputation and may impact their finances [GH12b]. The cost related to reputation may be assessed via marketing models, and it has to be considered if network providers want to have a complete picture of the problem.

Finally, the price of the service offered is the main source of income for an operator, and hence one of the most important variables to make profitable business. The price of a service can be obtained for the case of external customers directly via their payments. For the case of internal uses, it could be calculated by estimating the importance and the value of the services that run over the network. Pricing involves some task out of the scope of this thesis. However, there are some other tasks directly related with the addressed problem, e.g., the price has to be estimated based on compensating the infrastructure expenses, and in order to prevent that the cost of penalties make the business unprofitable.

This section shows that all the six pieces of the puzzle are interconnected and depend on each other. The current way to deal with SLAs in a business context has two weaknesses that make the solution of the proposed puzzle difficult. First, the companies analyze profitable business scenarios giving much more priority to marketing models than to the technical details behind, such as the SLA and the interval availability [Med10]. On the other hand, when a technical assessment is made, only few pieces are considered. The big challenge when profitable business in an SLA context are addressed is the ability to have models that consider simultaneously the six pieces presented in Figure 3. This thesis offers tools that help to solve the presented puzzle.

3. State of the Art and Open Issues in Guaranteeing SLA Availability Requirements

In spite of being a very relevant issue, to date, the task of guaranteeing availability according to SLA requirements is far from being solved. Network operators use modeling techniques to assess the availability of a network connection, making assumptions such as independence, steady state, or Markovian properties. The results obtained through those techniques may differ from real/operational situations, and they may not allow a proper assessment of the probability of failing the SLA. The resources provided to a network connection according to the SLA availability requirements may be handled in two ways [WZY04], [CMH+07]: i) If the customer is considered as "*very important*", or if the penalties stipulated in the SLA are very drastic, networks operators usually over-dimension the infrastructure provided to the customer, and in this way, they can react easily to any undesired event. However, this approach is unprofitable and not scalable. ii) When the customer has low priority or the penalty associated with the availability violation is week, operators usually do not put much effort in the "*in-advance*" planning, but they use restoration mechanisms that react after a failure event, without any guarantee.

Previous studies have provided tools to assess the probability of failing the SLA using protection mechanisms. The first study that formally defined the concept of SLA-risk (probability of failing the SLA) and SLA success probability $S(\tau,t)$ was made by Goyal and Tantawi in [GT88]. They realize that the probability of not meeting the availability stipulated in an SLA depends on the duration of the contract and it has two possible behaviors as Figure 4 illustrates. If the availability requirement α is larger than the *steady state availability* A, the probability of meeting the requirements decreases continuously with the observation period until it quickly reaches a value close to 0. However, they also showed that if α is smaller than the *steady state availability* A, the probability of meeting the SLA availability requirement starts with a value close to one for short SLA durations, then, it decreases until certain value between one and zero for intermediate SLA durations, and finally, it converges back to one for long observation periods (SLA durations). Their study was based on the assumption of Markovian properties. Figure 4 shows the behavior of $S(\tau,t)$ in a Markovian system with $A = 0.9972$. In the first case ($\alpha < A$) the value of α is 0.995. In the second case ($\alpha > A$) the value of α is 0.999.

With the issues raised by Goyal and Tantawi, and the increasing use of SLAs in telecommunications business, further works have tried to provide tools to make an appropriate SLA risk assessment. For instance, Waldman and Mello have developed a series of studies in [MSW05], [MQWS06], [WM08] and [WM09], in order to provide SLA-aware network connections. Assuming Markovian properties and a maximum of two link failures, they developed a matrix-based analytical approach to assess the risk under shared and dedicated protection schemes.

Empirical studies have shown that telecommunications networks usually do not have the Markovian properties. Therefore, new analyzes have to be made in order to include this fact. In this direction, Snow and Weckman in [SW07] and [SWG10], and

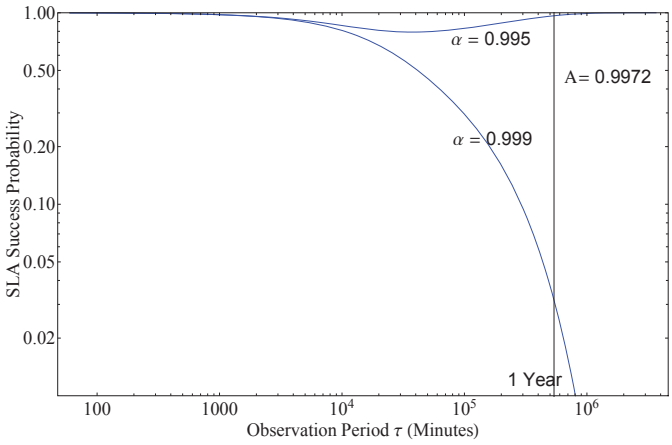


Figure 4. Behavior of the SLA success probability with the observation period τ in two cases: i) $\alpha < A$. ii) $\alpha > A$.

Mastroeni and Naldi in [MN11] have studied the probability of not meeting the SLA availability requirement, assuming the network connection as a single component ruled by a single on/off process, where the duration of up and down times may be Weibull or lognormal distributed, with different types of tails.

Finally, the work presented in [XTMM11] highlights the importance of assessing the SLA risk in WDM mesh networks, and the dangers implied by dealing only with steady state probabilities. They developed a method to assess the SLA risk when the stochastic properties of the networks are partially known. However, they assume independence between failure processes, Poison failure arrivals, and they do not consider the existence of overlapping failures.

This thesis identifies the following open issues that need to be addressed.

1 Network connections are not single components, but are compound systems that depend on the behavior of multiple processes. In addition, some of these processes may present dependencies and their duration may not follow exponential distributions, making their modeling very challenging. The behavior of the interval availability in compound systems under the existence of correlation and non exponential time distributions is unknown.

2 Having data that represents the dependability behavior of the network components is fundamental in order to evaluate the behavior of the interval availability. However, this information is not available. This thesis addresses this issue by developing from the very beginning the whole process needed in order to obtain such information.

3 The tasks that should be done in order to guarantee availability in network connections may be classified in two: Resource assignment and risk assessment on the assigned resources. The assignment of resources to a connection request

is usually addressed via routing or optimization techniques. However, how to make such assignment considering SLA interval availability constraints is still an open challenge.

4 Previous works on SLA risk have study the problem as a binary system, when the SLA is fulfilled or not. However, real SLAs are not binary. Every single failure before the SLA violation affects the reputation of the provider. In addition, SLAs usually use a continuous penalty scheme, where the refund price increases with the cumulated downtime. This non-binary view of the SLA, where reputation and cumulated downtime are considered is addressed by this thesis. It allows the implementation of methods not considered before; where the provider may estimate better his profit by using a model that captures closer the real features of the problem.

The following sections will discuss more in detail specific open issues in guaranteeing the SLA availability requirement as follows: First, assuming that the resources of a network connection are known, Section 3.1 presents the challenge of assessing the distribution of the interval availability. Then, the challenge of obtaining dependability features from operational networks is studied in Section 3.2. Finally, the problem of allocating connections in a telecommunication network is discussed in Section 3.3.

3.1 Modeling the Interval Availability

Traditionally, the assessment of the availability of a system, and its impact in an SLA may be made through the evaluation of one of the four different values.

- Steady State availability ($A = E[\hat{A}(\infty)]$).
- Expected Interval Availability ($E[\hat{A}(\tau)]$).
- PDF of the Interval Availability ($f_{\hat{A}(\tau)}(a)$).
- SLA Risk ($P[\hat{A}(\tau) \leq \alpha]$).

Each of these values is different. They offer a different description of the system, and obtaining them demands the use of different techniques, where the complexity may differ considerably.

The study of the steady state availability is the most common, and hence, the richer number of results and techniques can be found in this direction. There are complete results using Markovian models on single and compound systems using structural models [Sho76] and [BAK04]. One can say that steady state analysis is the simplest option for assessing the availability of a system, and thereby, the probability of failing the SLA. In spite of being the simplest, it is usually not an easy task. In fact, it may become very complex (see for instance [CH01], [MdSeSG89] and [KKM03]).

A very important characteristic of SLAs is that they are defined for a finite time interval τ. Therefore, asymptotic results are inappropriate. A possible alternative is the evaluation of the expected interval availability. However, this task is more demanding

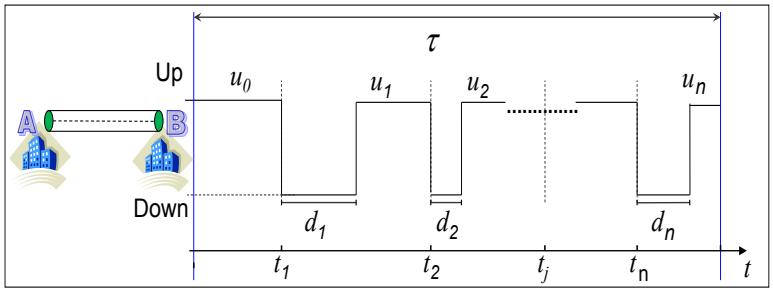


Figure 5. Behavior of an end-to-end network connection during an interval τ.

than the evaluation of the steady state availability and it does not offer much additional information. $E[\hat{A}(\tau)]$ just differs considerably from $E[\hat{A}(\infty)]$ for short observation periods. One can say that for the standard duration of an SLA, those two values are almost the same.

Considering the whole distribution of the interval availability yields a better SLA-risk assessment. According to [ITU94], the interval availability is the fraction of time in a specified interval when the system is able to deliver a given service. It has a rich stochastic behavior that to date is just partially known under Markovian assumptions, but it remains unknown when real/operational conditions are considered.

Figure 5 shows the general behavior of an end-to-end network connection. It considers failure and repair events on a connection that has an operational state which may be described by an on-off process. Failure j ($j = 1,2,....,n$) occurs at time t_j, where the downtime duration is denoted by d_j and n is the total number of failures during a time interval τ. After a repair event, the time when the connection is working properly before a new failure occurs will be defined as uptime and will be denoted as u_j i.e. $u_j = t_{j+1} - t_j - d_j$. In the depicted scenario, the observed interval availability $\hat{A}(\tau)$ may be calculated as follows:

$$\hat{A}(\tau) = \frac{\sum_{j=0}^{n} u_j}{\tau}. \tag{1}$$

In principle, expression (1) computes the interval availability after one observation period. However, assessing in advance this value is very complex due to: i) d_j and u_j are stochastic variables that may have distributions difficult to model. ii) The total number of failures (n) depends on the duration of u_j and d_j. For instance, if in a given scenario $u_0 > \tau$, the number of failures is zero and the interval availability equal to one. iii) The end-to-end connection shown in Figure 5 is composed by several routers and links. Therefore, its behavior depends on the individual state of multiple components.

Figure 6 illustrates the general behavior of the *distribution of the interval availability* based on results presented in [GH12a] and [Tak57]. It is defined as $f_{\hat{A}(\tau)}(a)$, and it may be used to calculate the SLA risk and success.

The *SLA Success* Probability is defined as:

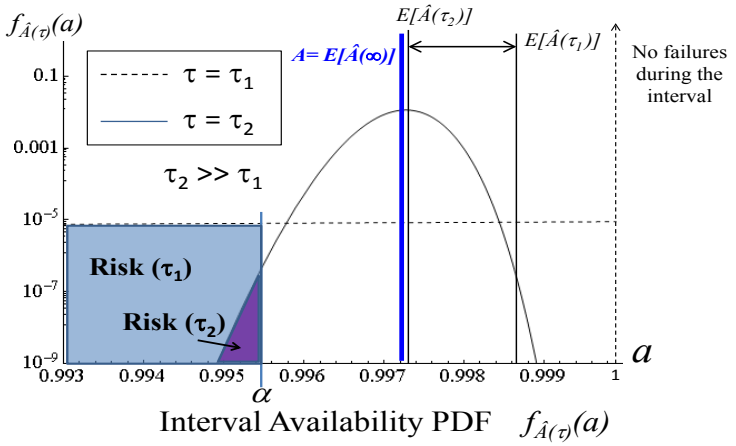


Figure 6. PDF of the Interval Availability.

$$S(\tau,\alpha) \;=\; P[\hat{A}(\tau) \geq \alpha] \;=\; \int_{\alpha}^{1} f_{\hat{A}(\tau)}(a)\,da. \tag{2}$$

The *SLA risk* is defined as the probability that the specified availability α will not be met, i.e., $1 - S(\tau,\alpha)$.

In addition, Figure 6 illustrates why the SLA risk may be considerable even if $E[\hat{A}(\tau)]$ is larger than the SLA requirement α.

$f_{\hat{A}(\tau)}(a)$ changes drastically with the duration of τ. Therefore, Figure 6 considers three different intervals (τ_1, τ_2 and ∞), where $\tau_1 \ll \tau_2 \ll \infty$, and it can be used to illustrate the differences between $E[\hat{A}(\infty)]$, $E[\hat{A}(\tau)]$, $f_{\hat{A}(\tau)}(a)$ and $S(\tau,\alpha)$, introduced at the beginning of this section.

Specific details of the interval availability distribution under specific scenarios will be widely discussed in the papers presented in the second part of this thesis. In this section, we will focus on the general challenges of modeling such distribution.

Network connections are compound systems. Therefore, an important issue, when their interval availability is assessed, is to define how detailed will be modeled the effects of the specific parts that compound the systems. In this section, we will discuss two of the most common approaches in the literature. First, a common simplification is to assume the system as a single global component, where only two global states are possible (operational and not operational). A second approach, considers in more detail each of the individual parts of the system, all the possible states, depending on the combination of the state of each individual component, and all the transitions between such states.

Regarding the first approach (single global component), the most relevant findings may be summarized as follows:

The accumulated down time $t(\tau)$ during the SLA contract period τ is associated with $\hat{A}(\tau)$ as: $t(\tau) \;=\; \tau[1-\hat{A}(\tau)]$. Here, $\Omega(\tau,t)$ is defined as the CDF of $t(\tau)$. A general expression for $\Omega(\tau,t)$ was derived by Takács in [Tak57] as follows

$$\Omega(\tau,t) = \sum_{n=0}^{\infty} D_n(t)[U_n(\tau-t) - U_{n+1}(\tau-t)] \tag{3}$$

where the failure and repair processes are described by i.i.d. up and down times with CDF $U(t)$ and $D(t)$ respectively, and the subindex n represents the $n-$fold Stieltjes convolution of a given function.

Equation (3) characterizes the problem with general distributions. However, it is difficult to compute for specific failure and repair processes due to the complexity posed by the $n-$fold convolution of generally distributed CDFs.

A complete result was obtained by Takács when the failure and repair processes are exponentially distributed as follows

$$U(t) = 1 - e^{-\lambda t}$$

$$D(t) = 1 - e^{-\mu t}$$

$$\Omega(\tau,t) = e^{-\lambda(\tau-t)}\left[1 + (\lambda\mu(\tau-t))^{\frac{1}{2}} \int_0^{\infty} e^{-\mu y} y^{\frac{1}{2}} I_1(2(\lambda\mu(\tau-t)y)^{\frac{1}{2}})dy\right] \tag{4}$$

where I_1 is the Bessel function of order 1.

In [FY94], $\Omega(\tau,t)$ is approximated for general distributions assuming short intervals as

$$\Omega(\tau,t) \approx 1 - [U(\tau-t)[1-D(t)]] \tag{5}$$

Regarding the second approach (compound system), the most relevant findings may be summarized as follows:

The interval availability of a compound system may be obtained by numerical methods using uniformization techniques as is presented in [GT88], [SEG86] and [RS95]. These approaches assume continuous-time homogenous Markov processes, where *the states of the system are divided in two subsets: the operational states and the unoperational states*, and through the use of the uniformized Markov chain the distribution of the interval availability may be assessed. These works represent an important advance in the computation of the interval availability in compound-Markovian systems, and show how challenging can be to obtain such solution.

As a conclusion, one can say that modeling accurately the transient solution of repairable compound-systems with failure and repair processes non-independent and non-exponentially distributed is still an open issue.

3.2 Dependability Features of Operational Networks

The analysis of real failure processes is compulsory in order to get the appropriate information for availability dimensioning and to deal with the risks of failing the SLA.

In spite of this, for a number of reasons, among them that failures of their network are not what operators like to have exposed in a competitive commercial marketplace, the access to such failure log information is very limited. Few studies based on operational data are published. In [CSKM07] a study of spatial and temporal failures and outages in an access network was performed to assess availability. Another study in [MVM02] estimates the time between failures and times to repair for elements in a large wireless access network, finding that they are not consistent with exponential distributions, but they may better be described by Weibull or two-stage hyper-exponential distributions. A study of the failure behavior in an operational backbone network is reported by Iannaccone et al. in [InCM$^+$02]. They examine the frequency and duration of failure events and discuss various statistics, for instance the distribution of inter-failure times and distribution of link failure durations. Nevertheless, due to confidentiality reasons the published values were normalized. This work was continued by Markopoulou et al. in [MIB$^+$08], where failure and repair events in the Sprint IP backbone Network are classified and analyzed. They perform a characterization of the different classes of failures, presenting also normalized values. In [KNR09], Kuusela and Norros analyze router failure logs from the Finnish academic network, FUNET.

In summary, for researchers without any direct relation with a telecommunication company, the dependability features of backbone operational networks are simply not available. Nevertheless, assuming that this challenge is overcome and the operational data of the network can be obtained, there are still a number of remaining challenges.

Failure logs are usually obtained using network management protocols such as SNMP. From this raw information the idea is to obtain more elaborated data that can be used in the assessment. For instance, basic values like: links and routers mean time to failure, mean time to repair or more detailed information such as the entire on/off behavior during a certain interval. The objective is to obtain filtered information from the raw data. However, the execution of these tasks demands effort, time, and programming skills.

If the on/off behavior of the network components is known, the next step is to develop a characterization process. Here, more sophisticated information such as the probability distribution of up and down times or the correlation between failure processes is obtained. For this, more advanced techniques such as distribution fitting and goodness of fit tests are needed.

Finally, the data obtained from the characterization process may be used in order to identify common patters and behaviors of the dependability features of the network components. To find a physical and logical explanation of the patterns and behaviors observed is a duty that demands multiple analysis and creativity.

3.3 Connection Allocation

A network is a collection of interconnected components to communicate several points. The network topology defines the way that those components are interconnected, offering a rich amount of connection possibilities. Networks are deployed and managed by operators who have control over the resources. Their aim is to get economical and strategic benefit from the infrastructure installed. Network providers

recurrently receive requests from customers that demand connectivity between two points, where some specific requirements are defined. The most common characteristics of a request are: source, destination, bandwidth demand, and the availability to be guaranteed. In this context, the question to be solved is: which path or group of paths should be assigned to the connection?

In conventional graph theory, this is a challenging question that may be solved in several ways [Gro04]. One of the most important solutions is the routing algorithm proposed by Dijkstra in [Dij59], which offers the best balance between simplicity and efficiency.

In some situations the simplicity offered by Dijkstra's algorithm may not be enough to satisfy the provider/customer needs. This algorithm is able to find a single shortest cost path. However, if the cost of the links cannot be added linearly of if more than one path have to be provided at the same time, alternative procedures have to be used. A common example of this is the multi commodity flow problem [KV06], where multiple request have to be allocated in a given network.

The problem of allocating multiple connections in a network, fulfilling bandwidth requirements at a minimum cost, has been addressed through the use of Integer Linear Programming (ILP). Formal optimization techniques have been successfully applied to allocate connections without protection and connection with dedicated path protection [ZZZ+07]. When specific availability requirements have to be fulfilled, the problem cannot be efficiently solved, due to non linear constraints are generated [LLY09]. An interesting proposal is found in [KL00], where by solving an ILP, a working and a shared backup path are obtained. However, this solution uses Shared Risk Link Group (SRLG) as the only availability constraint, and it cannot be used to fulfill specific availability goals. Finally, in [CSK02] and [LML10] are proposed two different ILP formulations in order to find diverse routes with minimum joint failure probability.

Two open challenges regarding the use of optimization techniques for finding the best resource allocation possible are: i) Previous works do not consider the distribution of the interval availability, but they assume steady state availability values. ii) Previous works do not target directly company revenues, neither the non-binary features of the penalty scheme.

4. Research Goals

At the highest level, this thesis deals with guaranteeing availability in SLAs. This specific target involves three research goals that will be explained in this section. Implementing new policies to meet the SLA availability requirements is an intuitive goal as a consequence of the main target. The second research goal is designing new methods to assess the distribution of interval availability, due to it is an open issue that needs to be addressed. The third research goal is to provide information about the behavior of failure and repair processes of operational networks for the validation of the new methods and policies proposed, and for its potential use in future works.

4.1 New Policies to Meet the SLA Availability Requirements

When a connection request is made, the provider has the freedom to allocate it in his network in the way that he wants. In addition, the fault management tasks are also actions up to the provider. What really matters here is that at the end of the contract, the SLA requirements will be fulfilled. For this reason, network providers should follow specific policies that help them to avoid the SLA violation. One of the research goals of this thesis is to provide SLA oriented policies for: i) The allocation of service requests. ii) The adequate use of protection schemes. iii) Fault management.

4.2 New Methods for the Assessment of the SLA Risk and the Interval Availability

The distribution of the interval availability of a network connection under real operational conditions is unknown. Therefore, new methods are needed to obtain an improved understanding of the distribution of the interval availability. Mathematical methods provide general information on how to obtain the interval availability. However, specific results are difficult to compute. The goal is to propose simplifications to obtain concrete results, without losing accuracy in the estimation. On the other hand, simulations methods are more flexible and allow the consideration of a wider set of conditions. However, the way to model and implement such conditions requires the development of new methods.

4.3 Provide Information about the Behavior of Operational Networks

The data that describes the dependability behavior of operational networks was a fundamental need at the beginning of this thesis, in order to validate the models that will be developed. Due to this information was not available; to obtain it becomes one of the thesis goals. The targets are: i) Validating the thesis results using operational data. ii) Justify some of the assumptions taken during the thesis. iii) Publishing real based data that can be used by any future related work.

5. Research Methodology

This section describes the methodology used to solve the described research problems. It includes the methods used and how they were evaluated and validated. The research effort started with an extensive literature review. Topics related to network dependability with an special focus on the task of guaranteeing availability in SLAs were studied. This initial step was fundamental for defining the state of the art, and to highlight the challenges that this thesis will address. This section will describe the methodology followed in order to find: i) The features of the interval availability. ii) The dependability behavior of operational networks. iii) The methods needed to use network resources efficiently, fulfilling SLA availability requirements.

The rich stochastic behavior on the interval availability, and its direct impact on the SLA risk, make its accurate assessment a matter of prime interest. There are two traditional options to evaluate the interval availability: simulation and mathematical modeling. In this thesis, we propose the use of mathematical methods for the assessment of the probability of failing the SLA availability requirement in a single-component system. The challenge of considering non-Markovian properties was addressed by proposing and approximation that makes the expression of the interval availability mathematically tractable. The challenge now is to show that this approximation fits closely the characteristics of the real system. Two different approaches were used to test the accuracy of the proposed approximation. First, given that there is an expression to model the interval availability in single-component Markovian systems, a direct comparison of that expression with the approximation proposed was made. Second, for the case of non Markovian systems, a Monte Carlo simulation that allows the assessment of the interval availability with negligible error was used for comparison.

The use of mathematical methods for the analysis of compound systems under different protection schemes, where correlation may exist, represent a demanding task due to the number of variables involved. A compound system may have a huge number of components that may make the mathematical formulations long and unpractical. Therefore, the assessment of the interval availability of these kinds of systems is made via simulation. The simulation of the behavior of a network requires a detailed evaluation of the characteristics to be considered in order to have trustable and relevant results. A methodology to simulate networks and the difficulties that it brings is presented in [FP01].

The simulations performed in this thesis are focused in assessing the probability distribution of the interval availability of an end-to-end connection. For this, Discrete Event Simulations were implemented in DEMOS [Bir03]. These simulations consider: the stochastic distribution of up/down times, different protection mechanisms, a model able to emulate the correlation values observed, and the failure handling by using the protection schemes studied. The execution time of the developed simulations allows the implementation of large number of replications (sample size) in reasonable time. Therefore, for the studies made during the thesis, the sample size was set large enough, in order to estimate with very narrow confidence bands the SLA risk. Finally, the simulation results were analyzed using software such as Matlab or Mathematica, where several properties were evaluated, e.g., expected up and down times, confidence bands, empirical probability density function, cumulative density function, hazard function, survival function, etc.

The lack of information about the dependability behavior of operational networks was addressed by taking advantage of the professional cooperation that NTNU has with the network provider UNINETT. It has a network that connects universities, colleges and research institutions in Norway, and it is a nonprofit organization that provides service to several hundred thousand users, carrying many critical applications. These conditions give the opportunity to obtain data form a network, where dependability is a fundamental issue, without commercial restrictions. The informa-

tion provided by UNINETT contains summaries of events based on raw SNMP data without any previous processing. A set of PERL scripts [PER] were implemented to obtain clean *on-off* processes for each network component, that were verified using alternatives means such as traffic logs and Customer Relationship Management (CRM) information.

Using the filtered information, the next step is to obtain the tentative characterized distribution of up/down times on each of the network components. Given that a good characterization leads to a trustable risk assessment, in this thesis, several techniques were used and compared in order to get as reliable as possible the features of the network processes. Initially, visual tools such as Q-Q plots were used to compare the sampled data with well known parameterized distribution. In addition, through the method of maximum likelihood estimation (MLE) the parameters that may fit a hypothesized theoretical distributions were estimated. Finally, the information obtained from the MLE was tested through the use of well known goodness-of-fit test i.e. Kolmogorov-Smirnov, Camer-von Mises and Anderson Darling.

The obtained *on-off* processes were also used to evaluate correlation. Here, several techniques were also used to get as close as possible information of the correlation between network processes. First, through the use of numerical methods was probed that the measured operational failure processes are neither renewal nor independent, using the Palm-Khintchine theorem. Visual methods such as scatter plots were also used in order to probe correlation between failure processes. Finally, a method to evaluate the correlation coefficient of two failure processes was proposed.

The efficient use of network resources, fulfilling SLA availability requirements is a challenging task. One of the best way to solve this problem is through the use optimization techniques, due to there is a guarantee that the solution obtained is the best possible. For this reason, this thesis uses optimization mechanisms, either to solve the connection allocation problem, or as a reference in the design of the heuristic procedures proposed. This thesis addresses the difficulty posed by the nonlinearities constraints using two methods. First, an expression that approximates the steady state availability of a network connection with different protection schemes is used in an heuristic algorithm. Such algorithm performs a systematic selection of the connections that may share resources, without violating any availability constraint. Second, due to a solution that considers the whole distribution of the interval availability is still needed, a novel solution was proposed based on the following concepts: i) The use of stochastic optimization methods. ii) The proposal of a novel method able to approximate in a linear expression the cumulated downtime of and end-to-end connection.

The topologies used in the case studies presented in the thesis include: First, typical networks commonly used in similar works as is presented in [SND] and [OPTW07]. Second, the UNINETT network topology [UNI12b].

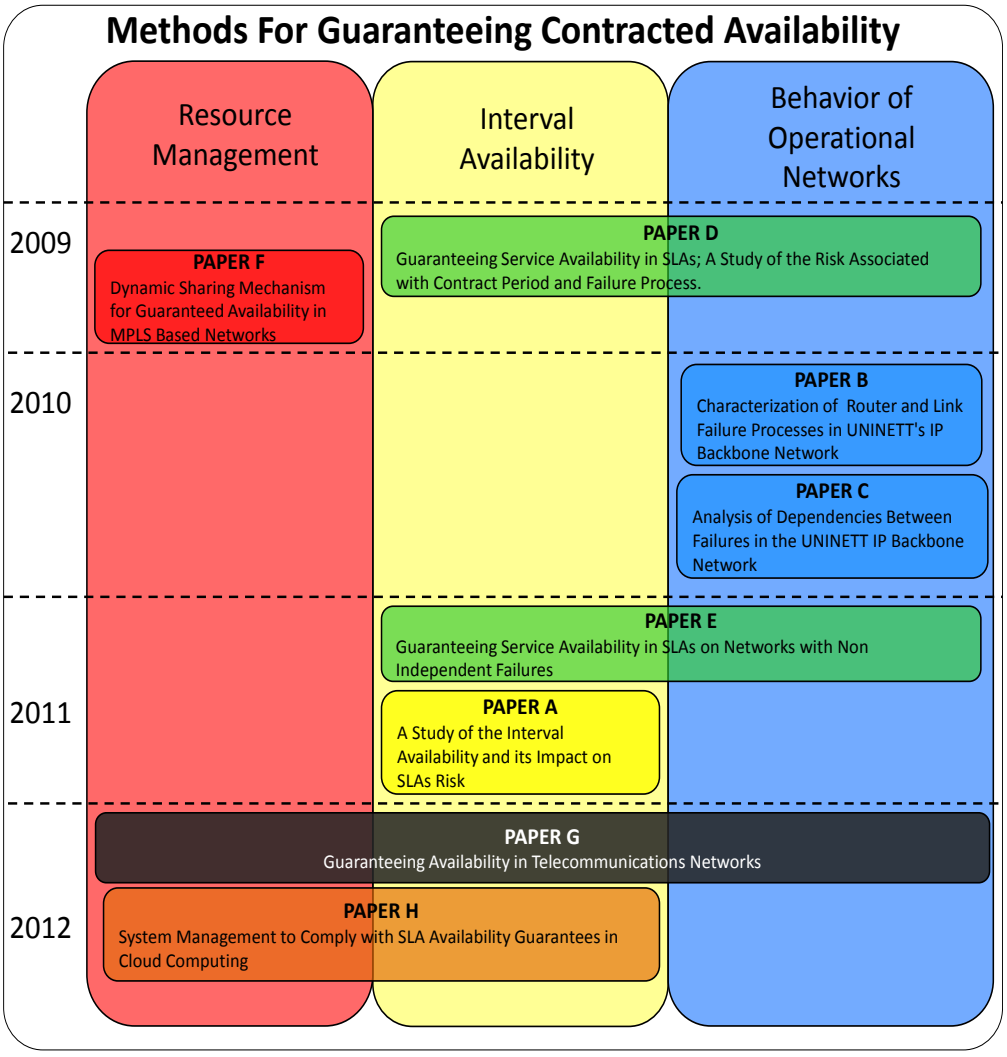


Figure 7. Overview of the written papers. Chronological framework and addressed topics.

6. Contributions

This section presents the main contributions of each of the eight papers included in the second Part II of the thesis. Those contributions have been published in international peer-reviewed conference proceedings and journals.

Figure 7 presents a chronological framework of the papers, and it shows their relation with the three *pillars* described in Section 1 (See Figure 2).

At the beginning of the thesis, the three pillars were addressed in paper D and F. Paper D analyzes the SLA risk in connections allocated in networks with non-Markovian failure/repair processes. Paper F designs mechanism to allocate con-

nections using an smart sharing scheme, and considering steady state availability constraints.

In this first stage, the parameters that characterize the behavior of the network components were assumed. As was explained in Section 3.2, many parameters are not available due to commercial restrictions. For this reason, in the next stage of the thesis, a study of the features of the UNINETT operational networks was performed. In paper B and C this results are summarized. Paper B focuses on the characterization of failure and repair processes in routers and links. Paper C studies the correlation and independence properties of such processes.

The last stage of the thesis tries to cover again the three pillars, with the advantage of working with the operational information obtained from paper B and C. Paper E studies the SLA risk and the distribution of the interval availability assuming non independent failures and realistic failure/repair distributions. Paper A develops a numerical method to compute the interval availability based on the distributions obtained in the characterization developed in paper B. In paper G, all the three pillars are articulated by using an stochastic optimization formulation.

Finally, all the knowledge acquired during the PhD was applied on Cloud Computing, which is one of the most growing telecommunication technologies to date. The results of this study are presented in paper H.

In all the papers, the thesis author had the original idea in cooperation with professor Helvik. However, for Paper B and C the cooperation with UNINETT was fundamental. UNINETT provided the log of failure and repair events in its backbone network for the period 2001-2009. In addition, in Paper C, interesting inputs and suggestions were received from Jon K. Hellan and Pirkko Kuusela from UNINETT and VTT Technical Research Center of Finland, respectively. The work reported in paper H was funded by Telenor and the Department of Telematics of the Norwegian University of Science and Technology.

6.1 Contributions of the papers

This section describes in detail the main contributions of the papers presented in the second part of the thesis.

PAPER A

A Study of the Interval Availability and its Impact on SLAs Risk
ICCSEA, SPRINGER Proceedings, 2012

The interval availability is commonly assessed via simulation or numerical methods under markovian assumptions. This paper complements previous works, proposing a numerical method to approximate the distribution of the cumulative downtime during a given interval, when up and down times are gamma or Weibull distributed. The numerical method applies for single-component system, and uses counting models and feasible Laplace transforms. The proposed approximation is very accurate in

a backbone network context, where the components have a steady state availability larger than 99% (e.g., average uptime in the order of weeks and average downtime in the order of minutes).

The second contribution of the paper is an illustrative description of the evolution of the interval availability distribution with the duration of the observation period. For short observation periods, the interval availability presents a dominant density concentrated in the probability of no failure, modeled using a Dirac delta function, and for the rest of the values, the probability density is almost uniformly distributed. When the observation period increases the probability of no failure and the probability of total failure (system always down) start to be considerably reduced, being this density redistributed around the expected interval availability value. Finally, for very long observations periods, the distribution of the interval availability tends to be normally distributed.

This paper also contributes to understand better the impact of the shape parameter of the failure/repair processes in the SLA risk. In addition, this paper gives two messages: i). The stochastic variations of the interval availability have a huge impact in the SLA success probability. ii). The duration of the contract τ modifies drastically the probability distribution of the interval availability.

PAPER B

Characterization of Router and Link Failure Processes in UNINETT's IP Backbone Network
IJSSC, INDERSCIENCE Journal, 2012

This paper yields an improved insight into the failure characteristics of an operational core network. Due to the lack of such results, it represents an important contribution to the analysis of network dependability. In addition, the obtained results are used to justify some of the assumptions made in the others papers that are part of the thesis.

This work offers a classification of the network devices based on the types of threats that may affect them. The differences and similarities among the defined groups were analyzed through the evaluation of the expected number of failures and the expected unavailability. The results show that the most unreliable devices are the links that interconnect far located places, both with respect to unavailability and number of failures. Three different orders of unavailability were observed. i) The routers and short distance links that present cumulative downtimes in the order of minutes. ii) Medium distance links with values in the order of hours. iii) Long distance links where the total cumulative downtime may be specified in days.

This paper analyzes the type of probabilistic distributions of failure and repair processes with their respective parameters. The parameterized information was published without normalizing the presented values. Some of the results confirm the assumptions of previous works, where the Weibull distribution is used to model failure processes of network components. Nevertheless, this assumption seems to be not valid for the case of links that cover long distances, where the gamma distribution is a

better option. This conclusion was verified by making a hazard analysis, observing that when long distance links have survive a long period, the probability of failing per unit of time does not decrease monotonically to zero, but after some survival time this probability get fixed in a constant value. On the other hand, for the case of routers and short distance links, the longer the survival time, the bigger the probability of having optimal operational conditions.

PAPER C

Analysis of Dependencies Between Failures in the UNINETT IP Backbone Network
PRDC, IEEE Proceedings, 2010

In network dependability studies, failure independence is commonly assumed, in part for mathematical convenience, in part due to lack of a better failure model, and in some cases due to ignorance. The main contribution of this paper is to present the behavior of dependencies in the failure and repair processes at a real operational network, and to show that the independence assumption commonly used to model network dependability may not be correct.

The basis for this work has been the log of failure and repair events in the UNINETT network. The objective is to get an insight into the correlation between failure processes in an operational network. In order to reach this, several methods were developed, obtaining the following specific contributions:

- A method to evaluate if measured failure processes are renewal and independent.
- Two methods to compute the correlation coefficient of two on/off processes.
- Several correlation values were presented, analyzed, and classified.
- Geographical distance has a significant impact on the correlation.

PAPER D

Guaranteeing Service Availability in SLAs; A Study of the Risk Associated with Contract Period and Failure Process
LATINCOM, IEEE Proceedings, 2009

This paper demonstrates the importance of taking into account the distribution of the interval availability and the duration of the contract for the definition of SLA availability requirements. Two different scenarios were studied. First, non protected connections composed by several links. Second, shared backup connections, where the resources of the backup path are shared by several connections.

This paper was written before having the operational data from UNINETT (see Figure 7). For this reason, the failure and repair processes were modeled only by Weibull distributions but not gamma. The Weibull distribution was selected due to it

can be used to model different kind of burstiness in up and down times, and in this way evaluate different behaviors and scenarios.

The SLA success probability as a function of the observation period has basically two different behaviors depending on whether the guaranteed value can be met in the asymptotic case or not. In the first case, it drops below one for a period, but converges back to one. In the second case, it decreases continuously until reach zero. This was already known for negatively exponentially distributed processes (see Figure 4). However, one of the contributions of this paper is to show that the same behavior occurs when the failure/repair processes are Weibull distributed. This paper goes beyond that, showing that the higher the burstiness of the failure process (i.e., small shape parameter of the Weibull distribution), the higher the risk of failing the SLA.

A last contribution of this paper is the proposal of a method to control the probability of fulfilling the SLA availability requirement, with the simplicity of assessing asymptotic values. For this, a *Safe-Guard Factor* is proposed. This is a correction factor that computes the proportion between the availability requirement and the steady state availability, in order to obtain the SLA success probability that the network provider desires.

PAPER E

SLA Success Probability Assessment in Networks with Correlated Failures
COMCOM, ELSEVIER Journal, 2012

This paper is the next step of paper D given that the success and risk probability are also assessed in protected network connections. However, this paper takes more realistic assumptions, by using the results obtained from paper B and C. Specifically, this paper considers the following operational features:

- Operational failure/repair distributions such as Weibull, gamma and empirical.
- Correlation between failure/repair processes.
- End to end connections, compound by several links and routers, under different protection schemes.

For the assessment were proposed two approaches. First, trace driven simulation combined with circular shifting. This approach is effective for detecting potential dangers on the SLA fulfillment. The implementation is made directly from the measured operational data, but irregular tendencies for intermediate observation periods were observed. Second, a discrete event simulation that generates correlated samples according to marginal distributions. This approach offers finer results that allow the better description of the behavior of the SLA success probability.

This paper not only presents values that describe the SLA success probability, but it also shows explicitly the entire distribution of the interval availability, contributing to the better understanding of the problem of guaranteeing contracted availability. The

study of the interval availability's PDF complements the observations obtained from the SLA success probability, and it shows the variations that correlation produces in the whole distribution. For the case of correlation in series, the distribution appears slightly shifted to the right of the distribution when independence is assumed, which represents that correlation in series produces a small improvement in the availability level offered to the connection. For the case of correlation in parallel, the distribution has a bigger variance and appears considerably shifted to the left of the distribution when independence is assumed, which means that if correlation in parallel exists, the convergence to the steady state is slower and the availability level offered to the connection is drastically worse.

Finally, this paper shows the importance of taking into account correlation between failure processes for defining SLA's availability parameters and for the provision of reliable network connections.

PAPER F

Dynamic Sharing Mechanism for Guaranteed Availability in MPLS Based Networks
CQR, IEEE Proceedings, 2010

Previous papers present contributions that give a deep insight into the behavior of the interval availability in established end-to-end connections. However, when a connection request arrives, the operator has to decide which network resources should be assigned to it. This paper proposes an allocation mechanism considering availability constraints. The proposed mechanism is flexible given that it may provide different solutions depending on the availability requirements and the behavior of the network components. In this way, an unprotected connection may be provided if the requirements are fulfilled. If this is not the case, a protected connection is offered. The proposed mechanism is able to find the best sharing setting that fulfill the constraints.

The main contribution of this paper is the design of a dynamic sharing mechanism. It uses a smart algorithm that finds the optimal sharing scheme fulfilling the availability requirements and using the minimum bandwidth possible. Considerable bandwidth amounts are saved under a wide range of availability requirements and network loads. Finally, the complexity of the proposed mechanism is approximately one order of magnitude bigger than the complexity obtained by the use of conventional sharing mechanisms.

PAPER G

Guaranteeing SLA Availability in Telecommunications Networks
NETWORKS, IEEE Proceedings, 2012

Figure 7 presents this work as a transversal paper that deals with the three main topics addressed in the thesis. The main target of paper G is to contribute with a

model able to integrate the results presented in previous papers. The way to make this integration is through an stochastic optimization formulation, where the objective function maximizes the network operator profit. The proposed model can be used when the following considerations and SLA requirements have to be fulfilled.

- Finding the end-to-end path.
- SLA finite duration.
- Stochastic behavior of operational network components
- Correlation between failure/repair processes.
- Cost generated by resources utilization.
- Different protection schemes (Unprotected, Shared Backup, Dedicated Backup).
- Refund penalties.
- Operator Reputation.
- Operator profit.

To the author knowledge, there is not an optimization model in the literature able to integrate all these factors, being this work the first in this direction. The proposed stochastic programs are solvable in CPLEX. Therefore, they can be used directly solve the problem and maximize the operator profit. The addressed problem is NP-Complete and the solution may not be obtained quickly. Hence, this proposal can be used as reference to verify the accuracy of any future heuristic procedure, when given the demands of a certain scenario, the solution is required in short time.

Another important contribution of the paper is the design of an expansion method, to represent in a linear expression the end-to-end connection downtime in terms of individual links downtime for the three different protection mechanisms. This general approach can be adapted in any other optimization mechanism where the end-to-end downtime connection is relevant.

PAPER H

System Management to Comply with SLA Availability Guarantees in Cloud Computing
CLOUDCOM, IEEE Proceedings, 2012

Through the development of this thesis, many solutions to solve the problem of guaranteeing SLA availability requirement are proposed. In cloud computing scenarios, SLAs are also widely used. Therefore, the last step of this thesis was focused on applying the insight acquired, to this emerging technology.

Cloud computing is a technology with huge and increasing popularity in ICT business. One of its uses is the provision of virtual machines with a given memory,

processor speed, and storage capacity that can be adapted according to the customer needs.

The main contribution of this paper is the proposal of two smart resource provision policies that reduce considerably the SLA risk, without investing in additional infrastructure. In the previous works of this research, the shape of the interval availability under different scenarios was carefully studied. This paper goes beyond these studies and proposes manage resources mechanisms that modify such distribution in order to reduce the SLA risk by implementing smart policies in: **i)** the assignment of spare servers when virtual machines are restored. **ii)** the dynamic use of different fault tolerance licenses.

This paper shows that the SLA risk may be reduced not only by adding resources, improving the quality of the components of the system, or by improving the protection scheme used, but also, by applying policies for the efficient use of the resources available.

6.2 Summary of the contributions

An overview of the contributions of the papers is given in Figure 8. The contributions of this thesis can be summarized in three main points as follows:

- Better understanding of the distribution of the Interval Availability: The main point of this thesis is that there are many stochastic variables that affect the probability of meeting the availability requirement stipulated in an SLA. The stochastic distribution of the interval availability of a connection is the one that determines the risk behind an SLA. However, to obtain the mentioned distribution is not trivial. In fact, this is clearly an open issue as was explained in Section 3.1. Therefore, a better understanding of these functions represents a valuable contribution.

- Improved insight into the behavior of real operational networks: Several methods to guarantee contracted availability may be designed based on theoretical assumptions. However, the applicability of those methods may be increased considerably if they are validated using operational data.

- Resource Management to reduce the SLA risk: This thesis offers tools to network providers in order to allocate resources, saving considerable amount of bandwidth, while being aware of the SLA requirements. The solutions proposed target the operator profit and include terms such as reputation and penalty schemes. Finally, in this thesis were proposed novel failure management policies able to manipulate the distribution of the interval availability in order to reduce the SLA risk

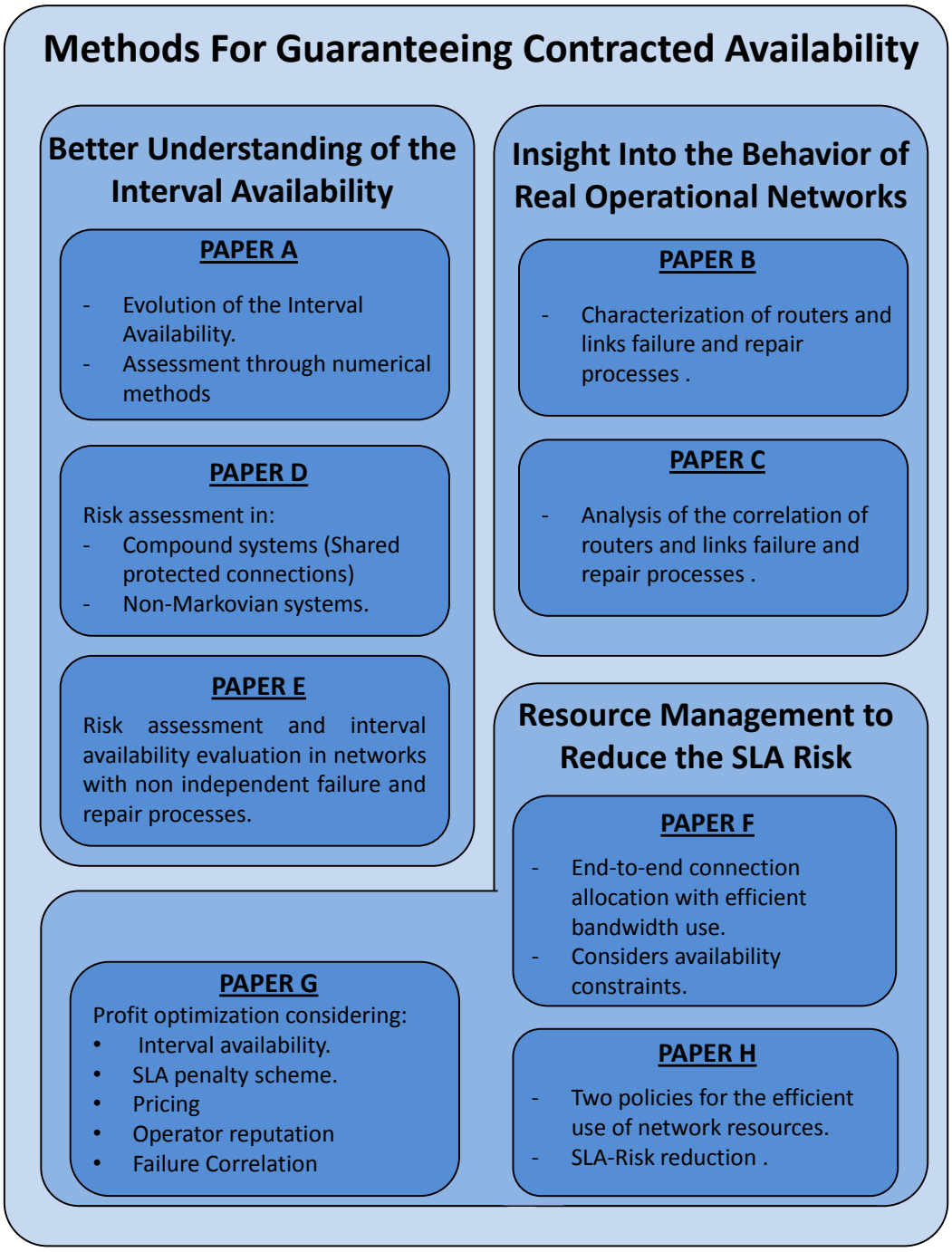


Figure 8. Summary of Contributions.

7. Conclusion

7.1 Thesis Finish Point

In Section 1 the thesis objective was presented. The need of a bridge to interconnect the SLA stipulation with its successful conclusion, where providers and customers are satisfied was illustrated in Figure 2. This challenge was addressed in this thesis by proposing new methods and concepts that contribute to the bridge construction. The final structure archived by this research is shown in Figure 9. When such structure is constructed, there is always place to improvements. However, this final result (Figure 9 structure) allows the connections of the two desired points, and the strengthening of the pillars that support such path.

The contributions to the bridge construction are summarized in two parts. First, the strengthening of three pillars by the different concepts studied in the papers is analyzed. Second, this section shows how the papers helped to build the path on top of the pillars.

Regarding strengthening of the three pillars, the following conclusions can be drawn:

The interval availability was characterized using numerical methods for single-component systems under Weibull and Gamma distributed processes. In addition, the evolution of its PDF with the duration of the observation period was studied (PAPER A).

The interval availability PDF was assessed via simulation in compound systems under Weibull, gamma and empirical distributed processes where correlation between failure processes may exist. An interesting observation is that the properties of the interval availability such as its evolution with the observation period, its convergence to a normal distribution, and the SLA risk trends are similar for both single-component Markovian systems and compound generally distributed correlated systems (PAPER E).

Regarding the characteristics of up and down times of network components, this thesis uses measurements from the UNINETT operational network. The measured processes do not have the Markovian properties. Instead, probabilistic distributions such as Weibull or gamma are better in order to model the real behavior of these components. Finally, this study shows that long distance links are the most unreliable components regarding the number of failures and the duration of downtimes (PAPER B). This thesis also shows that the presence of correlation between failure processes in operational networks is very likely. In addition, it was observed that geographical distance has a big impact in the magnitude of such correlation (PAPER C).

The contributions related to the efficient resource utilization pillar were: i) The proposal of a smart sharing mechanism that optimize the bandwidth utilization (PAPER F). ii) A method to calculate the end-to-end cumulative downtime path in terms of individual components downtime (PAPER G). iii) The structure of the SLA penalty was analyzed, allowing the inclusion of pricing and marketing models in the solution of the problem (PAPER G).

Regarding the path on top of the pillars, the following conclusions can be mentioned:

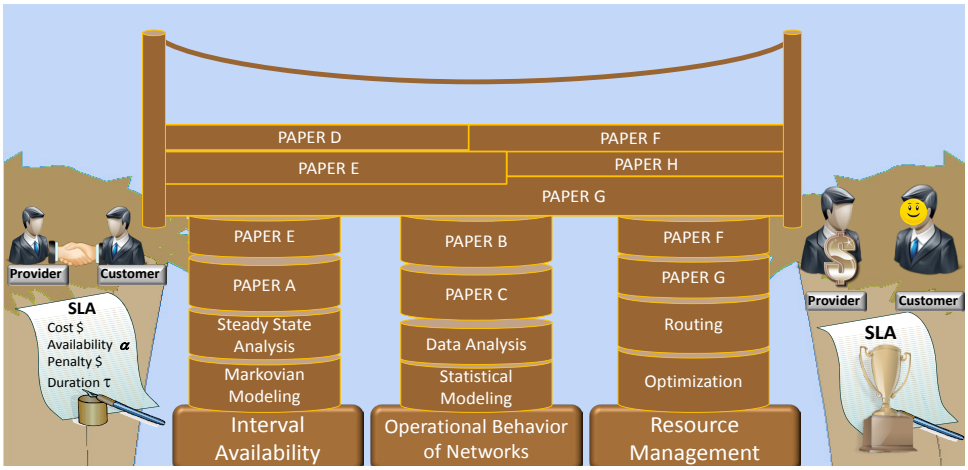


Figure 9. PhD Thesis Framework.

The path is made by five of the papers developed during the PhD. According to the chorology of the thesis, papers D and F were the first two developed works. These works may provide an end-to-end solution of the thesis problem in the following way. Paper D shows the importance of considering the entire distribution of failure and repair processes, and how sensible is the probability of failing the SLA to the duration of the contract. At the end of this work a safe guard factor is proposed. It allows network providers to control the SLA success probability, based on asymptotic results. On the other hand, Paper F is an approach that allows the allocation of network connections based also on asymptotic availability constraints. Therefore, they can be used as a joint solution to solve the problem addressed. However, given that this was the first approach developed, it does not consider important concepts such as the correlation in failure processes, the risk assessment under empirically distributed failure/repair processes, and the general formulation of the Safe Guard Factor.

Paper G is an ambitious proposal that tries to build the whole bridge. To the authors knowledge this is the first model able to consider most of the relevant concepts implied in the thesis problem. However, it is an initial proposal that still needs some refinements. One of the most urgent is the creation of heuristic procedures to obtain faster solutions.

Paper E is an study that proposes two methods to assess the SLA risk under realistic scenarios. Correlation is considered, as well as the use of real distributions in the failure and repair processes, where Weibull, gamma and even empirical distributions are addressed. In addition, these methods were used in order to observe the entire distribution of the interval availability and analyze its respective properties.

Finally, Paper H take advantage of all the knowledge acquired during the PhD in order to propose a way to reduce considerable the SLA risk in cloud computing scenarios, one of the most leading technologies in the telecommunication business nowadays.

7.2 Concluding Remarks

This thesis studies methods for guaranteeing contracted availability in connection oriented networks. A topic that becomes every day more relevant, due to: the increasing use of SLAs, and the huge commercial implications that they may have. The issues to be addressed and the state of the art were presented. Many open challenges were identified. Therefore, one can say that to date, network providers do not have enough tools in order to assess precisely the risk of failing the availability stipulated in an SLA.

This thesis shows that some of the most common simplifications used in reliability analysis may deviate considerably the risk value obtained from the assessment, leading to the definitions of SLAs that most likely cannot be met.

A basic policy to reduce the risk of not meeting the availability defined in an SLA is the assignation of highly redundant infrastructure. This thesis shows that through the use of the appropriate policies, the level of availability offered can be increased, and the SLA risk can be reduced considerably only by assigning the resources that are needed, leading to a better utilization of the infrastructure available, and to a profitable way to operate the network.

Assessing the probability distribution of the interval availability is fundamental in the SLA scenarios studied. This thesis offers a leading edge for such assessment under non independent and non Markovian systems. These two points were studied due to the operational data obtained from UNINETT shows that these are the two of the most common features of the behavior of operational networks. However, in the characterization of any other backbone network, new operational behaviors may be observed, which may demand the development of new techniques and methods that lead to an appropriate SLA risk assessment. One of the most interesting challenges addressed in this thesis was the modeling of correlated failures. The trace driven simulation proposed is a solution that may be used to model several kind of correlation, but some lack of accuracy was observed. On the other hand, the method that uses the Marshall and Olkin copula is a more accurate solution, but it only applies for the case of simultaneous failures. Operational networks may present a diverse kind of failure correlation. Therefore, the studies presented in this thesis are just a first step into the interesting task of modeling dependencies.

This thesis proposes a resource allocation mechanism using an stochastic optimization formulation that considers the most relevant issues of the addressed problem. Due to such proposal cannot guarantee fast solutions, an issue that should be addressed in future works is the design of heuristic procedures that provide solutions with better computational times. This proposal shows how interesting could be to have a model with a holistic view of the problem of guaranteeing SLA availability. In addition, this thesis tires to depict some of the most relevant pieces that may help to build a complete view of such problem. However, the difficulties and the magnitude of the challenges that need to be addressed were better understood during the development of this thesis. For this reason, the contributions presented in this dissertation represent just a first step, a guide, and a motivation in order to continue building improved and better solutions that allow the definitions of appropriate SLAs.

Our last remark is a tentative answer to the main research question addressed by this thesis: *How to assess the availability to be stipulated in the SLA, and which resources must be provided in order to guarantee that requirement?* The first thing that a provider must do is to obtain information that allows him to characterize as real as possible the operational behavior of its network. The next step is to use the appropriate policies for the resources assignment that satisfy the requirements in an economical and efficient way. These policies have to be aware of the distribution of the interval availability of the operational network connections. Finally, a general risk assessment should be done in order to verify if the resources assigned are enough in order to guarantee with a high probability a successful SLA development.

Part II

INCLUDED PAPERS

PAPER A

A Study of the Interval Availability and its Impact on SLAs Risk

Andres J. Gonzalez and Bjarne E. Helvik

Proceedings of the International Conference on Computer Science, Engineering and Applications (ICCSEA)

Delhi, India, May, 2012

A STUDY OF THE INTERVAL AVAILABILITY AND ITS IMPACT ON SLAs RISK

Andres J. Gonzalez, Bjarne E. Helvik
Centre for Quantifiable Quality of Service in Communication Systems
Norwegian University of Science and Technology,
Trondheim, Norway
{andresgm,bjarne}@q2s.ntnu.no

Abstract The obligations that telecommunications providers have with their customers are nowadays clearly specified in SLA contracts. The offered availability during the contract period is one of the most relevant variables in SLAs. Modeling accurately the transient solution posed by the need of considering the interval availability is still an open challenge. A common policy taken to make simpler models is the use of steady state assumptions. Nevertheless, this simplification may put on risk the contract fulfillment, as stochastic variations of the measured availability are significant over a typical contract period. This paper makes a theoretical study of the interval availability and proposes an approximation to evaluate the cumulative downtime distribution of a system component using renewal theory. We study the evolution of the distribution of the interval availability with the increase of the observation period i.e., duration of the contract, and show its respective impact in the SLA success probability. In addition, using the approximation proposed, we analyze numerically the behavior of the cumulative downtime distribution and the SLA risk under processes that do not follow Markovian assumptions.

1. Introduction

Real world networks are not fault free. A single failure has economic and reputation impact to the operator and incalculable consequences to the customers through the affected services. A common policy to handle this issue is the stipulation of the availability to be guaranteed in a business contract known as Service Level Agreement SLA. The promised availability must be commercially competitive, and it must fit the customer needs. In addition, it may imply huge costs in terms of the price of high reliable equipment and the penalties associated with the violation of the agreement. Therefore, the selection of the availability to be promised requires an accurate analysis. However, it is difficult to define due to the computational challenge posed by the transient solution and the stochastic variations of the interval availability. The use of steady state assumptions is a common policy taken to simplify the analysis. Nevertheless, this simplification may put on risk the contract fulfillment, as stochastic variations of the measured availability are significant over a typical contract period.

Our study is focused on network components that can be operational (up) or not operational (down), modeled as a two-state system. We start by proposing a numerical method to estimate the cumulative downtime distribution of individual network comopents. The definition of the cumulative downtime probability density function PDF in a component with general distributed failure and repair processes during a finite interval has been an open challenge for a long time. The formulation of this problem was first addressed by Takács [Tak57] and confirmed by Muth [Mut68] using different methods. They obtained a solution in terms of the convolution of the cumulative distribution function CDF of the failure/repair processes. However, an explicit solution is only given for exponentially distributed up/down times, due to the complexity posed by the n-fold convolution of CDFs with other kinds of distributions. Takács also proposed a tractable solution for general distributions, assuming long intervals ($t \to \infty$). On the other hand, Funaky et al. [FY94] proposed a good approximation assuming short intervals. Given that a typical SLA duration does not fit into the two mentioned solutions, we propose an approximation to calculate the distribution of the cumulated downtime in a SLA context, i.e., a finite contract-interval that last for several months.

The probability that a network operator *meets / do-not-meet* the contracted interval availability α will be referred as *SLA success probability / risk*. This concept was first raised for general systems in [GT88] by Goyal and Tantawi. They observed that if the promised availability is larger than the *steady state availability* (A), the success probability decreases continuously. However, they also showed that there is a considerable risk even when $\alpha < A$. They provide a numerical method to compute risk, assuming Markovian properties in the failure and repair processes. For the case of non-markovian and complex composed systems previous works such as [GH09], [MH09] and [GH11b] use simulation tools for the assessment.

This paper proposes a numerical method to characterize the entire distribution of the interval availability in network components that have failure/repair processes Weibull, gamma and negatively exponentially distributed. In addition, the evolution of the interval availability with the duration of the contract is studied and the effects of the shape parameter on the SLA success probability are shown.

This paper is organized as follows. In Section 2, we present our approach to calculate the two-state component cumulative down time during a finite interval. Section 3 describes the evolution of the distribution of the interval availability and studies the effect of the shape parameter in the SLA risk. Finally, Section 4 concludes the paper.

2. Distribution of the Cumulative Downtime During a Finite Interval

The operational status (up/down) of a network component can be modeled as a two-state system. The right modeling of such system is crucial in order to evaluate the unavailability and its associated penalty. In this section, we present a method to evaluate numerically the distribution of the accumulated downtime of a two-state system with general distributed ud/down times.

The network component state as a function of time can be modeled by a random process $O(T)$ defined as follows:

$$O(T) = \begin{cases} 1 & \text{If the component is working at time } T. \\ 0 & \text{Otherwise.} \end{cases} \quad (1)$$

The interval availability $\hat{A}(\tau)$ is a stochastic variable that measures the percentage of time that a network component has been working during τ.

$$\hat{A}(\tau) = \frac{1}{\tau}\int_0^{\tau} O(T)\,dT. \quad (2)$$

The network component behavior can be described as a sequence of failure (g) and repair (h) processes, i.e., $O(T)$ may be modeled as an alternating renewal process.

The accumulated down time over τ $t(\tau)$ is associated with $\hat{A}(\tau)$ as: $t(\tau) = \tau[1-\hat{A}(\tau)]$. $\Omega(\tau,t)$ and $\omega(\tau,t)$ are defined as the CDF and the PDF of $t(\tau)$ respectively. A general expression for $\Omega(\tau,t)$ was derived by Takács in [Tak57] as follows

$$\Omega(\tau,t) = \sum_{n=0}^{\infty} H_n(t)[G_n(\tau-t) - G_{n+1}(\tau-t)] \quad (3)$$

where the failure and repair processes are described by i.i.d. up and down times with CDF $G(t)$ and $H(t)$ respectively, and the subindex n represents the $n-$fold Stieltjes convolution of a given function.

Equation (3) characterizes a problem with general distributions. However, it is difficult to compute for specific failure and repair processes due to the complexity posed by the $n-$fold convolution of general distributed CDFs. In [FY94], $\Omega(\tau,t)$ is approximated for general distributions assuming short intervals. Nevertheless, this result cannot be applied in our problem given that the duration of a SLA is typically of several months.

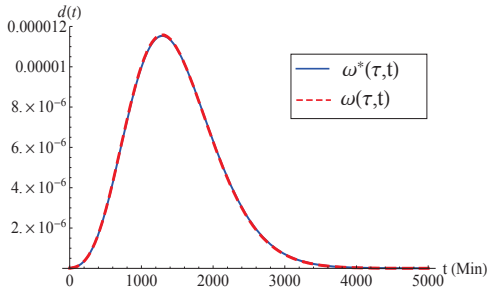
For the case of failure and repair processes exponentially distributed, a complete result was obtained by Takács as

$$\Omega(\tau,t) = e^{-\lambda(\tau-t)}\left[1 + (\lambda\mu(\tau-t))^{0.5}\int_0^t e^{-\mu y} y^{0.5} I_1(2(\lambda\mu(\tau-t)y)^{0.5})dy\right] \quad (4)$$

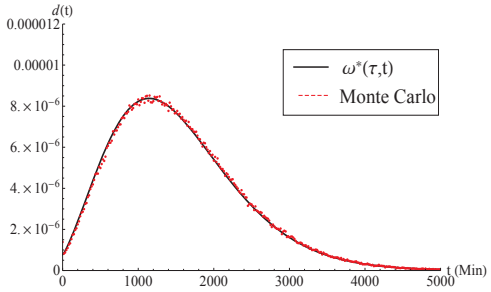
where λ and μ are the respective failure and repair rates and I_1 is the Bessel function of order 1.

Some studies (e.g., [MIB+08], [GH11a]) have shown that the Weibull and gamma distributions are representative to model real failure and repair processes. In the literature there is not an explicit expression that describes $\Omega(\tau,t)$ for these two distributions. In this paper, we suggest an approximation using renewal theory, making $\Omega(\tau,t)$ and $\omega(\tau,t)$ tractable and sufficiently accurate.

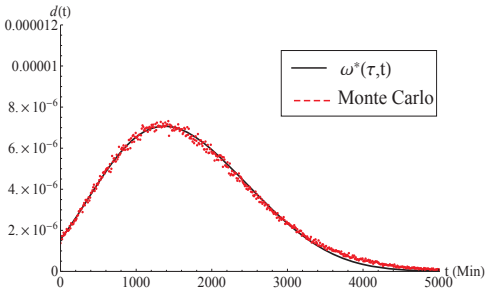
Assuming n down events during τ, the duration of each down period is assumed independent and identically distributed $h(t)$. The PDF of the total cumulated downtime



(a) $g(t)$ exponentially distributed.

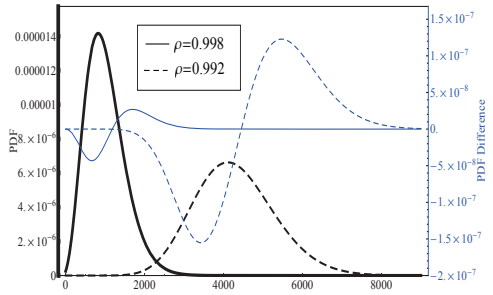


(b) $g(t)$ gamma distributed.

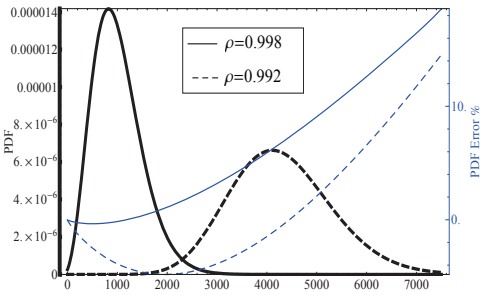


(c) $g(t)$ Weibull distributed.

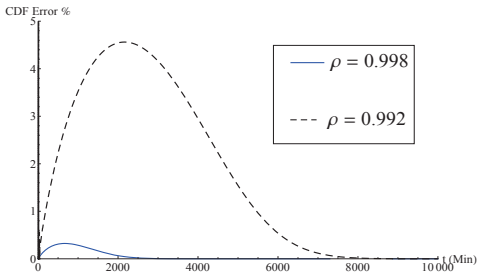
Figure 1. Approximation of the PDF cumulative downtime.



(a) **Left axis**: $\omega(\tau,t)$ density. **Right axis**: PDFs difference $(\omega^*(\tau,t) - \omega(\tau,t))$.



(b) **Left axis**: $\omega(\tau,t)$ density. **Right axis**: PDF percent error $(1-(\omega(\tau,t)/\omega^*(\tau,t)) \times 100\%)$.



(c) CDF percent error $(1-(\Omega(\tau,t)/\Omega^*(\tau,t)) \times 100\%)$.

Figure 2. Approximation error.

is given by the n-fold convolution $h_n(t)$. Hence, if the probability of n down events during τ $P(N(\tau) = n)$ is known, the problem can be solved.

Individual network components in backbone operational networks are highly reliable, with mean time to failure in the order of months and mean time to repair in the order of minutes to hours, obtaining steady state availabilities ρ usually larger than 0.999 (See for instance [GH11a]). In this context, one can assume that the downtime duration is very small compared to the uptime. Therefore, we can approximate the number of down events during τ considering only the number of renewals of the failure process.

This approximation overcomes the complexity of (3) by dividing the problem in two. First, by finding the n-fold convolution of only the PDF of the downtime. Second, by obtaining the number of down events ruled only by the uptime distribution. For this, renewal theory and counting models may be used. The approximated PDF of the total cumulated downtime is given as

$$\omega^*(\tau,t) = \sum_{n=0}^{\infty} P(N(\tau) = n) h_n(t) \tag{5}$$

The probability $P(N(\tau) = n)$ of n renewals during τ for exponentially distributed uptimes follows the Poisson distribution. The n-fold convolution of an exponential function can be easily obtained applying Laplace transform.

In [Win95], Winkelmann defines a count-data model that computes $P(N(\tau) = n)$ when the times are independent and identically gamma distributed:

$$P(N(\tau) = n) = G^i(\beta_g n, \theta\tau) - G^i(\beta_g n + \beta_g, \theta\tau) \tag{6}$$

where β_g and θ are the shape and scale parameter of the gamma distribution, respectively, and G^i is the incomplete gamma function. The n-fold convolution of gamma distributed downtimes is obtained straight forward by using the Laplace transform.

Finally, the Weibull distribution poses a bigger challenge, since its Laplace transform is analytically intractable. However, for the case of Weibull distributed uptimes, $P(N(\tau) = n)$ can be obtained by expanding the Weibull function in Taylor series. In [Lom66], for a renewal Weibull process with shape parameter β_W and scale parameter η equal to one, $P(N(\tau) = n)$ is defined as

$$P(N(\tau) = n) = \sum_{s=n}^{\infty} (-1)^{s+n} \frac{\tau^{\beta_W s}}{s!} \frac{b_k(s)}{\gamma(s)} \tag{7}$$

where $\gamma(s) = b_0(s) = \Gamma(\beta_W s + 1)/\Gamma(s+1)$ and $b_{k+1}(s) = \sum_{w=k}^{s-1} b_k(w)\gamma(s-w)$.

The convolution of Weibull distributed downtimes is approximated in [HH99] using the Saddle Point Approximation.

In order to see the accuracy of our approximation, Fig. 1 compares $\omega^*(\tau,t)$ with $\omega(\tau,t)$ on network components with three different uptime distributions with the same expected value. Fig. 1(a) shows a network component with exponentially distributed

uptimes with $\lambda = 1/30$ days ($E(g(t)) = 30$ days), and exponentially distributed downtimes with $\mu = 1/2$ hours. $\omega(\tau,t)$ was obtained using (4).

Fig. 1(b) compares $\omega^*(\tau,t)$ with a Monte Carlo simulation when network component uptimes are gamma distributed with shape parameter $\beta_g = 0.5$, scale parameter $\theta = 60$ days ($E(g(t)) = 30$ days), and exponentially distributed downtimes with $\mu = 1/2$ hours. Finally, Fig. 1(c) compares our approximation with a Monte Carlo simulation in a network component with Weibull distributed uptimes with shape parameter $\beta_W = 0.5$, scale parameter $\eta = 15$ days ($E(g(t)) = 30$ days), and exponentially distributed downtimes with $\mu = 1/2$ hours. From Fig. 1 not only the accuracy of the approximation can be appreciated, but also the variance of $\omega(\tau,t)$, which becomes bigger when the failure processes are not exponentially distributed but Weibull or gamma distributed with shape parameters shorter than one.

In order to estimate the magnitude of the error posed by our approximation, we use three different methods. First, we evaluate the PDF difference $(\omega^*(\tau,t) - \omega(\tau,t))$. Second, the percent error between PDFs $(1-(\omega(\tau,t)/\omega^*(\tau,t)) \times 100\%)$. Finally, the percent error between CDFs $(1-(\Omega(\tau,t)/\Omega^*(\tau,t)) \times 100\%)$. We use exponentially distributed processes, given that (4) can be used as reference.

Fig. 2 presents the results of applying these three methods in two different network components with ρ =0.998 and ρ =0.992, respectively. Fig. 2(a) illustrates that $(\omega^*(\tau,t) - \omega(\tau,t))$ is approximately three and two order of magnitude smaller than $\omega(\tau,t)$ in the network component with ρ =0.998 and ρ =0.992, respectively. Fig. 2(b) shows initially a negative error that becomes zero when the cumulated downtime is equal to $E[\omega(\tau,t)]$, and it becomes positive with monotonic increase for $t > E[\omega(\tau,t)]$. However, when the PDF error becomes considerable, the remaining probability mass is small $(1-\Omega(\tau,t) \to 0)$. In order to illustrate this better, Fig. 2(c) considers the remaining mass probability by estimating the error directly in the CDF.

Peak CDF errors equal to 1% and 5% are obtained for network components with ρ equal to 0.996 and 0.9915, respectively. The presented results show that our approximation will work well in backbone networks scenarios and general systems with mean time to failure in the order of months and mean time to repair in the order of minutes to hours.

3. SLA success probability

The interval availability becomes fundamental in business scenarios where a clear specification of the offered availability during a contract period has to be defined. Previous works try to assess the availability to be promised using simulation techniques. In this paper, we make use of the numerical results obtained and presented in the previous section, in order to present the behavior of the interval availability in different scenarios, using numerical methods. In this section, first we will define the success and risk of an SLA. Second, the evolution of the interval availability with increasing τ will be presented. Finally, the effect of the shape parameter of failure processes will be analyzed.

When a SLA is defined, the provider promises an availability guarantee α for a given period τ (the duration of the contract). Under this scenario, to know the

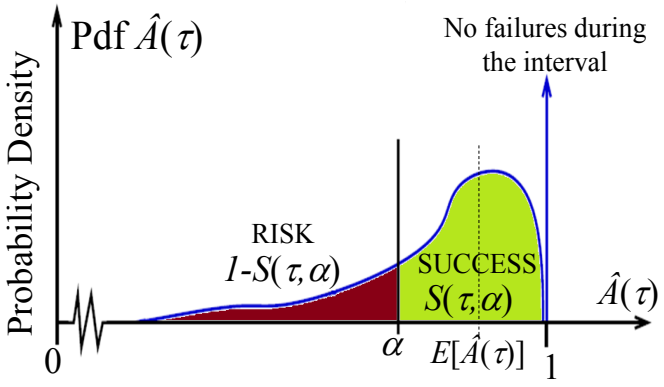


Figure 3. Interval Availability (General Shape).

probability that the availability after some observation period τ will be larger than or equal to the defined guarantee is crucial. The Success Probability is defined as follows:

$$S(\tau,\alpha) = Pr[\hat{A}(\tau) \geq \alpha] \tag{8}$$

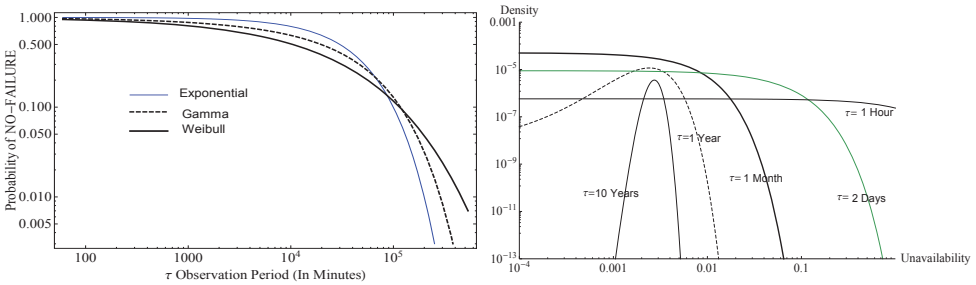
Additionally, the *risk* will be defined as the probability that the specified availability α will not be met, which can be expressed as $1 - S(\tau,\alpha)$. Fig. 3 shows the general shape of the PDF of the interval availability, and how the risk and the success of the SLA can be estimated from this information.

3.1 Interval Availability Evolution

The interval availability usually is modeled using simulation techniques. In this section, we will model accurately the shape of the interval availability, using the numerical methods described in Section 2.

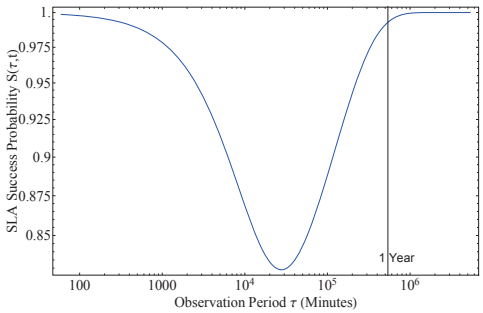
Fig. 3 shows that the density of the interval availability includes a Dirac delta function that represents the probability of no failure during the interval τ. This subsection shows explicitly the dimension of this probability i.e., the magnitude of the integral of the impulse. In Section 2 was presented how to compute the number of n down events during τ. Therefore we obtain the probability of no failures by computing $P(N(\tau) = 0)$. For the case of exponential failure processes this probability is reduced to a negative exponential function with parameter λ. For the case of gamma and Weibull distributed failure processes the probability of no failure will be computed using expressions (6) and (7) respectively.

Fig. 4(a) shows the results obtained for the probability of no failure for three different failure processes. First, negatively exponentially distributed failure processes with intensity $\lambda = 1/30$ days. Second, gamma distributed failure processes with shape parameter $\beta_g = 0.5$, scale parameter $\theta = 60$ days . Finally, Weibull distributed failure processes with shape parameter $\beta_W = 0.5$, scale parameter $\eta = 15$ days. The expected value for all the cases is the same ($E(g(t)) = 30$ days). We observe that for



(a) Probability of NO-FAILURES during τ.

(b) Interval Availability for different τ.



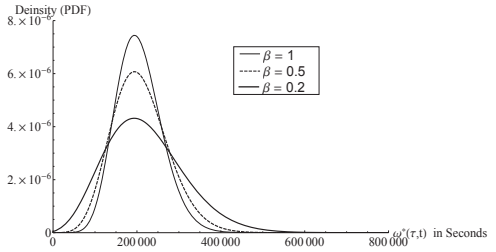
(c) SLA success probability.

Figure 4. Evolution of the Interval Availability with τ.

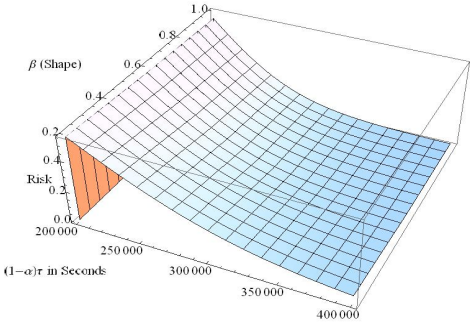
observations periods approximately shorter than 2 months (8x10^4 minutes) the Weibull and the gamma distribution present a higher reduction with increasing observation period τ. However, for observations periods larger than 2 months, the probability of no failure decreases faster under negatively exponentially distributed failures.

With the magnitude of the integral of the Dirac delta function defined, the next step is to study the rest of the distribution where interval availability values shorter than one are considered. For this, we use expression (4) in order to evaluate the interval availability in a component with negatively exponentially distributed uptimes with $\lambda = 1/30$ days, and exponentially distributed downtimes with $\mu = 1/2$ hours.

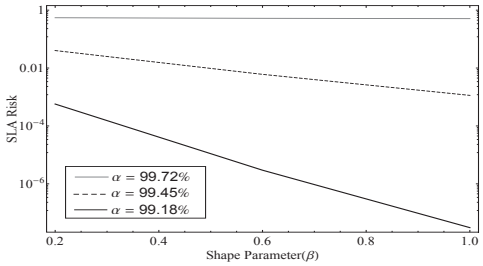
We select five different values for the observation period in order to be able to describe the evolution of the interval availability with τ. Fig. 4(b) shows the obtained results, using unavailability in the horizontal axes given that it offers a more illustrative presentation using a logarithmic scale. When the observation period is very short e.g., 1 hour (60 minutes), the density appears almost uniformly distributed being dominant the probability of no failure presented in Fig. 4(a) which is 0.9986. The next selected τ is two days (2880 minutes) in this case the probability of no failure is 0.9356. In addition the probability of high unavailability values (i.e., bigger than 0.1) is strongly reduced. When the observation period is equal to 1 month (43200 minutes) the probability



(a) $\omega^*(\tau,t)$.



(b) 3D-Risk under different values of α and β.



(c) Risk under different values of α and β.

Figure 5. The effect of β in $\omega(\tau,t)$ and $S(\tau,\alpha)$.

of no failure is reduced to 0.369 and the probability of having unavailability values higher than 0.1 and 0.01 becomes negligible and considerably reduced respectively. When τ is equal to 1 year (525000 minutes), the probability of no failure becomes very small ($5.4\text{x}10^{-6}$). In addition, the probability mass start to be concentrated near to the expected interval availability value ($E[\hat{A}(\tau)]$). Finally, when the observation period is very large i.e., 10 years (5250000 minutes), the interval availability presents a distribution close to normal as was mentioned by Takács in [FY94] and the probability of no failure becomes negligible with a value of $2.4\text{x}10^{-53}$

Finally, for the shake of illustration, Fig. 4(c) presents the shape of the success probability. This information agrees with the results shown in [GT88] and [GH09]. The information and analysis made from Fig. 4(b) combined with Fig. 4(c) provide a better understanding of the shape and stochastic variations of $S(\tau,\alpha)$.

3.2 The Effect of the Shape Parameter

Operational systems show higher occurrence of very short and long system uptimes than what is properly described by a negative exponentially distribution, e.g., [MIB+08], [GH11a]. They have found that Weibull and gamma with shape parameters less than one are more representative to model the behavior of the uptime of real systems.

We use the results obtained in Section 2 to show the effect of the uptime shape parameter in the SLA success-probability/risk.

Fig. 5(a) shows the probability distribution of the cumulative downtime for an observation period of 18 months in a two-state system with negatively exponentially distributed downtimes with $\mu = 1/2$ hours and three different gamma distributed uptimes with the same average uptime duration of 1 month (the steady state availability for all these systems is 0.997230). One can observe that when β_g is equal to 1 (negatively exponentially distributed uptimes) the distribution of the cumulative downtime appears approximately symmetrically distributed around the expected value ($E[\omega(\tau,t)]$). With the reduction of the shape parameter, the stochastic behavior of uptime duration becomes more bursty, presenting shorter and longer times. This property produces an increase in the variance of $\omega(\tau,t)$ and a loss in the symmetry around the expected value.

When an availability promise α is stipulated in an SLA, it means that the cumulative downtime has to be shorter than $(1-\alpha)\tau$. Therefore, an alternative notation for the risk would be:

$$\mathrm{RISK}(\tau,\alpha) = \int_{(1-\alpha)\tau}^{\infty} \omega(\tau,t)dt. \tag{9}$$

Fig. 5(b) shows the risk values obtained after applying equation (9) in several two-state systems with the same expected values that the cases illustrated in Fig. 5(a) and a fixed observation period of 18 months. We use different values of $(1-\alpha)\tau$ and shape parameters β_g from 0.2 to 1. Fig. 5(b) illustrates that the risk can be reduced considerably if the SLA availability promise α is shorter than the expected interval availability for all the evaluated shape parameters, and it shows the negative effects of the shape parameter in the magnitude of the SLA risk, for all values of α.

Finally, Fig. 5(c) shows specific cuts from Fig. 5(b) using three different points (α=0.9972, α=0.9945 and α=0.9918) and studying the effects of the different shape parameters on the risk. As observed before, the shorter the shape parameter value, the higher the magnitude of the risk. However, the more distant the promise is form the steady state availability, the higher the difference in orders of magnitude produced by the shape parameter.

4. Conclusion

We show that the cumulative downtime distribution of a two-state system can be accurately approximated in a backbone network context, using counting models and feasible PDF's Laplace transforms.

Pervious works have study the effects of the interval availability using simulation or via numerical methods by taking Markovian assumptions. This paper complements previous works, offering a numerical method to analyze non Markovian systems.

The detailed analysis of the distribution of the interval availability helps to understand better the implications of signing availability promises in SLAs. We observe that for short observation periods, the interval availability presents a dominant density concentrated in the probability of no failure, which can be modeled using Dirac

delta function. For the rest of the availability values the probability density is almost uniformly distributed. When the observation period increases, the probability of no failure and the probability of permanent failure (system always down) start to be considerably reduced, and this density start to be redistributed around the expected interval availability value. Finally, for very long observations periods, the interval availability distribution may be fitted to a Normal distribution.

With the reduction of the shape parameter for values below one (bursty uptimes) the risk increases considerably. In addition, the smaller the promised availability is, the higher the difference produced by the shape parameters, reaching differences of several orders of magnitude.

The assumption of steady state conditions may simplify the modeling of several dependability problems. However, as this paper shows, different stochastic variations may have a huge impact in the success of an SLA.

References

[FY94] Kenich Funaki and Kazuho Yoshimoto. Distribution of total uptime during a given time interval. *IEEE Transactions on Reliability*, 43(3):489 –492, Sep. 1994.

[GH09] Andres J. Gonzalez and Bjarne E. Helvik. Guaranteeing service availability in SLAs; a study of the risk associated with contract period and failure process. *Proceeding of the IEEE Latin-American Conference on Communications (LATINCOM)*, pages 1 –6, Sep. 2009.

[GH11a] Andres J. Gonzalez and Bjarne E. Helvik. Analysis of failures characteristics in the UNINETT IP backbone network. *Proceeding of the IEEE 7th International Symposium on Frontiers in Networking with Applications (FINA)*, Mar. 2011.

[GH11b] Andres J. Gonzalez and Bjarne E. Helvik. Guaranteeing Service Availability in SLAs on Networks with Non Independent Failures. *IEEE-IFIP International Workshop on Design of Reliable Communication Networks (DRCN)*, Oct. 2011.

[GT88] Ambuj Goyal and Asser Tantawi. A measure of guaranteed availability and its numerical evaluation. *IEEE Transactions on Computers*, Volume 37, Issue 1:25 – 32, 1988.

[Lom66] Z. A. Lomnicki. A note on the weibull renewal process. *Biometrika*, 53(3/4):pp. 375–381, 1966.

[MH09] Anders Mykkeltveit and Bjarne E. Helvik. Adaptive management of connections to meet availability guarantees in SLAs. In *Proceeding of the IFIP/IEEE International Symposium on Integrated Network Management, IM. Mini-Conference*, Jun. 2009.

[MIB+08] Athina Markopoulou, Gianluca Iannaccone, Supratik Bhattacharyya, Chen-Nee Chuah, Yashar Ganjali, and Christophe Diot. Characterization of Failures in an Operational IP Backbone Network. *IEEE/ACM Transactions on Networking*, 16(4):749–762, Aug. 2008.

[Mut68] Eginhard J. Muth. A method for predicting system downtime. *IEEE Transactions on Reliability*, R-17(2):97 –102, Jun. 1968.

[Tak57] Lajos Takacs. On certain sojourn time problems in the theory of stochastic processes. *Acta Mathematica Hungarica*, 8:169–191, 1957.

[Win95] Rainer Winkelmann. Duration dependence and dispersion in count-data models. *Journal of Business & Economic Statistics*, 13(4):pp. 467–474, 1995.

PAPER B

Characterization of Router and Link Failure Processes in UNINETT's IP Backbone Network

Andres J. Gonzalez and Bjarne E. Helvik

International Journal of Space-Based and Situated Computing (IJSSC)

Vol. 2, No 1, pp. 3 -11, 2012.

CHARACTERIZATION OF ROUTER AND LINK FAILURE PROCESSES IN UNINETT'S IP BACKBONE NETWORK

Andres J. Gonzalez, Bjarne E. Helvik
Centre for Quantifiable Quality of Service in Communication Systems
Norwegian University of Science and Technology,
Trondheim, Norway
{andresgm,bjarne}@q2s.ntnu.no

Abstract Backbone networks must be highly reliable. The offered availability can be predicted prior to operation if the stochastic behavior of network components is known. The aim of this paper is to provide information about failures and repairs processes in an operational network. Operational logs from the UNINETT's core network were analyzed to obtain distributions of the time between failures and downtimes of routers and links. The network components were classified according to their role in the network. The measured processes were fit with well-known distributions. The inter-failure times of routers and short distance links may be characterized by a Weibull distribution, but for the long distance links the gamma distribution yielded a better characterization. The difference is discussed using hazard analysis. The parameters of each network component are published, providing a detailed insight that may be used for dependability predictions and research.

1. Introduction

The design of highly dependable backbone networks has acquired a lot of importance in the last years. Users, companies and society as a whole become every day more dependent of the services carried by these networks. Due to their high capacity, millon of users may be affected by a single failure, generating economic and reputation impact to the operator, and incalculable consequences to their customer due to the affected services. Real world networks are not fault free, making the availability guarantee, as well as other dependability attributes, salient parameters inside the Service Level Agreement (SLA), which is a business contract between a network operator and is customers.

Several studies have aimed the provision of tools for an adequate SLA definition. In [GH10] is proposed an algorithm that allocates end-to-end connections in network's links with assumed steady state availabilities, fulfilling bandwidth and availability requirements. Other works on SLA assessment assume Markovian properties in failure and repair processes, due to mathematical convenience, or due to the lack

detailed of real-life data ([FZ02], [MH09]). However, as shown in [GH09], the real failure and repair processes (e.g., Weibull distributed time between failures instead of exponential) have a major impact on the probability of meeting a specified SLA target for the interval availability. Hence, the importance of having information about the real behavior of such processes.

This paper analyzes operational data, and it is focused on describing dependability characteristics of the elements of an IP backbone network, such as the distributions of time between failures, and availability parameters. The investigation is based on logged failures made available by the Norwegian academic network operator UNINETT [UNI12a]. The results shown in this paper are based on measurements made from January 2008 to December 2009, a period where the network did not undergo relevant changes neither in topology nor in supporting layer 2 infrastructure.

Previous works provide substantial information of the dependability in operational access networks ([CSKM07], [MVM02]). On the other hand, studies on backbone networks ([InCM+02], [MIB+08]) normalize the published values due to commercial restrictions. This paper publishes explicit parameters of the failure processes in an operational backbone network. It proposes an original device classification that fits the dependability properties of network's elements. Finally, it develops a hazard analysis of the characterized processes, helping to the better understanding of the obtained results, and offering a physical interpretation of the studied processes.

We classify the network devices according to their dependability features, obtaining four different groups. In addition, we evaluate availability values and failure intensities in order to identify the differences and similarities among the defined groups and to have an initial understanding of the dependability behavior of the UNINETT network. Based on the obtained results, we found that assuming perfect routers (which is a common policy) is not valid.

An important objective in this paper is to fit the observed processes with well-known distributions that may allow their easy replication in other studies. There are many procedures in order to have trustable fit of measured data. Initially, we apply Quantile-quantile plots (Q-Q plots) to have a visual and intuitive evaluation. In addition, we use the method of maximum likelihood to estimate specific distribution's parameters. Finally, we run tests that evaluate whether a data set is well-modeled or not by the estimated distributions. Given that fitting procedures are common in many experimental analysis, in [CDG+04] and [MPP+06] is presented a statistical toolkit that tries to standardize this procedures.

We found that the failure processes in short and long distance links have different features. We show that the Weibull distribution traditionally used for modeling link failures processes is not accurate for the case of long distance fibers. In this case, the gamma distribution is a better option. We use the concept of failure rate (hazard function) in order to explain the found difference.

This paper is organized as follows. Previous works are briefly reviewed in Section 2. Section 3 introduces the UNINETT's IP backbone network, the data collection method and the information used for the analysis. In Section 4, the failure intensity and unavailability of the network elements are analyzed. Routers, short (intrasite), medium

(regional) and long (intercity) links are identified as types of elements having differing characteristics. Section 5 outlines the techniques used to fit distributions, analyzes the obtained results by using of cumulative distribution and hazard functions, and presents the fitted parameter values for all elements. Section 6 concludes the paper.

2. Previous Work

In spite of the importance of using realistic failure processes for dimensioning network availability, the access to such information is limited. This is due to a number of reasons, among them that failures of their network are not what operators like to have exposed in a competitive commercial marketplace. However, there have been some recent studies on operational failure data giving valuable insights. In [CSKM07] a study of spatial and temporal failures and outages in an access network was performed to assess availability. Another study in [MVM02] estimates the time between failures and times to repair for elements in a large wireless access network, finding that they are not consistent with exponential distributions, but they may better be described by Weibull or two-stage hyper-exponential distributions. A study of the failure behavior in an operational backbone network is reported by [InCM$^+$02]. They examine the frequency and duration of failure events and discuss various statistics, like the distribution of inter-failure times and the distribution of link failure durations. Nevertheless, some information is missed given that for proprietary reasons they normalized the published values. This work was continued in [MIB$^+$08], where failures and repairs in the Sprint IP backbone Network are classified and analyzed. They perform a characterization of the different classes of failures found. [BL97] presents expected values of the Mean Time Between Failures (*MTBF*) and Mean Time To Repair (*MTTR*), collected from several network equipment and design a model to perform end-to-end availability studies. In [KNR09] Kuusela and Norros analyze router failure logs from the Finnish academic network, FUNET, and in [KN10] they describe a method that can be used to assess downtimes due to joint failures.

3. UNINETT Network Description

UNINETT is the network that connects universities, colleges and research institutions in Norway. It is a nonprofit organization giving us the opportunity to publish dependability related information without commercial restrictions. UNINETT provides service to several hundred thousand users, carrying many critical applications. Hence, availability is a major concern, as can be appreciated in the highly redundant network topology presented in Figure 1. The core of the network interconnects the main norwegian cities through optical fiber connections of 10 and 2.5 gigabit per second (Gbps). Full details about the UNINETT topology can be found in [UNI12b]. The studied backbone is operated using WDM technology and several brands of routers. IS-IS is used for intradomain routing.

In this paper, we are interested in analyzing dependability features of the core network. Hence, we have select the subset of connections considered by UNINETT as the backbone. They are at the same time the connections that interconnect the

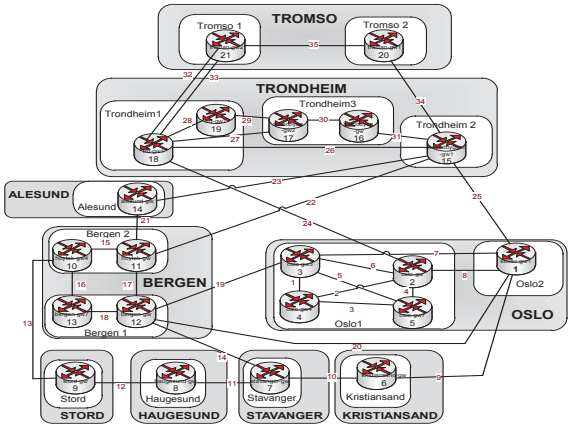


Figure 1. UNINETT Core Network Topology.

main cities in Norway as is shown in Figure 1. The failure and repair processes of the routers and links shown in this figure will be analyzed in the next chapters.

The failure logs were obtained by a network management system, recording all events at the IP layer. Logs are available since January 2001. However, the data that will be used in this paper is from January 2008 to December 2009. The reason for choosing this period is that the core network did not undergo any relevant change, neither in topology, nor in routers and links, which makes it likely that the processes are stationary and uninterrupted.

Router's and link's failures are registered with a precision of seconds. The data collection method follows SNMP standards, where for each new component installed SNMP agents enable the detection of changes in the network operation. Those changes may be identified either by periodical polling, using GetRequest/GetResponse messages generated from the central server, or by trapping techniques generated on the remote agent that uses notification messages able to capture every change online. Events that imply the whole router are identified by *"no-response/reachable"* messages, while link failures are reported on specific router interfaces by *"linkUp/linkDown"* messages. In case of a router failure, the network management platform does not register all the individual down events of each router interface/link.

The information provided by UNINETT contains summaries of events based on raw SNMP data without any previous processing. In addition, the information about some events may be replicated. Therefore, a set of PERL scripts were implemented to obtain clean *on-off* processes for each network component. The obtained processes were verified using alternatives means such as traffic logs and UNINETT Customer Relationship Management (CRM) information. Additional information about the UNINETT operations may be found in [UNI12a].

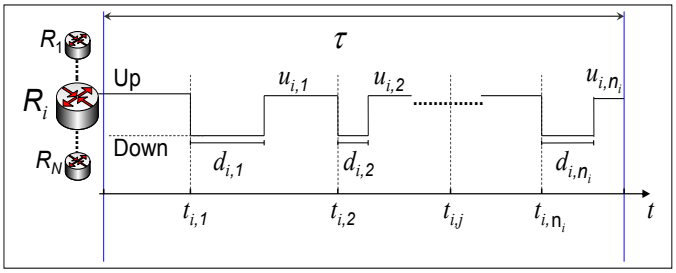


Figure 2. On-Off process of a Network component.

Table 1. Network Device Classification

Group ID	Group Name	Characteristic
G_1	Routers	Core Routers
G_2	Short Link	Intrasite (Few meters)
G_3	Medium Link	Regional (Few kilometers)
G_4	Long Link	Intercity (Hundreds of kilometers)

4. Availability of Network Components

This section regards failure intensities and unavailability of network elements. The elements analyzed are those shown in Figure 1, and the observation period (τ) is from 1 January 2008 to 31 December 2009.

The following notation is used. Failure and repair events of N network components are analyzed. Each element i ($i = 1,2,....,N$) has an operational state that may be described by an on-off process as illustrated in Figure 2. Failure j ($j = 1,2,....,n_i$) of element i occurs at time $t_{i,j}$, where the downtime duration is denoted by $d_{i,j}$ and n_i is the total number of failures of device i during τ. After a repair, the time when the device is working properly before a new failure occurs will be defined as uptime and will be denoted as $u_{i,j}$ i.e., $u_{i,j} = t_{i,j+1} - t_{i,j} - d_{i,j}$.

Network components are classified according to their role in the network, which determines their exposure for externally induced faults and operating conditions. Four different groups G_x ($x = 1,2,3,4$) are identified as shown in the Table 1. All routers have roughly the same role and operating conditions. Short distance links interconnect routers located in the same data-center (i.e., site). This means that they are short in their physical extent, they have an uncomplicated layer 2 and they are less exposed to environmental stress. Therefore, they are mainly affected by synchronization problems, electrical fluctuations, human failures and similar incidents occurring within a controlled access domain. Most failures, also physical impairments, are relatively easily rectified. Medium distance links connect routers located at different sites inside the same city. Most of the threats that affect short distance links also apply for medium distance links. However, the layer 2 handling of the links is often more complex, the environmental exposure to weather or civil engineering activities, (e.g., diggers) are

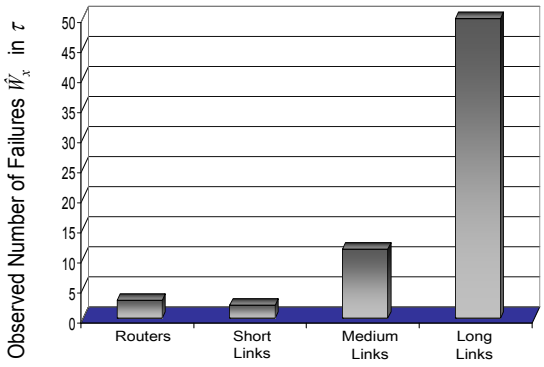


Figure 3. Observed Number of failures $\hat{W}_x$ during τ.

larger. Finally, long distance links, connecting cities hundreds of kilometers apart, are provided by layer 2 optical fiber infrastructures leased from other operators. Typically, these fibers are spun on electrical power lines, or located in ditches along the railways. They are exposed to a bigger number of threats giving the large number of variables implied in the wide areas covered.

The average number of failures per network element, $\hat{W}_x$ for G_x is obtained as:

$$\hat{W}_x = \sum_{i=1}^{M_x} \frac{n_i}{M_x}. \tag{1}$$

where M_x is the total number of elements that belong to G_x.

Figure 3 shows $\hat{W}_x$ for each of the groups of Table 1, during the observation period τ, and it may be interpreted as failures per τ. The number of failures in long distance links is approximately fifteen times bigger than in routers and short distance links, and five times than in medium distance links. This difference was expected from the discussion above. At the same time medium distance links present approximately four times more failures than routers and *intra-site* links. Note also that devices from group G_1 and G_2 have similar average number of failures. Hence, the common assumption that routers are (unconditionally) more reliable than links, and may be assumed to be failure free is not entirely true.

On the other hand, we will also analyze availability levels on routers and links during the observation period τ. The observed unavailability $\hat{U}^i(\tau)$ in a single component i will be calculated as follows:

$$\hat{U}^i(\tau) = \frac{\sum_{j=1}^{n_i} d_{i,j}}{\tau}. \tag{2}$$

Based on the classification made on Table 1, we compute the average unavailability per group G_x as indicates the next equation:

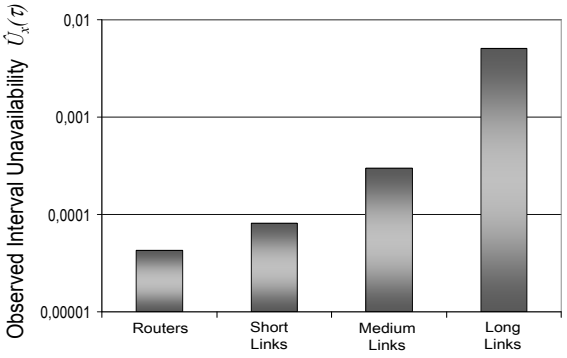


Figure 4. Observed Average Interval Unavailability $\hat{U}_x(\tau)$.

$$\hat{U}_x(\tau) = \sum_{i=1}^{M_x} \frac{\hat{U}^i(\tau)}{M_x}. \tag{3}$$

Figure 4 shows the obtained values of $\hat{U}_x(\tau)$. We may say that the average cumulative downtime of short distance links and routers during τ is in the order of minutes, for medium distance links is around 3 hours, and for long distance links is approximately two days. The difference between routers and long distance links is approximately two orders of magnitude. Therefore, we may say that the perfect routers assumption mentioned before, would be relatively valid in the study of a network with only these two kinds of components.

The ratio between $\hat{U}_x$ and $\hat{W}_x$ shows that links have more difficult repair processes than routers even for the case of short distance links. In Figure 5, we computed the average up and down time on each of the links and plot them against the physical length of the fiber. These results not only show the explicit values for mean up and down times, but also describe two tendencies: First, the mean uptime is worse (decrease) with distance. Second, the mean downtime is better (shorter) when the links are shorter. Figure 5 shows that the unavailability increase observed in long distance links is due to both; failure and repair process.

Finally, we are interested in evaluate if the duration of up and down times are independent. For this we calculate the Pearson correlation coefficient ρ in the following way: We define a vector U_i that contains uptime durations of link i ($U_i = \{u_{i,1}, u_{i,2}, .., u_{i,n_i}\}$) and a vector D_i that contains downtime durations of link i ($D_i = \{d_{i,1}, d_{i,2}, .., d_{i,n_i}\}$). We compute ρ using the next equation:

$$\rho_{U_i,D_i} = \frac{E[(U_i - \mu_{U_i})(D_i - \mu_{D_i})]}{\sigma_{U_i}\sigma_{D_i}} \tag{4}$$

Figure 6 shows the results of applying (4) on the UNINETT network's links. An important observation is that some links present high ρ values on all kind of links (intrasite, regional, intercity). With the information at hand was not possible to

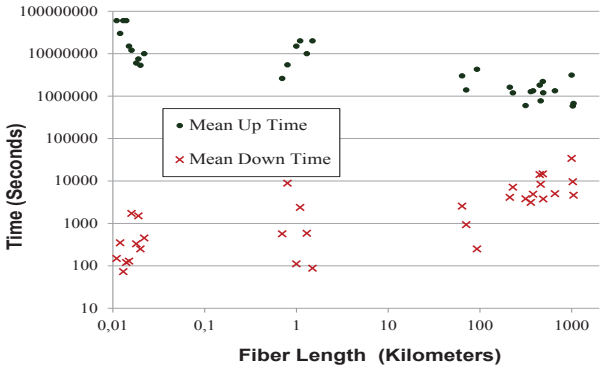


Figure 5. Mean UP/Down time Vs Fiber Length.

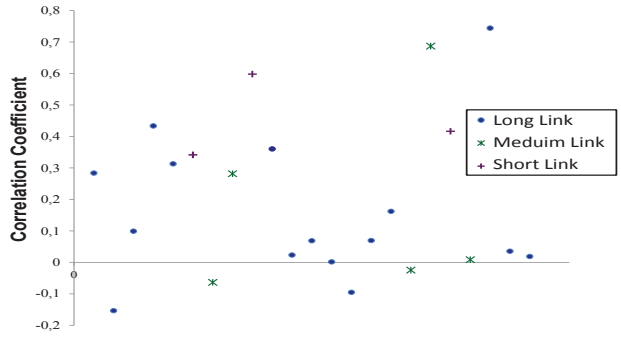


Figure 6. Correlation Coefficient of the UP/Down time duration.

observe additional patterns. However, the message of Figure 6 is that up and downtime processes are not always independent.

The average failure intensity, the average unavailability and the mean up/down time provide a first insight of the network dependability analyzed. In the next section, we will show additional information about the stochastic features of the failure and repair processes.

5. Time Distributions

This section shows stochastic characteristics of the UNINETT's network components through the estimation of up/down time distributions. First, we describe the fitting techniques used for estimate parameters distributions. Then, we show and explain the obtained values.

5.1 Distribution Fitting and Goodness of Fit

In order to know if parameterized distributions may be used to model the failure/repair processes of the UNINETT network, we use initially Q-Q plots. In this

visual tool, the pattern of points is used to compare two distributions. When the deviation between the two patterns is not considerable, we may assume that the empirical data fits the hypothesized distribution. Using the information obtained from the Q-Q plots, we apply the method of maximum likelihood estimation in order to obtain parameters that may fit the hypothesized theoretical distribution. We obtain 95% confidence intervals and evaluate if the empirical cumulative distribution function (CDF) does not lie beyond them.

The obtained parameters of the hypothesized CDF may be tested through the use of well known goodness-of-fit test e.g., Kolmogorov-Smirnov, Camer-von Mises and Anderson Darling. The application of these tests is analyzed for gamma distributions in [WVMD84], and for Weibull distributions in [SL99].

Given that this kind of procedures are widely used by the scientific community, the NIST in cooperation with other institutions have developed a handbook [NIS11] that tries to standardize many related issues. We use these standards and tools such as Matlab and Wolfram-Mathematica in order to obtain and verify results.

5.2 Uptime Fitting

Based on previous studies ([MVM02], [MIB+08]), we are interested in verify if the failure processes of the components that belong to the UNINETT core network may be modeled by a Weibull distribution. The probability density function (PDF) of this distribution will be defined as

$$f(t) = \frac{\beta(t)^{\beta-1}}{\theta^{\beta}} e^{-(\frac{t}{\theta})^{\beta}}, \forall t \geq 0, \theta > 0, \beta > 0. \tag{5}$$

where θ is the scale parameter in time units (e.g. seconds) and β the shape parameter (It is seen that when $\beta = 1$ the distribution becomes exponential).

We use the filtered on-off processes and apply the procedures explained in 5.1 in order to verify if the Weibull distribution may be used to model uptimes. In Figure 7, we provide an example that describes the typical behavior observed when a CDF-fit is applied for uptimes of routers, short and medium distance links. This figure shows maximum likelihood estimates (MLEs) for a gamma and a Weibull distribution with respective confidence bounds of 95%.

It can be noticed that the empirical CDF lies out of the upper bound of the gamma-fit for uptime values around 0.1 X 10^7 seconds. This observation shows the low accuracy of the gamma-fit and the convenience of the Weibull-fit. We verify this result by using goodness-of-fit tests, probing that uptimes of devices that belong to G_1, G_2 and G_3 are well described by a Weibull distribution.

The fitted parameters are shown in Table 2 for the case of links. We notice that the shape parameters (β) are less than 1 for all the cases, and the scale parameters (θ) are usually bigger than one month.

Table 3 shows the fitted parameters for the case of routers with enough number of samples in order to obtain trustable results and satisfy the goodness-of-fit test.

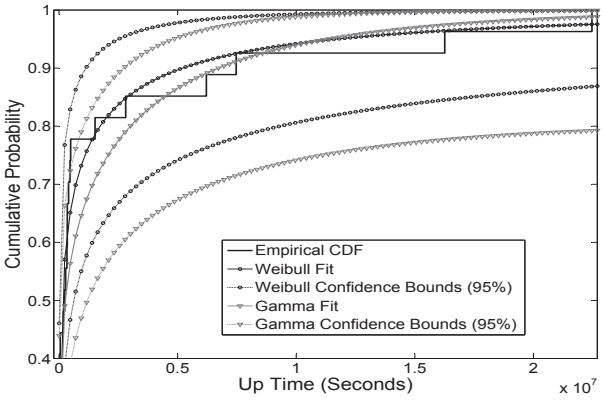


Figure 7. Cumulative Distribution fitting for Uptimes in routers and links with short and medium distance.

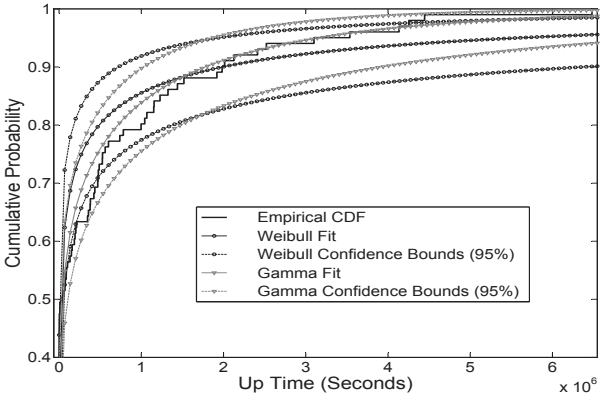


Figure 8. Cumulative Distribution fitting for Uptimes in links that connect different cities.

Table 2. Links Up-Times that fit a Weibull Distribution

Link ID	Distance	Scale Parameter (θ)	Shape Parameter (β)
15	Short	4450480	0,3645
16	Medium	1452318	0,3561
17	Medium	2684259	0,3573
18	Medium	400670	0,3254
26	Medium	1414095	0,331
27	Medium	4378835	0,4251

The Weibull-fit does not seem to be valid for links that interconnect far located cities. We found that the use of gamma distributions is a better alternative. Here, we show the procedures that lead us to this conclusion.

Table 3. Routers Up-Times that fit a Weibull Distribution

Link ID	Distribution	Scale Parameter (θ)	Shape Parameter (β)
9	Weibull	1603156	0.549
13	Weibull	1819710	0.460
16	Weibull	5910717	0.358

Table 4. Links Up-Times that fit a Gamma Distribution

Link ID	Distance	Scale Parameter (λ)	Shape Parameter (α)
9	Long	3419930	0,172
10	Long	6648830	0,175
11	Long	3104930	0,128
13	Long	2061830	0,138
14	Long	9526930	0,165
19	Long	11323500	0,154
20	Long	4162060	0,182
22	Long	3727250	0,348
23	Long	6125220	0,203
24	Long	8708750	0,244
25	Long	6200540	0,189

The notation used for the probability density function of the gamma distribution is:

$$f(t) = \frac{(t)^{\alpha-1}}{\lambda^{\alpha}\Gamma(\alpha)} e^{-(\frac{t}{\lambda})}, \forall t \geq 0, \lambda > 0, \alpha > 0. \tag{6}$$

where λ is the scale parameter in time units and α the shape parameter.

In order to show the difference with the Weibull-fit, Figure 8 shows the typical result obtained when a CDF-fit is performed for uptimes on links that interconnect far located cities (G_4). This figure also uses maximum likelihood estimates (MLEs) with confidence intervals of 95%. In this case, we observe how the empirical CDF lies out of the lower confidence bound of the Weibull-fit for uptime values around 0.5 X 10^6 seconds. We also performed goodness-of-fit tests that confirm that uptimes of long distance links may be characterized by a gamma distribution.

The estimated parameters are shown explicitly in Table 4. The values shown are based on (6). We observe that the shape parameters (α) are smaller than the obtained on the Weibull-fit, indicating a higher burstiness in the failure processes.

5.3 Differences Between the Weibull and Gamma processes

We will study the theoretical differences between the gamma and the Weibull distributions in order to get an explanation of the results shown above.

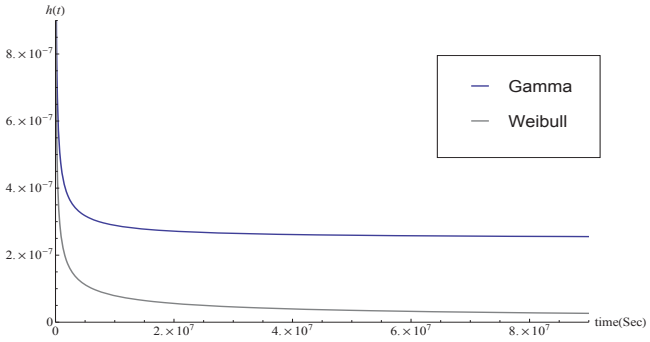


Figure 9. Hazard Function on Uptimes with gamma and Weibull Distributions.

The Weibull and gamma distributions may be used to model monotonically increasing or decreasing failure rates, also known as hazard function $h(t)$. It describes the probability per time unit that a system fails during a short interval after having been operational without failure up to time t.

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t < UP \leq t + \Delta t \mid UP > t)}{\Delta t} \tag{7}$$

The failure rate function of the Weibull and the gamma distribution are described by the following equations [AMZ87]:

For Weibull:

$$h_W(t) = \frac{\beta (t)^{\beta - 1}}{\theta^{\beta}} \tag{8}$$

For Gamma:

$$h_g(t) = \frac{(t)^{\alpha - 1} e^{-(\frac{t}{\lambda})}}{\lambda^{\alpha} \int_{\alpha}^{\infty} \frac{(t)^{\alpha - 1} e^{-(\frac{t}{\lambda})}}{\lambda^{\alpha}} dt.} \tag{9}$$

$h_g(t)$ involves the incomplete gamma function. Alternative techniques can also be used to model it. For instance in [HH99] is proposed the use of the saddlepoint approximation in order to compute *"relatively straightforward"* $h_g(t)$.

Figure 9 shows the hazard function of a gamma and a Weibull distribution with scale parameter of 46 days and shape parameter equal to 0.5 in both cases.

The main difference between the gamma and the Weibull distribution lies on the fact that the failure rate function of the gamma distribution gets stable and tends to a constant value for big uptimes, contrary to the Weibull distribution where for shape parameters bigger than one the failure rate always increase up to infinity and for values of β shorter than 1 decrease monotonically to 0.

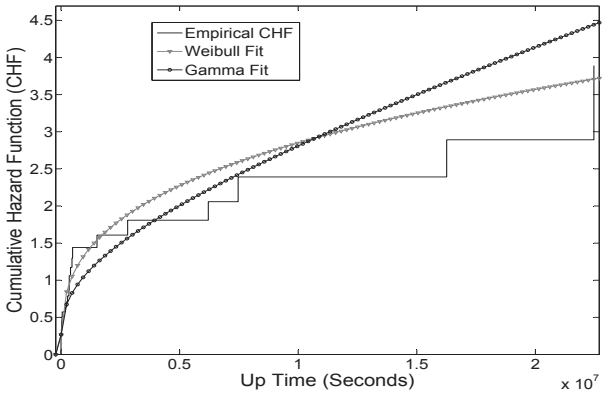


Figure 10. Cumulative Hazard Function (CHF) fitting for Uptimes in routers and links with distance 2 and 1.

In order to have a better illustration of the results observed in Figures 7 and 8, we make a fit using the Cumulative Hazard Function (CHF), which is the accumulation of the failure rate over time defined as follows:

$$H(t) = \int_{-\infty}^{t} h(x)dx. \tag{10}$$

In Figure 10, we show the typical behavior observed when a CHF-fit is performed for uptimes on devices that belongs to G_1, G_2 and G_3. The Weibull-fit shows a CHF with a big increase at the beginning of the curve, for short uptime values that gradually reduce the increase rate up to 0 ($dH_W(t)/dt \to 0$) for big uptimes, describing more precisely the empirical CHF. According to the hazard definition this indicates that when a devices has survive a long period, the probability of failing per unit of time decrease considerably to values close to 0. The devices that present this behavior (G_1, G_2 and G_3) are affected by failures which in some way may be controlled by the network operator i.e., if one of those devices have survived for a very long time, may imply that the threats have been controlled. We may say that for these kind of devices, the longer the survival time, the bigger the probability of having optimal operational conditions.

On the other hand, Figure 11 shows the typical behavior observed when a CHF-fit is performed for uptimes on links that interconnect far located cities (G_4). In this case, the gamma-fit shows a CHF with a more moderate increase rate than the Weibull-fit for short uptimes and gradually converge to a constant increase rate ($dH_g(t)/dt \to C$), describing more precisely the empirical behavior of this kind of links. This indicates that when those devices have survived a long period, the probability of failing per unit of time is not reduced, but it gets fixed in a constant value. We may say that for long distance links a long survival time imply just partial optimal operational conditions, while there are some remaining threats out of the control of the operators.

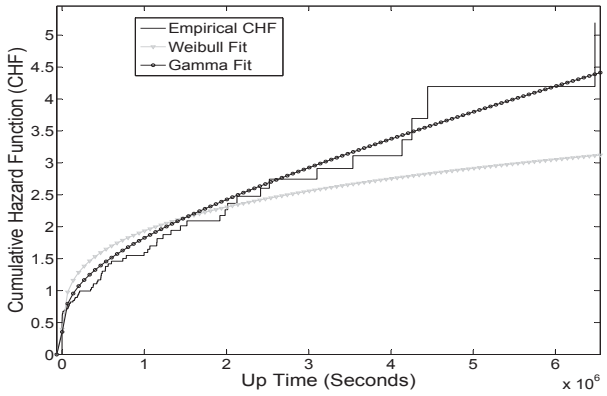


Figure 11. Cumulative Hazard Function (CHF) fitting for Uptimes in link that connect different cities.

The presented CHF figures not only illustrate better the results observed in Figures 7 and 8, but also they help to get a physical interpretation of the behavior of uptimes, given that the theoretical differences of the Weibull and the gamma distribution are directly highlighted.

5.4 Downtime Fitting

For the case of downtimes distributions, we observe different stochastic behaviors depending on the kind of device and its geographical location. Common to all of them is that they can not be fitted to any simple type of parameterized distribution. Figure 12 illustrates this issue for downtimes on link 21.

After analyzing the obtained downtimes, we notice that they are influenced by different processes that affect their stochastic properties. For the case of short downtimes, we conclude that synchronization problems have a huge impact. They are usually solved in a short period. On the other hand, when long downtimes exist, better planned repair procedures are needed. UNINETT organizes them according to priorities, depending on the impact of the failure and the availability of human and physical resources.

6. Concluding Remarks

This paper yields an improved insight into the failure characteristics at a real operational core network. First, a classification based on the types of threats that may affect the devices was made. The differences and similarities among the defined groups were analyzed through the evaluation of the expected number of failures and the expected unavailability. It was found that the most unreliable kind of devices are the links that interconnect far located places, both in unavailability and number of failures. Three different orders of magnitudes were observed for unavailability values. First, the routers and short distance links which present cumulative downtimes

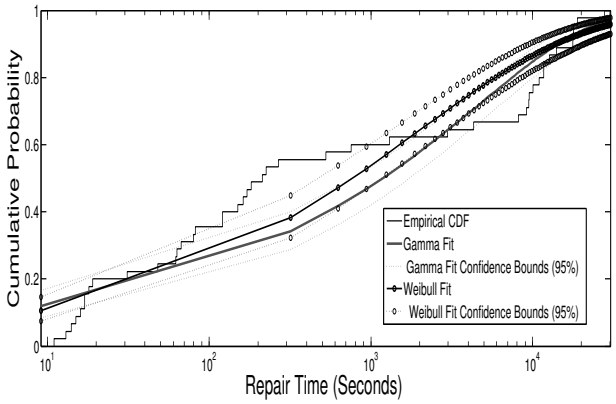


Figure 12. CDF in the Repair Processes.

in the order of minutes. Second, the medium distance links with values in the order of hours. Finally, the long distance links, where the total cumulative downtime in the observation period may be specified in days. When the mean uptime and the mean downtime are plotted against the physical distance of the link, a decreasing and an increasing trend was detected, respectively. That means that the better average availability in shorter links (worse in longer links) is due to both, failure and repair processes.

A very interesting observation is that routers and short distance links present similar dependability features, making very unprecise the assumption made in some related works, where routers are assumed to be failure free, considering only failures on links.

For some UNINETT devices, we confirm the findings of some previous works where the Weibull distribution seems to be a good option in order to model failure processes of network components. Nevertheless, this assumption seems to be not valid for the case of links that cover long distances, where the gamma distribution is a better option. This shows that when long distance links have survive a long period, the probability of failing per unit of time does not decrease monotonically to zero, but after some survival time this probability get fixed in a constant value. On the other hand, for short and medium distance links, the longer the survival time, the bigger the probability of having optimal operational conditions.

This work shows specific values that clearly describe the distributions that may fit uptimes of operational devices from a real network, where the obtained shape parameters for long distance links suggest a high burstiness on the failure processes. Additionally, the provided values do not have any kind of normalization. Therefore, this information may be used in future dependability studies.

For the case of downtimes at the UNINETT network, we could not find any simple type of parameterized distribution that fits successfully their empirical behavior. The repair procedures at the UNINETT network are organized according to priorities depending on the impact of the failure, and according to the availability of human and

physical resources needed for the repair, which affects the stochastic properties of those processes.

References

[AMZ87] Khalil Sheikh Anwar, Ahmad Munir, and Ali Zulfiqar. Some remarks on the hazard functions of the inverted distributions. *IEEE Transactions on Reliability*, 19(4):255 – 261, 1987.

[BL97] John R. Birge and Francois Louveaux. *Introduction to Stochastic Programming*. Springer Series in Operations Research and Financial Engineering. Springer, 1997.

[CDG+04] Pablo. Cirrone, Stefania Donadio, Susanna Guatelli, Alfonso Mantero, Barbara Mascialino, S. Parlati, Maria G. Pia, Andreas Pfeiffer, Alberto Ribon, and Paolo Viarengo. A goodness-of-fit statistical toolkit. *IEEE Transactions on Nuclear Science*, 51(5):2056 – 2063, Oct. 2004.

[CSKM07] Baek Young Choi, Sejun Song, George Koffler, and Deep Medhi. Outage analysis of a university campus network. *Proceedings of 16th IEEE International Conference on Computer Communications and Networks (ICCCN)*, pages 675 – 680, 13-16 Aug. 2007.

[FZ02] Maxim S. Finkelstein and Vladimir I. Zarudnij. Laplace-transforms and fast-repair approximations for multiple availability and its generalizations. *IEEE Transactions on Reliability*, 51(2):168 –176, Jun. 2002.

[GH09] Andres J. Gonzalez and Bjarne E. Helvik. Guaranteeing service availability in SLAs; a study of the risk associated with contract period and failure process. *Proceeding of the IEEE Latin-American Conference on Communications (LATINCOM)*, pages 1 –6, Sep. 2009.

[GH10] Andres J. Gonzalez and Bjarne E. Helvik. Dynamic Sharing Mechanism for Guaranteed Availability in MPLS Based Networks. *Proceeding of the IEEE International Communications Quality and Reliability (CQR)*, Jun. 2010.

[HH99] S. Huzurbazar and Aparna V. Huzurbazar. Survival and hazard functions for progressive diseases using saddlepoint approximations. *International Biometric Society Journal of Biometrics*, 55(1):pp. 198–203, 1999.

[InCM+02] Gianluca Iannaccone, Chen nee Chuah, Richard Mortier, Supratik Bhattacharyya, and Christophe Diot. Analysis of link failures in an IP backbone. *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurment (IMW)*, pages 237–242, 2002.

[KN10] Pirkko Kuusela and Illka Norros. On/Off process modeling of IP network failures. In *Proceeding of the IEEE/IFIP the 40th Annual International Conference on Dependable Systems and Networks (DSN)*, 2010.

[KNR09] Pirkko Kuusela, Ilkka Norros, and Pertti Raatikainen. Report on modelling the reliability of an ip-network and strategies for improving the reliability. Technical report, A report of the IPLU-II project, Jun. 2009. Available at http://iplu.vtt.fi.

[MH09] Anders Mykkeltveit and Bjarne E. Helvik. Adaptive management of connections to meet availability guarantees in SLAs. In *Proceeding of the IFIP/IEEE International Symposium on Integrated Network Management, IM. Mini-Conference*, Jun. 2009.

[MIB+08] Athina Markopoulou, Gianluca Iannaccone, Supratik Bhattacharyya, Chen-Nee Chuah, Yashar Ganjali, and Christophe Diot. Characterization of Failures in an Operational IP Backbone Network. *IEEE/ACM Transactions on Networking*, 16(4):749–762, Aug. 2008.

[MPP+06] Barbara Mascialino, Andreas Pfeiffer, Maria Grazia Pia, Alberto Ribon, and Paolo Viarengo. New developments of the goodness-of-fit statistical toolkit. *IEEE Transactions on Nuclear Science*, 53(6):3834 –3841, Dec. 2006.

[MVM02] Steven M. Matz, Lawrence G. Votta, and Mohammad Malkawi. Analysis of failure and recovery rates in a wireless telecommunications system. *Proceedings of the IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, pages 687 – 693, 2002.

[NIS11] NIST/SEMATECH. e-handbook of statistical methods, [online]. available at: http://www.itl.nist.gov/div898/handbook/. 2011.

[SL99] Toshiyuki Shimokawa and Min Liao. Goodness-of-fit tests for type-i extreme-value and 2-parameter weibull distributions. *IEEE Transactions on Reliability*, 48(1):79 –86, Mar. 1999.

[UNI12a] UNINETT. The Norwegian Research Network. Downtime Statistics. [online]. Available at: http://drift.uninett.no/downs/. 2012.

[UNI12b] UNINETT. The Norwegian Research Network. Network Topology. [online]. Available at: http://drift.uninett.no/stat-q/load-map/uninett,,traffic,peak. 2012.

[WVMD84] Brian W. Woodruff, Philip J. Viviano, Albert H. Moore, and Edward J. Dunne. Modified goodness-of-fit tests for gamma distributions with unknown location and scale parameters. *IEEE Transactions on Reliability*, R-33(3):241 –245, Aug. 1984.

PAPER C

Analysis of Dependencies Between Failures in the UNINETT IP Backbone Network

Andres J. Gonzalez, Bjarne E. Helvik, Jon K. Hellan and Pirkko Kuusela

Proceedings of the IEEE 16th Pacific Rim International Symposium on Dependable Computing PRDC

Tokio, Japan, December, 2010

ANALYSIS OF DEPENDENCIES BETWEEN FAILURES IN THE UNINETT IP BACKBONE NETWORK

Andres J. Gonzalez,[1] Bjarne E. Helvik,[1] Jon K. Hellan,[2] and Pirkko Kuusela,[3]

[1] *Centre for Quantifiable Quality of Service in Communication Systems*
Norwegian University of Science and Technology,
Trondheim, Norway
{andresgm,bjarne}@q2s.ntnu.no

[2] *UNINETT, The Norwegian Research Network,*
Trondheim, Norway
jon.kare.hellan@uninett.no

[3] *VTT, Technical Research Center of Finland,*
Helsinki, Finland
pirkko.kuusela@vtt.fi

Abstract Dependencies between failures in operational networks may have a huge impact on their reliability and availability. In this paper, we analyze failure logs to identify simultaneous and potentially correlated failures in routers and links of an IP backbone network. We show that the actual behavior of failure processes does not support the independence assumption commonly used in theoretical studies. Scatter plots are presented to visualize the failure processes, observing that geographical adjacency has a pronounced effect in the correlation. The existence of high correlation coefficients and high autocorrelation in some failure processes was observed. A formal analysis confirms this. The consequences of these dependencies on the provisioning of guaranteed availability are briefly discussed.

1. Introduction

Availability is a significant element for provisioning a good QoS, and it is one of the most important parameters in setting up a service level agreement (SLA) between providers and users of a network service. Analysis of real failure processes in networks are mandatory in order to get the appropriate information for availability dimensioning and to deal with the risks associated with SLA agreements. In spite of this, for a number of reasons, among them that failures of their network are not what operators like to have exposed in a competitive commercial marketplace, the access to such failure log information is very limited, and hence few studies based on data from operational networks are performed. In [CSKM07] a study of spatial and temporal

failures and outages in an access network was performed to assess the availability. Another study in [MVM02] estimates the time between failures and times to repair for elements in a large wireless access network, finding that they are not consistent with exponential distributions, but they may better be described by Weibull or two-stage hyper-exponential distributions. A study of the failure behavior in an operational backbone network is reported by Iannaccone et al. [InCM+02]. They examine the frequency and duration of failure events and discuss various statistics, for instance the distribution of inter-failure times and distribution of link failure durations. This work was continued by Markopoulou et al. in [MIB+08], where failures and repairs in the Sprint IP backbone Network are classified and analyzed. They perform a characterization of the different classes of failures found. In [KNR09] Kuusela and Norros analyze router failure logs from the Finnish academic network, FUNET and in [KN10] they describe a method that can be used to assess downtimes due to joint failures.

This paper is based on real operational data, and it focuses on finding correlations between failure processes. The investigation is based on 9 years of logged failures made available by the Norwegian academic network operator UNINETT [UNI12a].

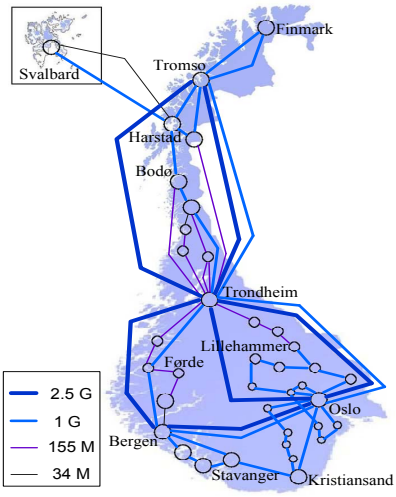
In the analysis of network reliability, independence between failure events is commonly assumed, in part for mathematical convenience, in part due to lack of a better failure model, and in some cases due to ignorance.

An objective in this paper is to show that independence assumptions are incorrect, and in this respect, our findings confirm the initial observations of Markopoulou et al. [MIB+08]. The main contribution of this study is to show that the correlation of the failures of network elements is strongly related to their geographical distance. This is demonstrated by visualization and by formal analysis.

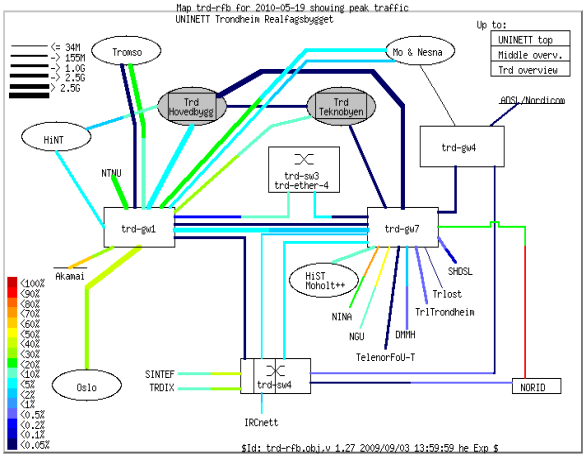
This paper is organized as follows. First, the UNINETT's IP backbone network, the information collection method and the data used for the analysis are presented in Section 2. In Section 3, the dissimilarity between the observed failure process and the process that would result if network elements failed according tho independent renewal processes, is studied. Section 4 describes three different methods used for the study of failure dependence, and the respective obtained results are illustrated and analyzed. Finally, Section 5 describes some potential future works and concludes the paper.

2. UNINETT Network Description

UNINETT is the network that connects universities, colleges and research institutions in Norway. The core of the network interconnects the main norwegian cities through optical fiber connections of 10 and 2.5 gigabit per second (Gbps), forming rings to ensure that the loss of a single link does not cause any loss of connectivity. In addition, there are at least two disjoint paths between the major universities. A large and increasing part of the network is connected to this core through 1 Gbps links and for locations with both a smaller number of users and relatively high costs of establishing links the capacity is typically 155 Mbps and some few with 34Mbps. Figure 1(a) shows a global overview of the UNINETT topology. Each *"node"* in this



(a) Global Overview of the UNINETT Network.



(b) Detailed Configuration of the Trondheim Node.

Figure 1. UNINETT Network Topology.

map represents a geographical area that may contain several interconnected buildings and smaller locations. For instance, Figure 1(b) illustrates more explicitly the network details inside the node Trondheim (see Fig. 1(a)) in order to illustrate that each node in the main overview may have several router and links.

The failure logs were obtained through a centralized network management controlled form the UNINETT NOC (Network Operations Center) in Trondheim, since January 2001 until October 2009. During this period, the global design of the network was conserved, but many individual changes were performed e.g., router replacements, new optical fiber installation, etc. Therefore, in the performed studies, we select relatively network stable intervals. Router's and link's failures are registered with a

precision in the order of seconds. The data collection method follows SNMP standards, where for each new component installed, SNMP agents enable the detection of changes in the network operation. Those changes may be identified in two ways: Either by periodical polling, using GetRequest/GetResponse messages generated from the central server, or by trapping techniques generated on the remote agent that uses notification messages able to capture every change on line. The router state is identified by *"no-response/reachable"* messages, and the link state is reported as failures on specific router interfaces associated with a link using *"linkUp/linkDown"* messages.

In some cases, trap messages are generated by network agents when a device return to an operational state, reporting to the central server the local log of the failure. This information may help to correct and define more precisely a down-time obtained by polling. For this reason, the use of an intelligent mechanism able to perform this correction is necessary. It is also important to clarify that for the case of links, the failures are identified in the SNMP management system as failures in specific routers interfaces associated with a link name. A logical link contains two physical fibers that in most of the cases report failures simultaneously from each side. However, our study avoids trivial conclusions by not considering as simultaneous two interface-events that are reported very close in time in different places, if they belong to the same link. Finally, the information obtained from UNINETT contains summaries of events based on SNMP data without any previous processing. Therefore, a very important phase in our work was the implementation of PERL scripts in order to obtain a clean UP/DOWN state in time for each network component.

The studied backbone is operated using WDM technology, routers from several brands, and using IS-IS as routing protocol. A more detailed information of the UNINETT network can be found in [UNI12a].

3. Empirical Behavior of Aggregate Failure Processes

A first objective in this paper is to compare the real stochastic behavior of failure events with the Poisson assumption that has been used during many years in the field of network dependability.

The following notation and considerations are used. Routers and links will be analyzed separately. Failure events within a fixed observation period T will be considered, where N network components are regarded. Each device i ($i = 1, 2, ..., N$) has an operational state that may be described by an UP/DOWN signal as Figure 2 illustrates. A failure j occurs at time $t_{i,j}$ and the downtime duration is denoted by $d_{i,j}$, where n_i is the number of failures of device i during T and $j = 1, 2, ..., n_i$.

For more than four decades mathematical models based on Poisson assumptions have been developed. An argument for this is the Palm-Khintchine theorem which states that the aggregation of a large number of processes tends to be Poisson distributed if: each individual process is renewal, no process is *dominant* (failure intensity very high compared to the others), and they are independent from each other [Cox67]. Under these assumptions the aggregation of the N failure processes will become a Poisson process with parameter λ where

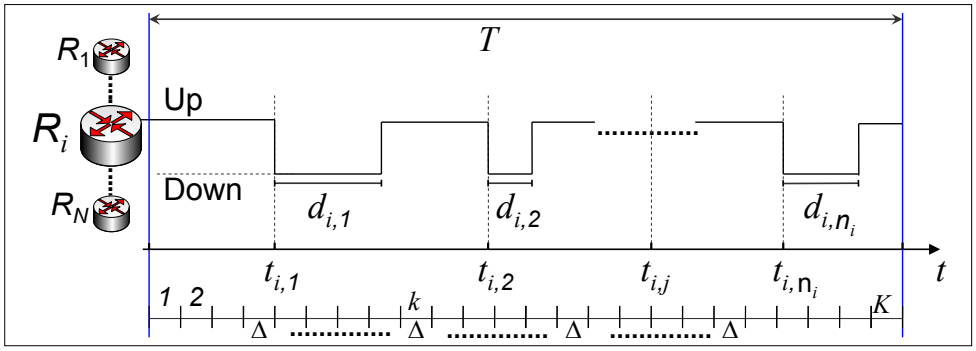


Figure 2. Behavior of a Network Component.

$$\lambda = \sum_{i=1}^{N} \frac{n_i}{T}. \tag{1}$$

We want to find the coincidence of M failures from different devices. Therefore, the occurrence of two or more failure events $P(M > 1)$ within a given short time interval Δ is of interest. From a theoretical point of view, using the Poisson parameter obtained in (1) this value may be written as:

$$P_{Poiss}(M > 1) = e^{-\lambda\Delta} + \lambda\Delta e^{-\lambda\Delta} \tag{2}$$

On the other hand, we can obtain empirical values for $P(M > 1)$ using the information from the failure logs. First, we split T in K small slots Δ as is shown in Figure 2. Then, we evaluate the number of failures m_k on each slot k ($k = 1, 2, ..., K$) as follows:

$$m_k = \sum_{i=1}^{N} \sum_{j=1}^{n_i} I\Big((k-1)\Delta < t_{i,j} \leq k\Delta\Big), \tag{3}$$

where $I(x)$ is the Indicator function. [1]

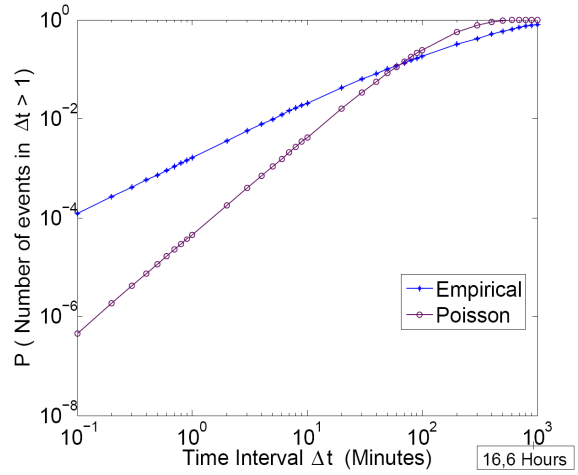
Considering all the network components $(1, 2.., N)$ the empirical $P(M > 1)$ is obtained as follows:

$$P_{emp}(M > 1) = \frac{\sum_{k=1}^{K} I\Big(m_k > 1\Big)}{K}, \tag{4}$$

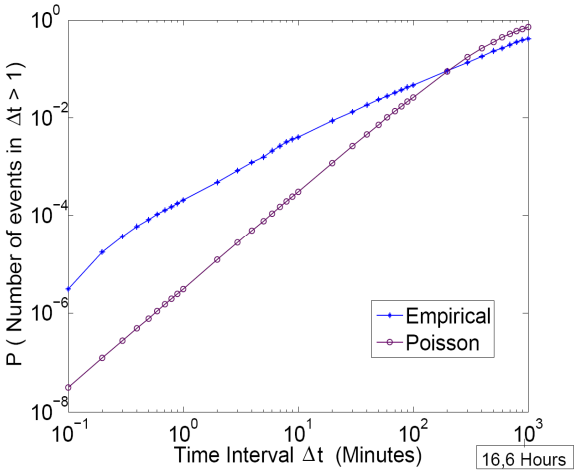
where $K = T/\Delta$ is the number of slots in the evaluation period.

Figure 3 illustrates $P(M > 1)$ under different Δ on links and routers separately during the year 2007. It shows that the empirical probability is larger for almost all Δ smaller than the inverse of the average failure intensity $(1/\lambda)$, which is 95 minutes for

[1] $I(x)$ takes value 1 if condition x is fulfilled and 0 if not.



(a) Probability of two or more Link-Down events in a time interval.



(b) Probability of two or more Router-Down events in a time interval.

Figure 3. Coincident events based on empirical logs and Poisson assumptions during 2007.

links, and 304 minutes for routers. The difference is two orders of magnitude for small Δ'*es*, clearly showing that failures in the real life tend to occur more coincidentally than what is expected under the Poisson assumption.

According to the Palm-Khintchine theorem, this difference may be due to the failure processes not being renewal or independent, or (most likely) both. This fact suggests a dependency between failure events that has to be verified through more specific methods in the next sections.

4. Dependence Analysis

Three different methods are used to identify and analyze dependencies among failure events. The first method is visual. The second one identifies simultaneous events depending on geographical distance. Finally, the last method regards the evaluation of auto-correlation and correlation coefficients of the failure processes.

4.1 Scatter Plot Analysis

Scatter plots enable us visually to identify patters in routers and links failures. In this method the data is displayed as a 2-dimensional collection of points where each of them represents a $t_{i,j}$ on the horizontal axis, and a device index i on the vertical axis.

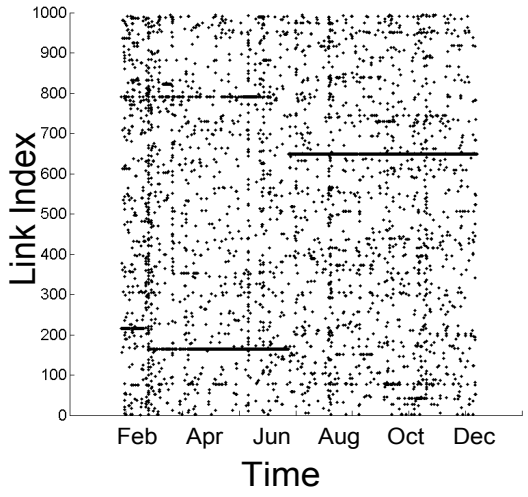
The scatter plots in Figure 4 contain link failures during year 2007. They show corresponding results to the observed in [MIB+08], where horizontal and vertical patterns may be identified, suggesting the presence of dependencies in space and time. For instance, the broken horizontal stripes indicate burst of failures from a component i over the duration of the stripe. Therefore, these processes are unlikely to be renewal and autocorrelation may exist. On the other hand, the formation of vertical patterns is due to the existence of simultaneous failures and potentially correlated events of different components. This suggests that a failure in one device may have influence on others.

An advantage of the scatter plots is that we may observe changes in the overall pattern by changing the ordering of the y axis, i.e., the indexing of devices. An interesting result is obtained when the links are indexed according to their geographical location as can be observed in Figure 4. In 4(a), the links are indexed randomly, while in 4(b) the links are indexed according to geographical location from the north to the south of Norway[2].

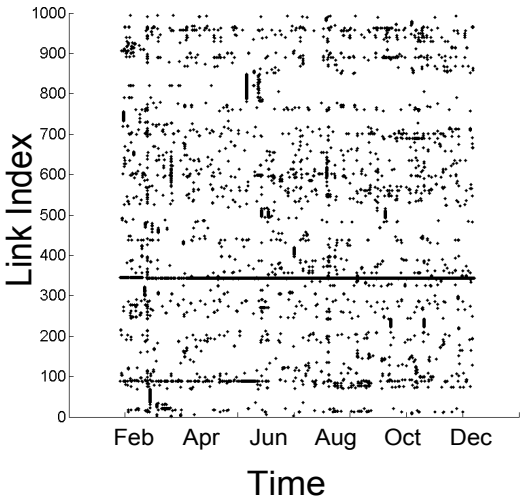
An initial observation is that the dots exhibit a more pronounced clustering in Figure 4(b), where some areas are characterized by a low or a high dot density, representing geographical zones and time periods with links relatively stable or with high amount of failures respectively. An important observation obtained from Figure 4(b) is the presence of more compact and clearly defined vertical stripes, suggesting not only that failures may occur simultaneously, but also that geographical adjacency has a high relevance. Similarly, horizontal stripes are identified, indicating that this behavior also tend to coincide in the same geographical area. Note for instance the horizontal stripe across the entire year in Figure 4(b), which is seen to stem from at least three different links by comparison with Figure 4(a).

A similar scatter plot for routers is presented in Figure 5. It shows the presence of horizontal and vertical patterns as well. The comparison of Figures 5(a) and 5(b) yields similar observations, where geographical location is an important actor.

[2]The number of dots is the same on both Figures (4(a) and 4(b)).



(a) Random Ordering of Link Index.

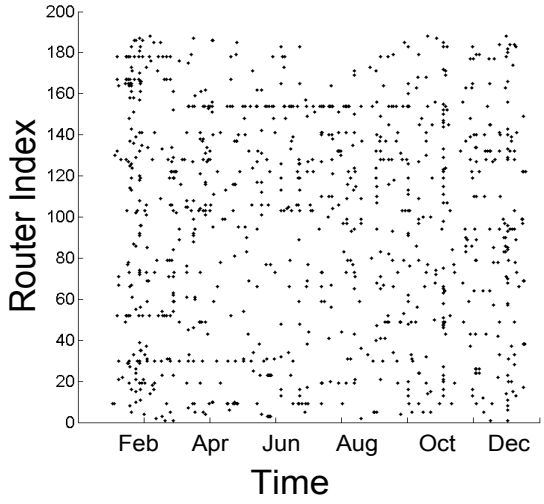


(b) Link Index Ordered by Location.

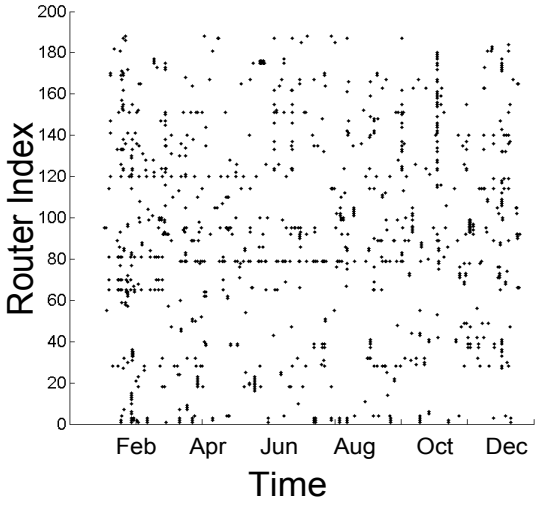
Figure 4. Links Down Events During 2007.

We may conclude that failure processes are highly unlikely to be renewal or independent, and that the coincidence in failures is significant in the UNINETT network. The dependence is stronger between devices that are geographically close.[3]

[3]In the scatter plots, the limited printing resolution makes impossible to discriminate closely located points. Our study will use complementary methods to verify the results obtained.



(a) Random Ordering of Router Index.



(b) Router Index Ordered by Location.

Figure 5. Routers Down Events During 2007.

4.2 Simultaneous Events Analysis

In this section, we introduce a method to identify simultaneous and potentially correlated failure events. First, the notion of simultaneous event has to be clarified. Ideally, a simultaneous event occurs when two down-events are reported exactly at the same time. Nevertheless, due to clock differences, delays in the system and also due to the propagation time of one failure over the others, this requirement has to be relaxed. Hence, we introduce a flexibility gap Δ, where two failures are considered

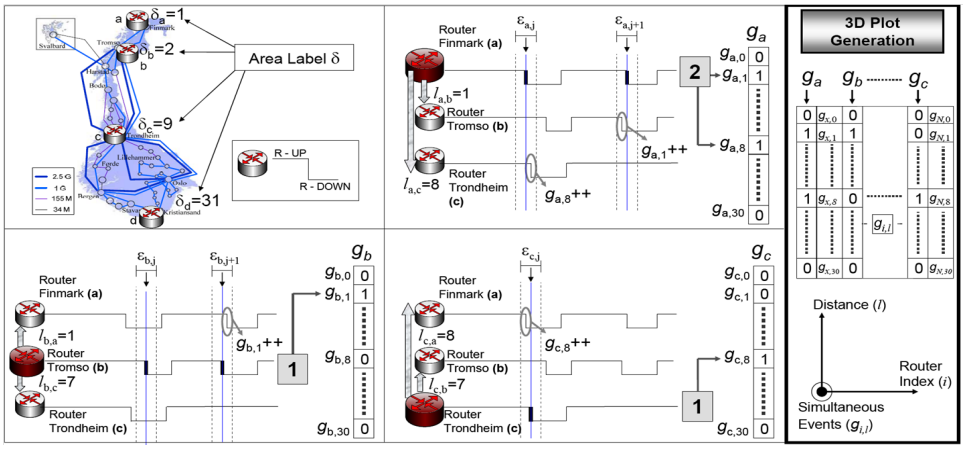


Figure 6. Methodology for the Construction of the Simultaneous 3D Graph.

simultaneous if they occur within Δ, i.e., $t_{i,j}$ and $t_{x,y}$ are simultaneous if $|t_{i,j} - t_{x,y}| \leq \Delta/2$. The magnitude of this gap will be determined using the information in Figure 3, selecting a Δ where $P_{Poiss}(M > 1)$ is very small. More specifically a value of Δ equal to 30 seconds was used for routers since the ratio $P_{emp}(M > 1)/P_{Poiss}(M > 1)$ has its approximate maximum at this gap value. Using a similar approach, for the case of links, the selected value of Δ was 10 seconds.

Motivated by the results obtained from the scatter plot analysis in Section 4.1, we want to evaluate the existence of simultaneous failures in two different network components, and the influence of geographical distance in such effect.

Figure 1(a) illustrates that network components are grouped in different geographical areas clearly defined and well delimited. Therefore, we can easily group all the components inside them to have an initial idea about the distance between elements, in a practical and organized way. For this reason, each area will be labeled with a value δ, using a geographic sweep from the north to south of Norway as is shown in the upper left side of Figure 6. The concept of *distance* l used in this paper will be the difference between the label of components i and x denoted by $l_{i,x} = |\delta_i - \delta_x|$. In this way, a router i located in Finmark will get a label $\delta_i = 1$, and it will have a distance $l_{i,x} = 8$ from a router x in Trondheim which has label $\delta_x = 9$.

An appropriate way to illustrate the number of simultaneous failures of network elements and the *distance* between them, is by a three-dimensional plot.

The procedure to generate the mentioned 3D plot is illustrated in Figure 6 and works as follows: The occurrence of simultaneous failures will be evaluated on each network element i separately. When a down event j occurs at $t_{i,j}$, a gap $\varepsilon_{i,j} = [t_{i,j} - \Delta/2\ ,\ t_{i,j} + \Delta/2]$ will be used to search which other component x suffered a failure f within this interval, taking into account the distance $l_{i,x}$ between the two affected devices.

When all the failures n_i in a network element i are checked, a vector g_i with the number of simultaneous events that occur per distance l will be obtained. If all these

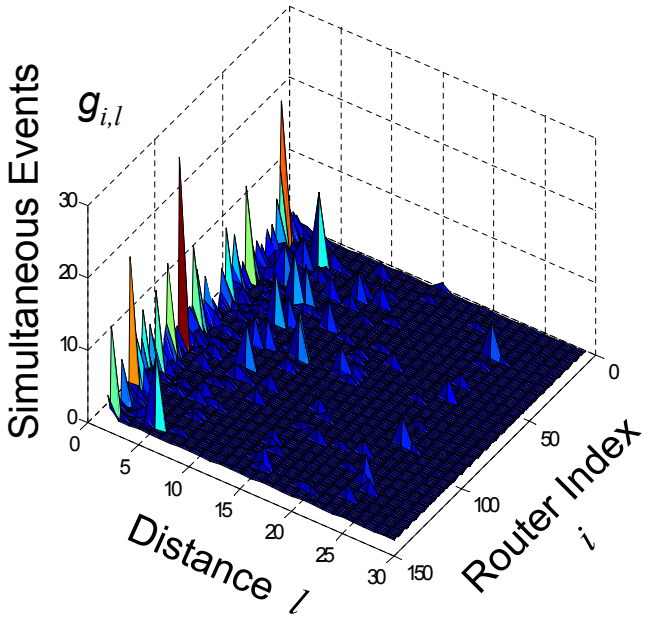


Figure 7. Simultaneous Down Events in Routers During 2007.

vectors are organized and plotted together, a 3D graph that shows the distribution of simultaneous down events will be obtained where each cell will have a value $g_{i,l}$ given by the next equation.

$$g_{i,l} = \sum_{j=1}^{n_i} \sum_{\forall x \neq i} \sum_{f=1}^{n_x} I\Big(t_{x,f} \in [\varepsilon_{i,j}] \wedge l = l_{i,x}\Big)^4 \tag{5}$$

Figure 7 shows the results regarding routers, where the x axis represents the *distance* l between network elements, the y axis contains the router index i, and in z is located the number of simultaneous events $g_{i,l}$. A similar 3D-plot for links is shown in Figure 8, where the same kind of patterns are observed. In this case the procedure described in Figure 6 was used as well, but taking into account the link considerations described in Section 2.

Figures 7 and 8 show a major concentration of events for the case of short distances. Specially, for the value 0, which means that the components are located inside the same data center.

There is a clear dominant presence of simultaneous events within components located in the same geographical area. However, this is more pronounced for the case

[4] $\wedge$ represents logical conjunction (AND operator).

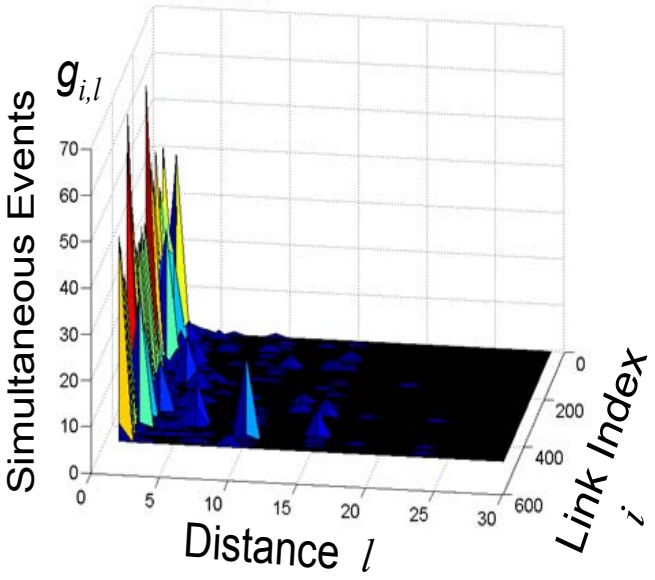


Figure 8. Simultaneous Down Events in Links During 2007.

of links. Finally, when simultaneous events occur in links, the number of elements involved is much larger.

4.3 Autocorrelation and Correlation Coefficient

The results obtained with the second method are very illustrative. They explain better the vertical patterns observed in the scatter plots. Nevertheless, we will apply a third method in order to verify and observe closely the dependence in failure events.

We analyze the vertical patterns of Figure 4 and 5, evaluating the correlation between failure processes. Here, the Pearson's correlation coefficient ρ will be calculated using two methods as is illustrated in Figure 9. In the first method, the state of a network element i will be modeled as a working/failed signal that is divided in slots of size Δ_1 during a defined period T, obtaining a binary vector $\mathbf{X_{i,k}}$ that takes value 1 if the element is up in the k interval ($k = 1, 2, ..., \frac{T}{\Delta_1}$) or 0 otherwise .

$$X_{i,k} = \begin{cases} 0 & t_{i,j} \leq (\Delta_1 \cdot k) < t_{i,j} + d_{i,j} \qquad \exists\, j:\ 1, ..., n_i \\ 1 & \text{otherwise.} \end{cases} \tag{6}$$

The value of Δ_1 is critical for this method. Therefore, we run several calculations with values between 1 and 100 seconds obtaining coherent and similar correlation values. Nevertheless, the huge size of the vectors obtained with this method is a considerable drawback for its application. On the other hand, bigger slots may be used

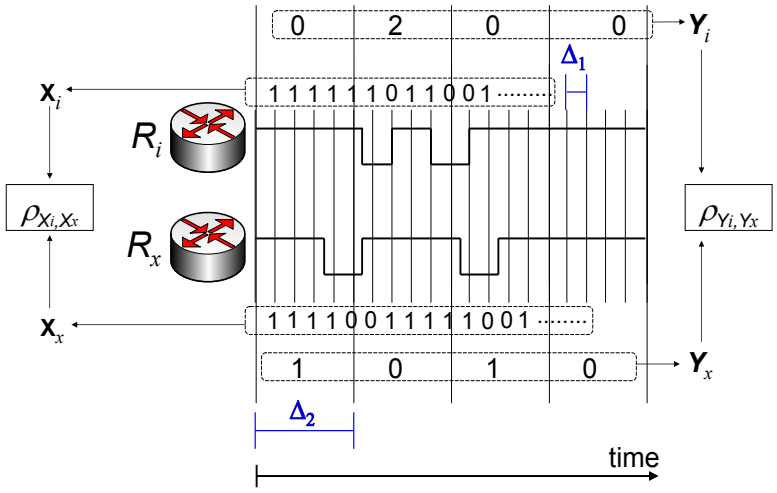


Figure 9. Methodology used to find ρ.

if a second method is applied in order to obtain a vector $\mathbf{Y_{i,k}}$ that count the number of down-events of component i within a given interval k ($k = 1, 2, ..., \frac{T}{\Delta_2}$) of size Δ_2.

$$Y_{i,k} = \sum_{j=1}^{n_i} I\Big(t_{i,j} \in [\Delta_2 \cdot k\ ,\ (\Delta_2 + 1) \cdot k]\Big) \tag{7}$$

In this study, we choose Δ_2 values between 5 and 60 minutes, reducing the size of the vectors considerably.

The correlation should be obtained for every pair of elements (i,x), using the Pearson's formula:

$$\rho_{Y_i,Y_x} = \frac{E[(Y_i - \mu_{Y_i})(Y_x - \mu_{Y_x})]}{\sigma_{Y_i}\sigma_{Y_x}} \tag{8}$$

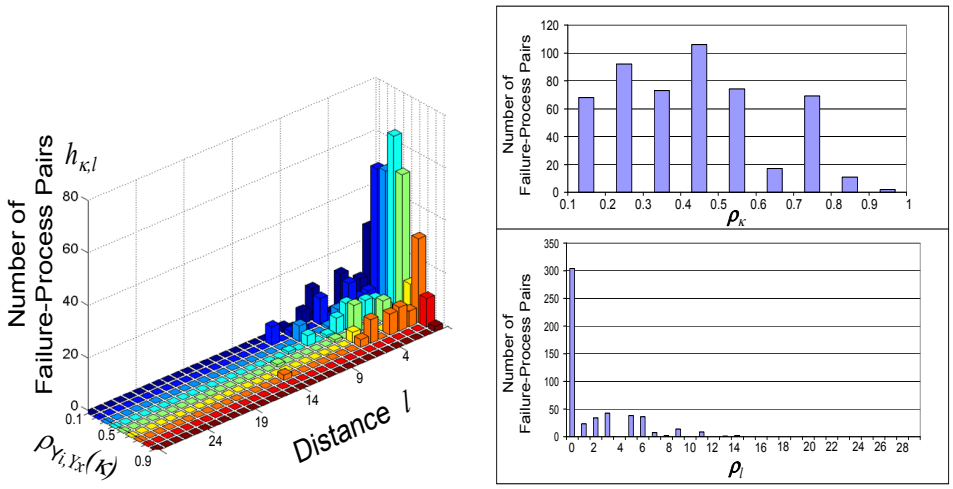
The obtained correlation values in approximately 96% of the cases were very close to 0. However, in our study, we put special interest on the values that were larger than 0.1.

Each evaluated pair has two features. First, the distance $l_{i,x}$. Second, the corresponding correlation coefficient ρ_{Y_i,Y_x} between the respective vectors obtained according to the procedure explained in Figure 9. Therefore, the use of a 3D plot is also appropriate for this case.

For this elaboration, the values of correlation will be grouped in nine discrete groups, defined by gaps $\rho_{Y_i,Y_x}(\kappa)$ as is described in the next equation.

$$\rho_{Y_i,Y_x}(\kappa) = \kappa \cdot 0.1 \leq \rho_{Y_i,Y_x} < (\kappa + 1) \cdot 0.1 \qquad \kappa = 1, ..., 9 \tag{9}$$

The correlation is evaluated on every vector-pair with distance l, and assigned to the respective $\rho_{Y_i,Y_x}(\kappa)$. A vector h_κ with the number of correlation values that belongs



(a) 3D plot for $\rho_{Y_i,Y_x}(\kappa)$.

(b) Marginal Plots for $\rho_{Y_i,Y_x}(\kappa)$.

Figure 10. Number of router-failure process pairs with $\rho_{Y_i,Y_x}(\kappa)$ larger than 0.1.

to $\rho_{Y_i,Y_x}(\kappa)$ per distance l is obtained. If all these vectors are organized and plotted together, a 3D graph that shows the distribution of ρ_{Y_i,Y_x} will be generated, where each cell will have a value $h_{\kappa,l}$ given by the next equation.

$$h_{\kappa,l} = \sum_{i=1}^{N} \sum_{\forall x \neq i} I\left(\rho_{Y_i,Y_x} \in [\rho_{Y_i,Y_x}(\kappa)] \wedge l = l_{i,x}\right) \tag{10}$$

Figure 10 shows the results for $\rho_{Y_i,Y_x}(\kappa)$, using Δ_2 equal to 60 minutes for failures in routers during year 2007.

Additionally, the 2-dimensional plots in Figure 10(b) help to visualize better how are distributed the different correlation coefficients calculated. The values illustrated in these marginal plots are calculated as follows:

$$\rho_\kappa = \sum_{\forall l} h_{\kappa,l} \qquad l = 0,...,30 \tag{11}$$

and

$$\rho_l = \sum_{\forall \kappa} h_{\kappa,l} \qquad \kappa = 1,...,9 \tag{12}$$

The results obtained with this method show again a clear dominant presence of correlated events within components located in the same geographical area.

On the other hand, we also studied the horizontal patters observed in Figures 4 and 5 through the formal evaluation of autocorrelation in individual components failures. This is made using a procedure where for each component i is obtained a vector θ_i that contains the time between failures during T as is described in Equation (13).

$$\theta_{i,c} = t_{i,c+1} - t_{i,c} \qquad c = 1,2,...(n_i - 1) \tag{13}$$

Using θ_i, we applied equation (14) to estimate the autocorrelation $\rho_{\theta_i,\theta_i}(\tau)$.

$$\rho_{\theta_i,\theta_i}(\tau) = \frac{\mathrm{Cov}(\theta_{i,c}, \theta_{i,c-\tau})}{\sigma^2_{\theta_i}} \tag{14}$$

In some cases, a high autocorrelation was found, as is shown in Figure 11(a), given that for most of the lags, the obtained values exceed the 95% confidence bounds that indicate acceptable values if the failure events were independent. These cases may be easily associated with the very pronounced horizontal patterns observed in the scatter plots. There are also some cases where medium levels of autocorrelation are observed in the sense that there are few lags that exceed the confidence bounds (Fig.11(b)). Finally, there are many failure processes that seem to be independent as may be observed in Figure 11(c). The important conclusion is that the presence of high and medium autocorrelation levels is not a negligible.

To conclude this section, we may say that after applying these three different methods, the obtained results show clearly that the distance among elements has a huge impact in the failure correlation. Nevertheless, there is a small portion from the possible correlated events that occur when they are geographically far. Although this portion is small, this finding is very interesting given that it was not expected.

Finally, it is important to clarify that the same patterns and behaviors were observed in all chosen periods among the available downtime logs (2001-2009), where many different network configurations may be present due to the dynamics of the network caused by the installations and changes of components, giving to our findings a wider validity.

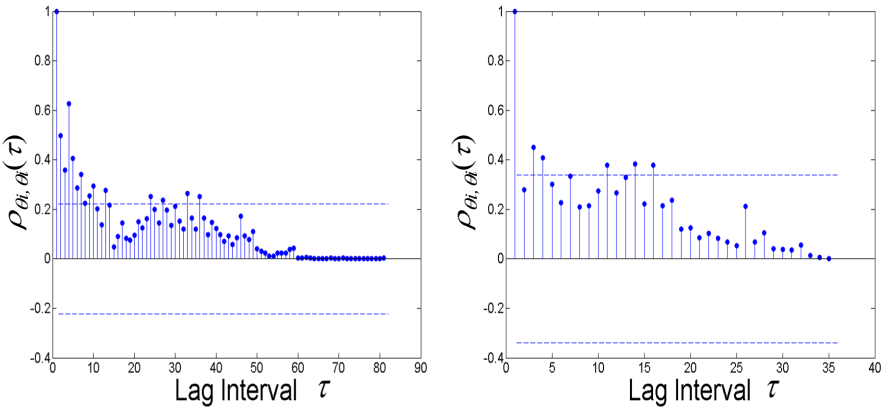
5. Conclusions and Future Work

This paper yields an improved insight into the failure processes at a real network. Correlation between failures are pronounced in both time and space. A main result obtained is that geographical distance has a significant impact. Therefore, this effect should be considered in dependability studies and network design.

The independence assumption commonly used to model network dependability is incorrect, at least for the case of the UNINETT backbone network. Nevertheless, we believe that this conclusion may be easily extended to a wider amount of networks.

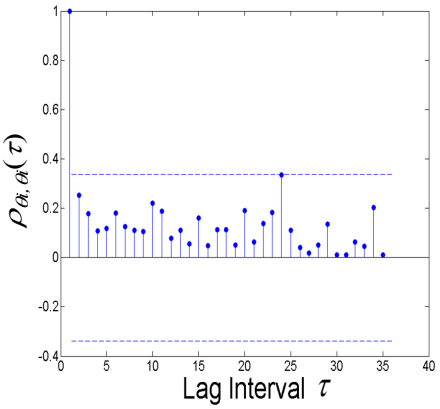
We found that there is a small portion of coincident failures that do not fit with the geographical location explanation. Therefore, a deeper study of this kind of events may be analyzed in future works.

The results shown in this paper have direct implications in network design and backup assignations techniques such as Shared Risk Link Group (SRLG) used to have robustness under single link failures, which states that the connections affected by one failure can not share any backup resource [RBS+01]. For instance the UNINETT topology shown in Figure 1 is vulnerable since Trondheim is an area that forms a "*bridge*" between southern and northern Norway. The "*bridge*" is well designed and has no single point of failure, but the geographical closeness increases the risk of



(a) High Autocorrelation.

(b) Medium Autocorrelation.



(c) Low Autocorrelation.

Figure 11. Autocorrelation in Failure Processes.

simultaneous failures. Another use of the results is to complement the theoretical models that have been developed for assess network reliability in presence of interdependence between the component failures e.g., [SK04] and [ND08]. These studies assume the existence of some correlation in failure events, but empirical information that describe such correlation is missed.

We have not looked into the potential causes of correlated failures, e.g., failures of the power grid, but we have a clear understanding that best practices are applied to avoid common cause failures. On the other hand, the *distance* concept used was based on geographical areas given that this makes easier the organization of information. Nevertheless, this is rather crude approach, which may be refined.

In this paper, there were described two initial methodologies to analyze the correlation of two different failure processes. Nevertheless, there were found some limits

given the size of the vectors obtained and the computational efficiency to use them. Therefore, new and innovative ways to obtain correlation values may be defined.

This work is an initial research that shows specific results in the analysis of dependability logs in the UNINETT network.

Acknowledgment

The authors would like to thank UNINETT for providing the log of failure and repair events of its backbone network for the period 2001-2009. Special help in the understanding of the logs was received from Jon Kåre Hellan.

References

[Cox67] David R. Cox. *Renewal Throry*. Methuen, 1967.

[CSKM07] Baek Young Choi, Sejun Song, George Koffler, and Deep Medhi. Outage analysis of a university campus network. *Proceedings of 16th IEEE International Conference on Computer Communications and Networks (ICCCN)*, pages 675 – 680, 13-16 Aug. 2007.

[InCM+02] Gianluca Iannaccone, Chen nee Chuah, Richard Mortier, Supratik Bhattacharyya, and Christophe Diot. Analysis of link failures in an IP backbone. *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurment (IMW)*, pages 237–242, 2002.

[KN10] Pirkko Kuusela and Illka Norros. On/Off process modeling of IP network failures. In *Proceeding of the IEEE/IFIP the 40th Annual International Conference on Dependable Systems and Networks (DSN)*, 2010.

[KNR09] Pirkko Kuusela, Ilkka Norros, and Pertti Raatikainen. Report on modelling the reliability of an ip-network and strategies for improving the reliability. Technical report, A report of the IPLU-II project, Jun. 2009. Available at http://iplu.vtt.fi.

[MIB+08] Athina Markopoulou, Gianluca Iannaccone, Supratik Bhattacharyya, Chen-Nee Chuah, Yashar Ganjali, and Christophe Diot. Characterization of Failures in an Operational IP Backbone Network. *IEEE/ACM Transactions on Networking*, 16(4):749–762, Aug. 2008.

[MVM02] Steven M. Matz, Lawrence G. Votta, and Mohammad Malkawi. Analysis of failure and recovery rates in a wireless telecommunications system. *Proceedings of the IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, pages 687 – 693, 2002.

[ND08] Maurizio Naldi and Giuseppe D'Acquisto. A normal copula model for the economic risk analysis of correlated failures in communications networks. *Journal of Universal Computer Science*, 14(5):786–799, 2008.

[RBS+01] Ramu Ramamurthy, Zbigniew Bogdanowicz, Shahrokh Samieian, Debanjan Saha, Bala Rajagopalan, Sudipta Sengupta, Sid Chaudhuri, and Krishna Bala. Capacity performance of dynamic provisioning in optical networks. *Journal of lightwave technology*, 19(1):40–48, Jan. 2001.

[SK04] Nozer D. Singpurwalla and Chung-Wai Kong. Specifying interdependence in networked systems. *IEEE Transactions on Reliability*, 53(3):401–405, 2004.

[UNI12] UNINETT. The Norwegian Research Network. Downtime Statistics. [online]. Available at: http://drift.uninett.no/downs/. 2012.

PAPER D

Guaranteeing Service Availability in SLAs; A Study of the Risk Associated with Contract Period and Failure Process

Andres J. Gonzalez and Bjarne E. Helvik

Proceedings of the IEEE Latin-American Conference on Communications LATINCOM-2009

Medellin, Colombia, September, 2009

GUARANTEEING SERVICE AVAILABILITY IN SLAs; A STUDY OF THE RISK ASSOCIATED WITH CONTRACT PERIOD AND FAILURE PROCESS

Andres J. Gonzalez, Bjarne E. Helvik
Centre for Quantifiable Quality of Service in Communication Systems
Norwegian University of Science and Technology,
Trondheim, Norway
{andresgm,bjarne}@q2s.ntnu.no

Abstract

Service Level Agreements (SLAs) are a common means to define the obligations of network/service providers and users in business relationships. The terms that define the guaranteed availability for a given period are important elements of these contracts. The appropriate selection of values is difficult due to the large number of variables involved, the complexities of the network and service provision, and the computational challenge posed by the transient solution. A common policy taken to solve it, is using the steady state availability as a reference. Nevertheless, this simplification may put on risk the contract fulfillment, as stochastic variation of the measured availability is significant over a typical contract period. This paper analyzes the relevance that the interval availability analysis has on SLAs, and provides suggestions to the network providers on the selection of adequate availability guarantees. The interval availability of unprotected and shared protected connections is studied under exponential and Weibull failure and repair distributions. It is observed that for a single path scenario, a small reduction of the guaranteed availability below the steady state value improves considerably the probability of meeting the requirements. The same situation was observed for connections with shared backup protection. However, performing this analysis in the transient domain is quite demanding. Hence, to simplify it, we propose the use of a *safeguard factor* to control the availability guarantee, using steady state values. For the Weibull distributed times between failures, where the shape factor is less than one, as observed in operational networks, the probability of meeting a guaranteed availability over a finite contract period decreases more radically than for the commonly assumed Poisson failure process. This increases the importance of making a transient analysis.

1. Introduction

Network operators and customers use Service Level Agreements (SLAs) to define a contracted QoS, where availability is a significant element. Violation of the agreed

value may have large performance, economic and reputational consequences for both parts.

This paper studies how to guarantee availability on SLAs, considering the transient working-failed behavior during a contract period, i.e., the distribution of the observed interval availability. Under Markov assumptions, the interval availability can be obtained by numerical methods using uniformization techniques as is presented in [MSW05, RS95, SEG86]. The analysis of the probability that a network operator meets the contracted interval availability α is of special interest. Such probability will be referred as *SLAs success probability*, and it will be evaluated in network connections, where links are exposed to failure and repair events.

This problem was first raised for general systems in [GT88] by Goyal and Tantawi. They observed that if α is larger than the *steady state availability* (A) of a service, the success probability decreases up to zero, which is quite intuitive. However, it was also shown that there is a considerable risk even when $\alpha < A$. In [MQWS06], it was suggested how to use the interval availability for economic planning to maximize the revenue from SLAs. A recent paper [MH09] studied the transient behavior under adaptive management strategies to control penalties and fairness.

Connections with shared paths protection are popular nowadays. They offer a considerably improvement in the availability, while they maintain an efficient resource utilization [CMH$^+$07, MH07], [MH08a]. On this issue, most of the works have modeled the availability through bounds that approximate the asymptotic solution. This approximation is needed to overcome the complexity involved in finding exact solutions, due to the dependencies that exists when links are shared [LLY09], [FTUF02], [HTH08], [HLS07], [MH08b]. This paper extends these results by analyzing the transient behavior of these scenarios. We found that the results offered by a conservative approximation are much safer and simpler to be obtained than the results obtained by using a precise estimation of the asymptotic availability.

During the last years, some studies have shown that the link interfailure and repair/recovery times may be more accurately modelled by a Weibull distribution than the more commonly used negative exponential distribution, due to the higher occurrence of very short and long times, e.g., [MIB$^+$08]. Hence, the success probability is analyzed under failure and repair processes with Weibull distributed inter-event times. An interesting relation between the Weibull shape parameter (β) and the reduction of the SLA success probability is observed. It highlights the importance of considering realistic, Weibull like, failure and repair time distributions in setting the interval availability guarantees in the SLAs.

Analytical/numerical tools for analyzing interval availability are found just for trivial systems (single elements), or systems with Markov properties. Hence, simulation is used to obtain the results, using Möbius [San08], [SO93]. The simulations were run with a relative confidence interval of 0.1 and a confidence level of 99% was obtained in all the scenarios.

To make easier the risk analysis, and reduce the SLA risk, a *safe guard factor*, σ, is introduced to allow the use of steady state availabilities, taking into account the stochastic transient effects posed by the finite duration of the contract.

This paper is organized as follows. In Section 2, issues related to the distribution of the interval availability and their relation to SLAs are introduced and discussed. Section 3 studies the success probability of connections under shared protections schemes, i.e., the probability than the observed availability will meet the SLA requirement. Section 4 presents the effect of having failure and repair/restore inter event times that follow a Weibull distribution. Section 5 discusses issues on how "safe" guarantees of availability in SLAs may be obtained. Finally, Section 6 concludes the paper.

2. SLA Success Probability

A *network connection* is defined as a group of interconnected links that provide end to end service. Maintaining it operational (up/working) is salient to offer a good quality of service. Its performance as a function of time can be modeled according to the random function $\hat{I}(t)$ defined as follows:

$$\hat{I}(t) = \begin{cases} 1 & \text{if the connection is working.} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Specifically on SLAs scenarios, the Interval Availability $\hat{A}(\tau)$ has to be considered. It is a stochastic variable that measures the percentage of time that the connection has been working during a defined period τ. $\hat{A}(\tau)$ can be described by the following equation:

$$\hat{A}(\tau) = \frac{1}{\tau}\int_0^{\tau} \hat{I}(t)\,dt. \tag{2}$$

When the transient behavior of a connection is studied, it is common to evaluate the expected interval availability $A(\tau) = E[\hat{A}(\tau)]$. For instance, Figure 1(a) shows that for a connection with independent link failures exponentially distributed with mean of one year, and repair times with an expectation of 12 hours, $A(\tau)$ converges in few weeks to the steady state availability A defined as:

$$A = \lim_{\tau\to\infty} A(\tau) \tag{3}$$

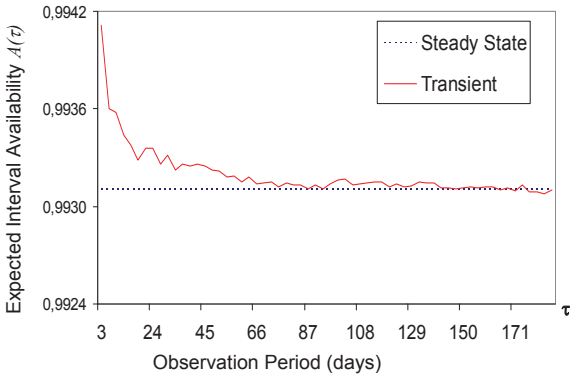
Nevertheless, the measure of $A(\tau)$ is not descriptive enough, since it does not consider the probabilistic variation that may have a critical influence on the connection quality.

From an SLA point of view, when the observed interval availability distribution is taken into account, and given an availability guarantee α, it is of special interest obtain the probability that the availability after some observation period τ will be larger or equal than the defined guarantee. This will be defined as the Success Probability:

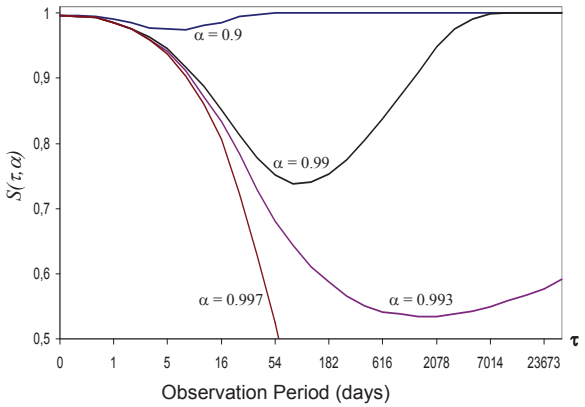
$$S(\tau,\alpha) = Pr[\hat{A}(\tau) \geq \alpha] \tag{4}$$

Additionally, the *risk* will be defined as the probability that the specified guarantee availability will not be met, which can be expressed as $1 - S(\tau,\alpha)$.

There are some numerical methods that may obtain $S(\tau,\alpha)$ through the use of uniformization techniques, assuming Markovian properties [MSW05, RS95, SEG86].



(a) Expected Interval Availability.



(b) Success Probability under different values of α for a system with $A = 0.9931514$.

Figure 1. Interval Availability Behavior.

Additionally, for a single item, the failure and repair processes may be regarded as an alternating renewal process when independence are assumed, obtaining interval availability results for general distributions [Cox67]. However, as we address compound systems under non-Markovian assumptions, i.e., Weibull failure and repair processes, $S(\tau,\alpha)$ is obtained via simulation.

The Success Probability as a function of the observation period was first discussed by Goyal et al. in [GT88]. It has basically two different behaviors, depending on whether the guaranteed value can be met in the asymptotic case or not. See Figure 1(b) for an illustration. In the first case, $\alpha \leq A$, it drops below one for a period, but it converges back to one. The time until it converges and the depth of the dip, depend on the ratio between the guarantee α and the asymptotic availability A. In the second case, $\alpha > A$, it decreases continuously until reach zero.

Taking into account those observations, in this paper is defined a *responsible promise* as the stipulated guarantee on the SLA that lies below A. The example in Figure 1(b) shows this for a network connection with five independent links, having Poisson failure processes with the expectation of one failure per year, and i.i.d. negative exponentially distributed down times with repair expectation of 12 hours (the path steady state availability is 0.99315). It is observed that the risk may become very large. This can be avoided if the contract period is very short or very large.

However, these safe observation periods are out of the range of typical SLA contract periods, and hance it is necessary to look for any other type of solution, like for instance, the use of guarantees smaller than the asymptotic availability. Note that the worst case occurs when α matches exactly the steady state availability, where we have a null recurrent stochastic process with a risk of 0.5 for an infinitely period.

3. Shared Protected Connections

To meet high availability requirements and to be able to guarantee it, telecommunications networks implement mechanisms that protect connections trough the reservation of additional resources that can be used when the primary fails. This includes the assignation of a dedicated principal path (W) and a shared backup path (B). We study the first and more demanding case, where backup links may be shared between several connections. This mechanism is known as *shared backup protection*, and it allows the combination of dependability improvement via protection and a more efficient resource utilization.

Under this approach, both paths (main and backup) must be failure disjoint. The connection, by default uses the primary path, but if it is affected by a failure, the backup path may be used instead, if it is working. However, as it is not dedicated, its availability also depends on the behavior of other connections.

One can say that the total availability of a given connection depends on the availability of the main paths of the other connections which share links on the backup. Hence, the availability prediction becomes more demanding than for the unprotected or dedicated path protection case. This is further complicated due to the connections in a network are continuously changing (istalled/uninstalled), so the set of "partners" sharing a backup link is dynamic [MH08b]. For this study we use a lower bound that approximates the asymptotic solution.

To predict the total asymptotic availability of a connection (C_n) that uses shared protection scheme, the unavailability of its main path is considered. It contains a number of links (L_i) with independent availabilities A_i, and the unavailability of the backup path with links (L_j) with availabilities A_j. They (links L_j) only can be used when all the working paths of the group of connections (C_s) that share that link are available. This yields.

$$A^{C_n} \geq 1 - (1 - \prod_{\forall i | L_i \varepsilon W} A_i).(1 - \prod_{\forall j | L_j \varepsilon B} A_j.[\prod_{\forall s | C_s} A_W^{C_s}]) \tag{5}$$

For the transient analysis, the problem becomes more complex. Therefore, it is more appropriate the use of simulation. In the case shown in Figure 2 two connections

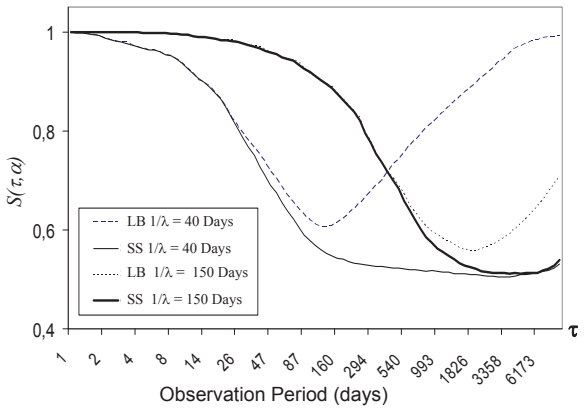


Figure 2. Success Probability with α =Lower Bound (LB) and $\alpha \approx$ Near to the Steady State (SS) under different links quality.

are considered, each with a dedicated principal path and an independent and disjoint backup path. Both backup paths have one link in common that have to be shared in case of failure. This link can be used by the first connection that need it, if it is not being used. Otherwise, the current connection that request the service has to wait until the working path of the other become again available, or if another link (different from the shared) on the backup of the connection that is currently using it, goes down, making useless the reservation of the shared link.

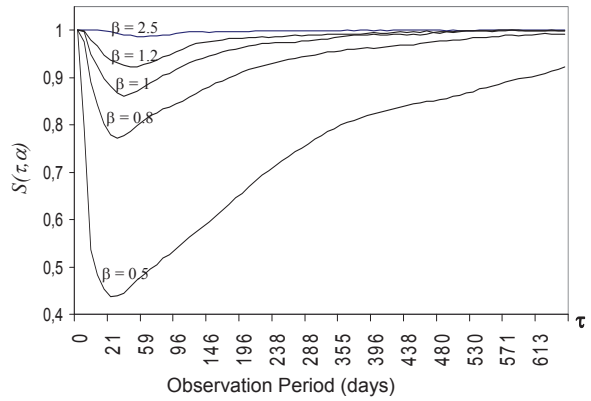
Figure 2 shows the results obtained in the simulation, for links with two different expected link interfailure times. First, using a guarantee α as close as possible to the steady state availability of the system. Second, α equal to the lower bound obtained from equation (5).

In general, the properties of $S(\tau, \alpha)$ remain for the shared protected scenarios as well.

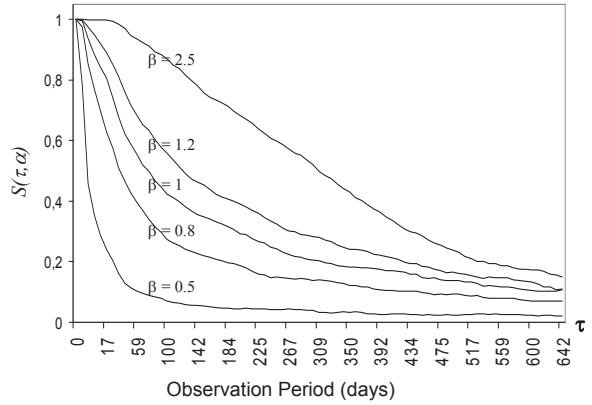
As was expected, according to the results in the previous section, the lower bound is always a safer decision for a network provider. Specially, the safety margin is more pronounced when the links do not have a good quality. Even though the lower bound is safer, from a SLA success probability point of view, the risk of not meeting a guarantee may be also significant.

4. Weibull Analysis

Measurements of operational systems show higher occurrence of very short and long link interfailure and repair/recovery times than what is properly described by a negative exponentially distribution, e.g., [MIB+08, UNI12a]. Hence, they are more accurately modelled by a Weibull distribution with a shape parameter, β, less than one, i.e., we assume i.i.d times T_x between events of type x, where x may be a failure or a repair of a link, and $P(T_x > t)$ is defined as:



(a) Success Probability under different Failure Link β values and $A \geq \alpha$.



(b) Success Probability under different Failure Link β values and $A < \alpha$.

Figure 3. $S(\tau, \alpha)$ Behavior under Weibull Processes.

$$P(T_x > t) = 1 - F_{T_x}(t) = e^{-(\frac{t}{\theta})^{\beta}}, \forall t \geq 0, \tag{6}$$

where θ is the scale parameter. It is seen that when $\beta = 1$, the distribution becomes exponential and the results obtained previously is a special case.

As above the success probability is obtained for unprotected and shared protected connections. The value of θ is set in order to have always an expected time between failures of one year, for all the values of β used. The repair times used were also Weibull distributed with the shape parameter less than one and an expectation equal to 12 hours for all cases. Hence, the asymptotic values are the same in all cases, and identical to those presented in Section 2.

Figure 3 shows some results, given a value of $\alpha = 0.98$ for a responsible promise and $\alpha = 0.997$ for the other case. One can observe that exponentially distributed link failure times ($\beta = 1$) may incur a smaller risk than the cases with a Weibull process with shape factor below one. Figure 3(a) shows that if the value of β is very small, the failure process may generate a very steep and large dip in $S(\tau, \alpha)$. Additionally, the time needed to return to an acceptable low risk may be longer than the exponential case as well as the typical contract periods, making less likely that the SLA agreement is met. When $A < \alpha$, one can observe from Figure 3(b) that the effect of β is also significant when the contract can not be met in the asymptotic case.

The effect of the shape of the repair distribution (its β parameter) has also been thoroughly investigated. We found that if the value of β is very small, the dip in $S(\tau, \alpha)$ is smoothed and therefore the risk may be reduced. This effect has inverse tendency as for the time between failures. However, we observe that the negative effects of the failure process are more significant.

The above findings show the importance of a proper characterization of the failure processes, due to their influence in the distribution of the observed interval availability, and in the risk of not meeting the availability contracted in an SLA.

5. Meeting Availability Guarantees

The analyzes made on the last sections show that the stochastic behavior of the interval availability may put the contract fulfillment at serious risk, and actions should be taken to counteract these effects.

Knowing that is very demanding for a network provider to have an SLA success probability equal to one, it must be planned a risk value that takes into account the commercial and technical costs implied. For this propose a guarantee value α^* should be specified in the SLA, to have a controlled risk.

To obtain this value, a *Safe-Guard Factor* σ is suggested as the correction value that multiply the steady state availability in order to obtain a desired success that take into account different consequences that imply the violation of the contracted parameters, i.e.

$$\alpha^* = \sigma \, . \, A \tag{7}$$

In this way is possible to assess the asymptotic solution A, and multiply it by σ to know the value to be promised in the SLA.

To have a better idea about the magnitudes implied to have an adequate SLA definition, the connection interval availability for a fixed observation period, on paths with multiple links, that follow independent failure and repair Weibull distributed processes was simulated.

Figure 4 shows the results obtained for the Safe Guard Factor tendency in an SLA with one year duration, using different network conditions. The values of β shown on the figure correspond to the failure process.

Analyzing the effects of the link quality on the safeguard factor is important. In Figure 4 can be also appreciated that in order to have correction values very close to

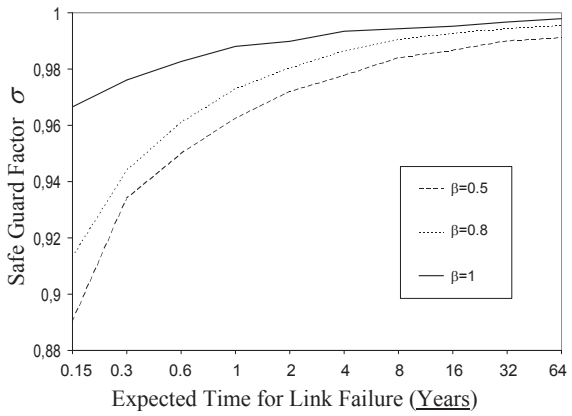


Figure 4. Safe-Guard Factor needed to obtain a 99% Success Probability.

one (small deviation from A), it is necessary to have links with extremely low failure rate (high expected time for link failure). In the practice, obtain a link with those characteristics may imply a huge effort. In most of the existing networks, the deviation that put σ may be considerable.

A particular example that may illustrate better the effect that the use of the Safe-Guard Factor has in the SLA specification may be obtained, assuming a contract for a given connection defined for a fixed observation time of one year, where it is previously known that it's steady state availability A is 0.99452 (expected cumulated uptime of 363 days). If the links failures on the network follow a weibull distribution with β equal to 0.5 and expected time to link failure of one year, using Figure 4, can be obtained that $\sigma = 0.96$, and the optimal α^*, to have an SLA success probability of 0.99 is equal to 0.956 (around 16 days of expected cumulative downtime per year). Hence, if this is analyzed under a business model, it is possible that the commercial image may be considerably affected. On the other hand, if the provider offers guarantees that are not fulfilled, the reputation consequences may be even worst in the long time. This dilemma raises the challenge of having a model able to balance this two important tendencies.

There are many variables that have an implication on the SLA success probability. From the business point of view, there are three facts that are relevant for Network Providers regarding the network dependability and SLA specification. The first is the cost implied in the installation, configuration and assignment of network resources in order to provide high quality connections, which includes the selection of a high reliable infrastructure and the assignation of resources with the appropriate protection. The second is the commercial impact that may have to offer a guaranteed availability that may be improved for the competitors (as was shown before, the difference may be of several days). Finally, the third one is the cost generated by the refund payment caused when the requirement are not met, additionally to the image impact that is generated when the customers assimilate those events as a weakness.

One of the ideas behind the Safe-Guard factor and all the other analysis made previously; is to provide tools in order to have a complete model, able to include the facts already mentioned, in order to offer a good quality service, and to obtain the best profit. The precise setting of the model variables may differ for each provider. However, they can be easily obtained, based on the particulars technical and marketing conditions that the network companies have.

6. Conclusions

This paper demonstrates the importance of taking the distribution of the interval availability into account for defining the SLA.

We show that when a provider offers an availability equal to the steady state availability, the dip in $S(\tau, \alpha)$ is extended for an infinite period with a value of 0.5. To have a guarantee in the SLA that is likely to be met, it must be below the steady state value. For the case of shared protection, it was confirmed that the lower bound obtained in previous works is sufficient to decide an appropriate guarantee α. It was shown that $S(\tau, \alpha)$ has similar behavior under Weibull failures and repair process, but additionally, we found that the shape parameter of the failure process has a major impact on the SLA success. For values of β less than 1, as found in operational data, the risk increases considerably fast and the duration at the high risk period is larger than the typical SLA contract period.

Finally, we show a procedure to control the SLA success probability, using easy implementable procedures.

References

[CMH+07] Piotr Cholda, Anders Mykkeltveit, Bjarne E. Helvik, Otto J. Wittner, and Andrezej Jajszczyk. A survey of resilience differentiation frameworks in communication networks. *IEEE Communications Surveys & Tutorials*, 9(4):32–55, Fourth Quarter 2007.

[Cox67] David R. Cox. *Renewal Throry*. Methuen, 1967.

[FTUF02] Andrea Fumagalli, Marco Tacca, Ferenc Unghvary, and Andras Farago. Shared path protection with differentiated reliability. In *Proceeding of the IEEE International Conference on Communications (ICC)*, volume 4, pages 2157–2161, 28 Apr.–2 May. 2002.

[GT88] Ambuj Goyal and Asser Tantawi. A measure of guaranteed availability and its numerical evaluation. *IEEE Transactions on Computers*, Volume 37, Issue 1:25 – 32, 1988.

[HLS07] Changcheng Huang, Minzhe Li, and A. Srinivasan. A scalable path protection mechanism for guaranteed network reliability under multiple failures. *IEEE Transactions on Reliability*, 56(2):254–267, Jun. 2007.

[HTH08] Pin-Han Ho, J. Tapolcai, and A. Haque. Spare capacity reprovisioning for shared backup path protection in dynamic generalized multi-protocol label switched networks. *IEEE Transactions on Reliability*, 57(4):551–563, Dec. 2008.

[LLY09] Hongbin Luo, Lemin Li, and Hongfang Yu. Routing connections with differentiated reliability requirements in WDM mesh networks. *IEEE/ACM Transactions on Networking*, 17(1):253–266, Feb. 2009.

[MH07] Anders Mykkeltveit and Bjarne E. Helvik. Provision of connection-specific availability guarantees in communication networks. In *Proceedings of the IEEE 6th International Workshop on Design of Reliable Communication Networks (DRCN)*, Oct, 2007.

[MH08a] Anders Mykkeltveit and Bjarne E. Helvik. Comparison of schemes for provision of differentiated availability-guaranteed services using dedicated protection. In *Proceeding of the IEEE Seventh International Conference on Networking (ICN)*, Apr. 2008.

[MH08b] Anders Mykkeltveit and Bjarne E. Helvik. On provision of availability guarantees using shared protection. In *Proceeding of the IEEE/IFIP 12th Conference on Optical Network Design and Modelling (ONDM)*, Mar. 2008.

[MH09] Anders Mykkeltveit and Bjarne E. Helvik. Adaptive management of connections to meet availability guarantees in SLAs. In *Proceeding of the IFIP/IEEE International Symposium on Integrated Network Management, IM. Mini-Conference*, Jun. 2009.

[MIB+08] Athina Markopoulou, Gianluca Iannaccone, Supratik Bhattacharyya, Chen-Nee Chuah, Yashar Ganjali, and Christophe Diot. Characterization of Failures in an Operational IP Backbone Network. *IEEE/ACM Transactions on Networking*, 16(4):749–762, Aug. 2008.

[MQWS06] Darli A. Mello, G. Quiterio, Helio Waldman, and Dominic A. Schupke. Specification of SLA survivability requirements for optical path protected connections. In *Proceeding of the National Fiber Optic Engineers Conference on Optical Fiber Communication Conference (OFC)*, page 3pp., 5–10 Mar. 2006.

[MSW05] Darli A. Mello, Dominic A. Schupke, and Helio Waldman. A matrix-based analytical approach to connection unavailability estimation in shared backup path protection. *IEEE Communications Letters*, 9(9):844–846, Sep. 2005.

[RS95] Gerard0 Rubino and Bruno Sericola. Interval availability analysis using denumerable markov processes: application to multiprocessor subject to breakdowns and repair. *IEEE Transactions on Computers*, Volume 44, Issue 2:286 – 291, Feb. 1995.

[San08] William H. Sanders. *Mobius Manual. Version 2.2.1. [online]. Available at: https://www.mobius.illinois.edu/manual/MobiusManual.pdf*, Nov. 20, 2008.

[SEG86] De Souza E Silva and H.R. E. Gail. Calculating cumulative operational time distributions of repairable computer systems. *IEEE Transactions on Computers*, Volume: C-35, Issue: 4:322–332, Apr. 1986.

[SO93] William H. Sanders and II Obal, Douglas. Dependability evaluation using UltraSAN. In *Proceeding of the Twenty-Third International Symposium on Fault-Tolerant Computing FTCS-23. Digest of Papers*, pages 674–679, 22–24 Jun. 1993.

[UNI12] UNINETT. The Norwegian Research Network. Downtime Statistics. [online]. Available at: http://drift.uninett.no/downs/. 2012.

PAPER E

SLA Success Probability Assessment in Networks with Correlated Failures

Andres J. Gonzalez and Bjarne E. Helvik

ELSEVIER Computer Communications Journal.

(ACCEPTED)

SLA SUCCESS PROBABILITY ASSESSMENT IN NETWORKS WITH CORRELATED FAILURES

Andres J. Gonzalez, Bjarne E. Helvik
Centre for Quantifiable Quality of Service in Communication Systems
Norwegian University of Science and Technology,
Trondheim, Norway
{andresgm,bjarne}@q2s.ntnu.no

Abstract

Service Level Agreements (SLAs) are used to define obligations between network/service providers and customers in business relationships. The terms that define the guaranteed availability for a given period are fundamental to these contracts. The appropriate selection of the availability to be promised is still an open challenge for network operators due to: i) SLAs are defined for finite periods, and hence the stochastic properties of the availability have to be considered. ii) Real operational networks have not the Markovian properties. iii) The way that correlation affects the interval availability in operational networks is unknown. In this work, we show the impact of dependent failures on SLAs, based on operational failure data obtained from the UNINETT network. Using these data, we simulate the behavior of network connections that use shared backup protection. We evaluate the SLA success probability using two different methods. First, we apply trace driven simulation combined with random circular shifting. Second, we develop a model that uses Monte Carlo techniques. This approach includes the characterization of up and down times of each network component and the use of a model that generates correlated samples based on fitted marginal distributions. Finally, we analyze the probability density function of the interval availability for different observation periods under independent and correlated failures.

1. Introduction

Network operators and customers use Service Level Agreements (SLAs) to define a contracted QoS where availability is a significant element. The violation of the agreed value may have economic and reputation consequences for both parts. Under this scenario, a natural question for network operators is: How to assess the availability level that my network is able to provide? In order to answer this question, many factors have to be considered. First of all, it is important to notice that the availability offered to a connection during a contract (SLA) is a stochastic variable. In this case, the assessment of expected values, assuming steady state conditions, will provide misleading estimates that may dramatically increase the risk of failing the SLA. Therefore, the study of the entire distribution of the interval availability is necessary.

Under Markovian assumptions, the interval availability can be obtained by numerical methods using uniformization techniques, as is presented in [RS95].

The probability that a network operator meets the contracted interval availability α can be calculated directly from the interval availability Probability Density Function PDF as $P(\alpha \leq \text{Interval Availability} \leq 1)$, and it will be referred as *SLAs success probability*. This concept was first raised for Markovian systems in [GT88] by Goyal and Tantawi. They observed that if α is larger than the *steady state availability* (A), the success probability decreases continuously. However, they also showed that there is a considerable risk even when $\alpha < A$. Real operational networks do not follow Markovian assumptions. The inter-failure times are not exponentially distributed, neither any restoration time. In addition, the failure processes of two or more network components may be correlated. In such scenario, the assessment of the interval availability and the SLA success probability represents an open challenge that needs to be addressed. The way of addressing that challenge is the main point of this paper.

The work presented in [XTMM11] highlights the importance of assessing the SLA risk in WDM mesh networks, and the dangers implied by dealing only with steady state probabilities. They developed a method to assess the SLA risk when the stochastic properties of the networks are partially unknown. However, they assume independence between failure processes, Poison failure arrivals, and they do not consider the existence of overlapping failures.

Analysis of real network failure processes is mandatory in order to get the appropriate information for availability dimensioning and to deal with the risks associated with SLAs. In spite of this, for a number of reasons, among them that failures of their networks are not what operators like to have exposed in a competitive commercial marketplace, the access to such failure log information is very limited. A study of the failure behavior in an operational backbone network is reported by Iannaccone et al. [InCM$^+$02]. They examine the frequency and duration of failure events and discuss various statistics, like the distribution of inter-failure times and distribution of link failure durations. This work was continued by Markopoulou et al. in [MIB$^+$08], where failures and repairs in the Sprint IP backbone Network are classified and analyzed.

The objective of this paper is to evaluate the SLA success probability of network connections that use shared backup protection, based on operational data obtained from the UNINETT's network management system [UNI12b]. For the assessment, we use trace driven simulation *(TDS)* and Monte Carlo methods, taking into account dependencies between events. To the authors' knowledge, this is the first time that a study of the SLA success probability is made based on real correlated failure processes.

Our first approach (TDS) uses filtered ON/OFF processes measured during two years. We extend the number of observations using randomly generated circular shifts. With this method, we capture directly the properties of failure and repair processes. Additionally, when correlation is detected, the circular shift is made in blocks. This approach is a bootstrapping technique to gain more information out of the logged network behavior. Nevertheless, in order to obtain a more accurate estimation of the success probability, we propose the use of a Monte Carlo method that generates failure and repair times based on fitted models. In addition, the correlation is handled by

using the Marshall and Olkin copula proposed in [MO67]. Finally, we analyze the behavior of the probability density function of the interval availability for different observation periods.

This paper is organized as follows. In Section 2, issues related to the distribution of the interval availability and their relation to SLAs are introduced and discussed. In Section 3, the UNINETT's IP backbone network and the information collection method are presented. Section 4 shows the simulation setup and the criteria used in order to allocate connections in the UNINETT's network. Section 5 describes the methods used to estimate dependencies in failure and repair processes. Section 6 studies the success probability in connections that use shared backup path protection, using trace driven simulations. Section 7 describes the Monte Carlo approach used to evaluate the success probability and show the evolution of the probability density of the interval availability with the interval duration. Finally, Section 8 concludes the paper.

2. SLA Success Probability

A *network connection* is a group of interconnected routers and links that provide end to end service. Maintaining it operational (up/working) is salient to offer a good quality of service. Its performance as a function of time can be modeled with the random process $O(t)$ defined as follows:

$$O(t) = \begin{cases} 1 & \text{If the connection is working.} \\ 0 & \text{Otherwise.} \end{cases} \tag{1}$$

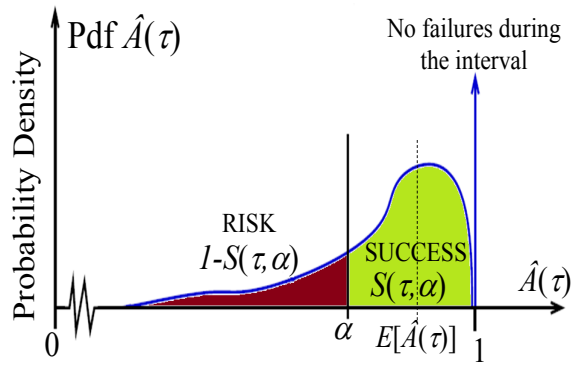
The Interval Availability $\hat{A}(\tau)$ is an important element of the SLA. $\hat{A}(\tau)$ is a stochastic variable that measures the time that the connection has been working during a defined period τ, i.e.:

$$\hat{A}(\tau) = \frac{1}{\tau}\int_0^{\tau} O(t)\,dt. \tag{2}$$

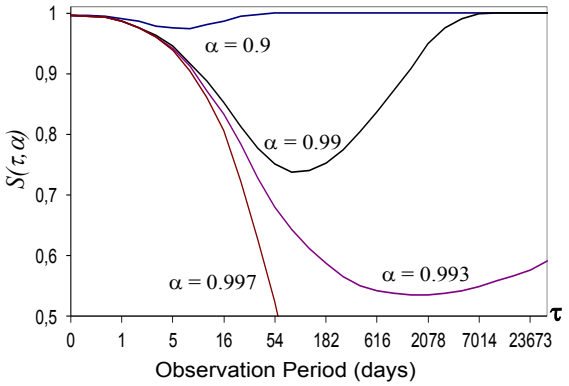
When the transient behavior of a connection is studied, the expected interval availability $A(\tau) = E[\hat{A}(\tau)]$ is usually evaluated, given that the first moment of a random variable is simpler to obtain than the entire probability distribution. $A(\tau)$ converges always after a long period to the steady state availability A defined as:

$$A = \lim_{\tau\to\infty} A(\tau) \tag{3}$$

When an SLA is defined, the provider promises an availability α for a given period τ (the duration of the contract). Under this scenario, to know the probability that the availability after some observation period τ will be larger than or equal to the defined guarantee is crucial. In this case, the evaluation of the expected interval availability is not enough for the estimation of such probability, and hence the entire probability distribution has to be considered. In this paper, the SLA Success Probability is defined as follows:



(a) Interval availability.



(b) Success probability vs contract period under different values of α for a connection path with $A = 0.99315$.

Figure 1. Interval availability and the success probability.

$$S(\tau,\alpha) = Pr[\hat{A}(\tau) \geq \alpha] \tag{4}$$

Additionally, the *risk* will be defined as the probability that the specified availability α will not be met, which can be expressed as $1 - S(\tau,\alpha)$. Fig. 1(a) shows the general shape of the probability density function (PDF) of the interval availability and how the risk and the success of the SLA can be estimated from this information.

The Success Probability as a function of the observation period was first discussed by Goyal et al. [GT88]. It has basically two different behaviors depending on whether the guaranteed value can be met in the asymptotic case or not. In the first case, $\alpha \leq A$, $S(\tau,\alpha)$ drops below one for a period, but converges back to one. In the second case, $\alpha > A$, it decreases continuously to zero. In order to illustrate this effect, we simulate the behavior of a network connection, evaluating the success probability for different values of τ. Fig. 1(b) shows the behavior of $S(\tau,\alpha)$ for a network connection with five independent links having Poisson failure arrivals with expectation of one year, and i.i.d. negatively exponentially distributed down times with expected repair time of

12 hours, obtaining a path steady state availability of 0.99315. From Fig. 1 one can observe that the shape of the interval availability PDF changes considerably with the duration of the SLA. In Section 7.3 the evolution of the probability density of $\hat{A}(\tau)$ with the increase of τ will be described.

$S(\tau,\alpha)$ converges to one or zero after certain value of τ. The convergence speed depends on the ratio between the guarantee α and the asymptotic availability A, as well as on the burstiness of the failure and repair processes that affect the network devices. For instance, in [GH09], a study of $S(\tau,\alpha)$ under Weibull distributed time to failure and time to repair processes shows that the risk increases considerably faster in both cases ($\alpha \leq A$, $\alpha > A$) when the shape parameter is shorter than one.

The accumulated down time over τ $t(\tau)$ is associated with $\hat{A}(\tau)$ as: $t(\tau) = \tau[1-\hat{A}(\tau)]$, where $\Omega(\tau,t)$ will be defined as the CDF of $t(\tau)$. A general expression for $\Omega(\tau,t)$ was derived by Takács in [Tak57] as follows

$$\Omega(\tau,t) = \sum_{n=0}^{\infty} H_n(t)[G_n(\tau-t) - G_{n+1}(\tau-t)] \tag{5}$$

where the failure and repair processes are described by i.i.d. up and down times with CDF $G(t)$ and $H(t)$ respectively, and the subindex n represents the $n-$fold Stieltjes convolution of a given function.

Equation (5) characterizes a problem with general distributions. For the case of failure and repair processes negatively exponentially distributed, a complete result was obtained by Takács as

$$\Omega(\tau,t) = e^{-\lambda(\tau-t)}\left[1+\sqrt{\lambda\mu(\tau-t)}\int_0^{\infty} e^{-\mu y}\sqrt{y}I_1(2\sqrt{\lambda\mu(\tau-t)y})dy\right] \tag{6}$$

where λ and μ are the respective failure and repair rates and I_1 is the Bessel function of order 1. However, $\Omega(\tau,t)$ is difficult to compute for other kind of distributions due to: i) $G(t)$ and $H(t)$ represent the CDF of up and down times of the entire connection which depends on the corresponding failure and repair distributions of all the network elements involved in the connection. ii) To obtain the $n-$fold Stieltjes convolution of a generally distributed function is complex. As we will explain in detail in Section 7.1, operational networks do not follow Markovian assumptions but present more complex distributions in the failure/repair processes such as Weibull, gamma or empirical distributions that can not be fitted to any simple type of parameterized function. In addition, the failure/repair processes are not independent. Therefore, $S(\tau,\alpha)$ will be assessed via simulation using trace driven and Monte Carlo methods.

3. UNINETT Network Description

UNINETT [UNI12b] is the network that connects universities, colleges and research institutions in Norway. The core of the network interconnects the main norwegian cities through optical fiber connections of 10 and 2.5 Gigabit per second (Gbps)

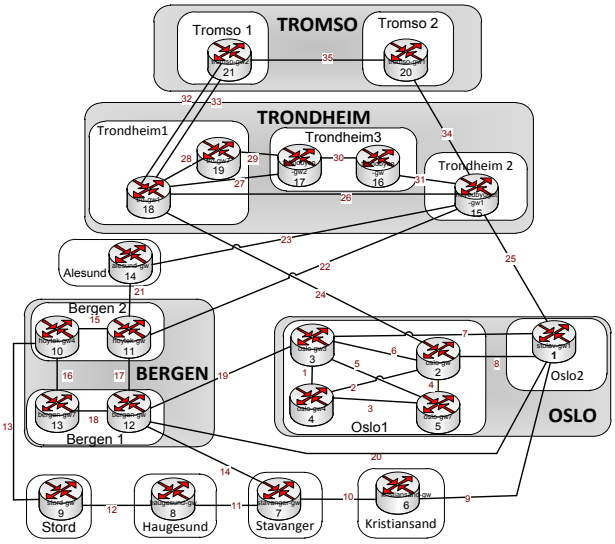


Figure 2. UNINETT's core network.

forming rings to ensure that the loss of a single link does not cause any loss of connectivity.

We are interested in studying the failure and repair processes that affect the core network. For this reason, in our study, we select the subset of connections that belongs to the UNINETT's backbone (Fig. 2).

The failure logs were obtained through a centralized network management system controlled from the UNINETT NOC (Network Operations Center) in Trondheim. It registers irregularities in the network at the IP level. The analysis performed in this paper is based on logs from January 2008 to December 2009. This period is used since the core network did not undergo significant changes neither in topology nor in equipment. Router's and link's failures are registered with a precision of seconds and the data collection method follows SNMP standards that enable the detection of anomalies in the network operation through the use of polling and trap mechanisms.

We obtain filtered $O_i(t)$ processes from the original raw log files for each network component i via PERL scripts. In addition, we verify the obtained information through the analysis of traffic logs and corporate quality assurance reports.

The studied backbone is operated using WDM technology and various brands routers. Full details about the UNINETT network can be found in [UNI12b].

3.1 Causes of Failures

From the failure logs, we observed that 45% of the downtimes are shorter than 5 minutes. These short-time failures usually are self-repaired or repaired by a simple logical restart e.g. shut/no shut. In principle, this kind of failures can be associated to transient software problems or short overload periods, but from the UNINETT

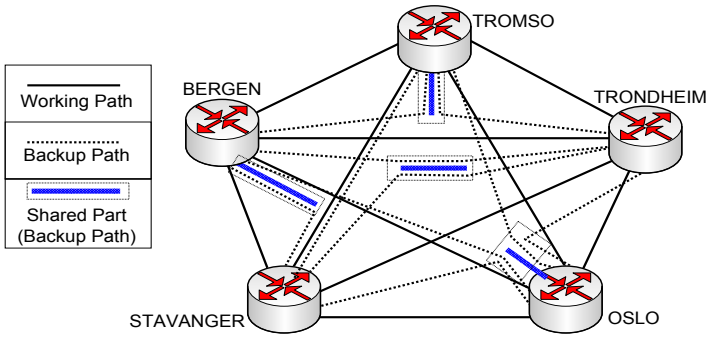


Figure 3. Connections set.

operation point of view there is not detailed register of the possible cause of the problem (if the failure is not repetitive).

The data collected automatically by the network management platform may be complemented by additional information registered by manual operations made by troubleshooting engineers, who identify the causes of the failures that hit the network. For reasons out of the scope of this paper, there is a percentage of failures larger than 5 minutes with unknown reason. In this paper, we provide information about the reason of the remaining failures (known reason and larger than 5 minutes) based on the troubleshooting reports, obtaining the following classification.

- 33%: Hardware Failure: Card problems, router problem, hardware malfunction, cable degradation.
- 31%: Electricity Failure: UPS, electricity flipping, other.
- 29%: Configuration Problems: Human Intervention, software problems.
- 7%: Fiber Cut. Digging or cuts generated by external agents.

As stated before, this classification does not include the unknown failures, but for the sake of information, we consider illustrative to present these results.

4. Simulation Setup

The success probability is evaluated using Discrete Event Simulation through a platform developed in DEMOS [Bir03].

We regard a set of virtual connections superimposed on the UNINETT's core network (Fig. 2). These connections use shared backup protection which is a standard strategy applied to meet dependability requirements with a high probability and to use efficiently the network resources. We need to implement a mechanism that establishes N connections C_n $(n = 1, 2..., N)$ composed by working (W) and backup (K) paths among the nodes of the main Norwegian cities, obtaining the fully connected network shown in Fig. 3.

The paths were found using the algorithm proposed in [GH10]. The $S(\tau,\alpha)$ assessment includes all the connections shown in Fig. 3 where all backup paths share resources with other connections defined by the sharing group G_n^s.

The obtained working and backup paths are disjoint, avoiding the existence of single point of failure. Our solution also considers the Shared Risk Link Group (SRLG) policy which is widely used by network designers when reliable connections have to be provided. This policy is used to prevent that connections affected by one failure in their working path will share resources in the backup path e.g., if the failure in a link affects the working path of connection n and m the backup paths of n and m have to be disjoint.

The algorithm in [GH10] needs as input a cost on each link (used for routing) and the steady state availability A_i of each network component i. The costs used on links are based on operational values set by UNINETT and the A_i values were estimated empirically from the failure logs.

The obtained working paths are composed of devices i with respective values A_i. The availability A_n^W of the working path of a given connection n is calculated using the following equation:

$$A_n^W = \prod_{\forall i \in W_n} A_i \tag{7}$$

Using this information, the total availability of a connection with shared backup A_n^S (including W and K) may be approximated by a lower bound as follows:

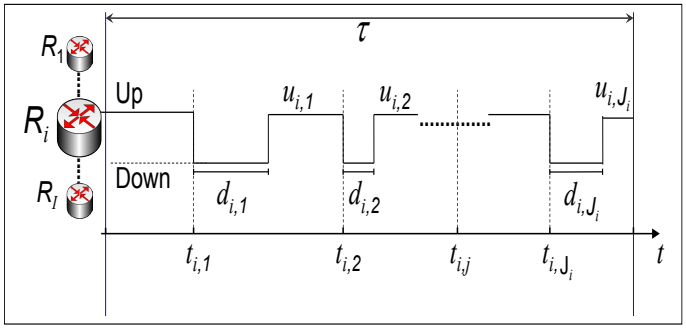
$$A_n^S \geq 1 - \left(1 - \prod_{\forall i \in W_n} A_i\right) \cdot \left(1 - \prod_{\forall i \in K_n} A_i \cdot \prod_{\forall C_x \in G_n^s} A_x^W\right) \tag{8}$$

The values obtained from equation (8) will be used in the simulation for the definition of availability guarantees where $\alpha \leq A_n^S$ (as defined in Section 2).

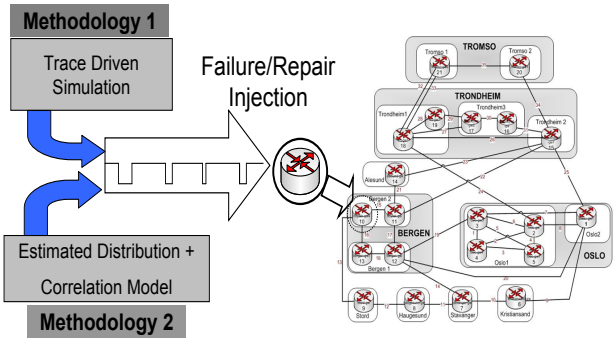
An interesting observation regarding equation (8) is that it defines a lower bound for the connection availability in steady state under independence assumptions. Therefore, in the next sections we will show the effects of those assumptions when transient and non independent scenarios are taken into account.

The simulation will handle the incoming failures and repairs in the following way: When a failure affects a working path W_n, the connection stays operational by using the backup path K_n if it is available, inhibiting at the same time the backup paths of all other connections that belong to G_n^s. A connection is registered as non operational when a failure affects a working path and the backup path is inhibited, or when a failure affects the backup path when it is being used.

Failure events will be considered during τ, where I network components are regarded. Each device i ($i = 1,2,....,I$) has an operational state that may be described by an ON/OFF process $O_i(t)$ as illustrated in Fig. 4(a). The jth failure of device i occurs at time $t_{i,j}$, where the downtime duration is denoted by $d_{i,j}$ and J_i is the total number of failures of device i during τ ($j = 1,2,....,J_i$). After the repair, the time when the device is working properly before a new failure occurs will be defined as uptime, and it will be denoted as $u_{i,j}$; i.e. $u_{i,j+1} = t_{i,j+1} - t_{i,j} - d_{i,j}$.



(a) On-Off processes of network components.



(b) Simulation methods.

Figure 4. Simulation setup.

Finally, failures and repairs are introduced in the DEMOS platform. First by trace driven simulation. Second by Monte Carlo methods. Fig. 4(b) illustrates the simulation methods.

5. Dependencies Between Failure Processes

When a failure affects a network device, it is possible that other devices are affected at the same time. These events may occur by coincidence, but in most of the cases they happen due to the existence of correlation between the failure processes of the different affected devices. The reason of these correlation may be among others due to the existence of a common cause of failure, overload in some devices after a failure, or due to network design problems. A preceding study [GHHK10] shows that the failures of the network components in the UNINETT network are correlated. In order to assess the described correlation, in [GHHK10] was defined a mechanism that captures the correlation of failure processes, but does not take into account the time spent in the repair processes. We extend that study by the implementation of the following mechanism:

We will evaluate the correlation between different failure/repair processes. Here the Pearson's correlation coefficient ρ will be calculated using the method illustrated

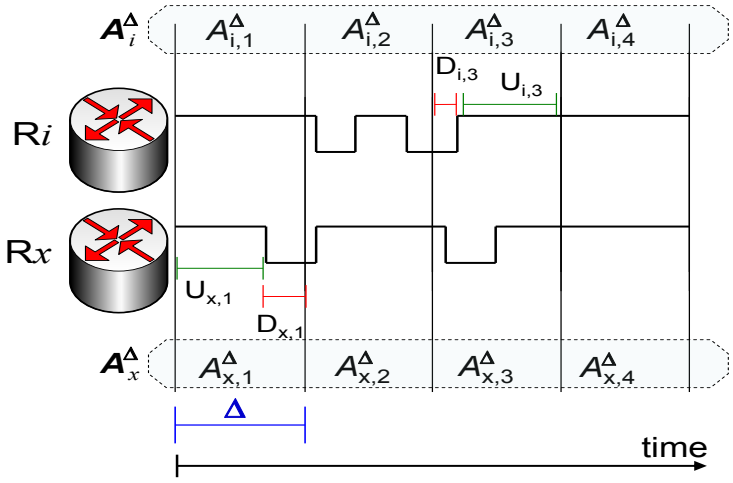


Figure 5. Methodology used to find ρ.

in Fig. 5. For this, the measured state ($O_i(t)$) of a network element i is divided in time slots of size Δ during the observation period τ. The idea is to evaluate the up time $U_{i,k}$ and down time $D_{i,k}$ in order to calculate the availability $A_{i,k}^\Delta$ on each slot k ($k = 1,2,..,\frac{\tau}{\Delta}$), as the next equation states:

$$A_{i,k}^\Delta = \frac{U_{i,k}}{\Delta} \tag{9}$$

For the implementation of this mechanism, we set Δ equal to 60 minutes (reducing the size of the vectors proposed in [GHHK10]). Then, we calculate the Pearson's correlation coefficient of every pair of vectors (A_i^Δ,A_x^Δ) using the following equation.

$$\rho_{A_i^\Delta,A_x^\Delta} = \frac{E[(A_i^\Delta - \mu_{A_i^\Delta})(A_x^\Delta - \mu_{A_x^\Delta})]}{\sigma_{A_i^\Delta}\sigma_{A_x^\Delta}} \tag{10}$$

where $\mu_{A_i^\Delta}$ and $\sigma_{A_i^\Delta}$ are the mean value and variance of vector A_i^Δ respectively.

We obtained ρ values close to zero in approximately 96% of the analyzed pairs. Nevertheless, we will focus on the remaining cases where ρ is larger than 0.1. We evaluate the relation between network components with high correlation coefficient ($\rho > 0.1$) and their geographical distance. We found that in 73% of the cases the devices are allocated inside the same data center. Nevertheless, the remaining 27% of the cases are devices that are geographically far (even far located cities). Additional analysis of the correlation in the UNINETT network in terms of distance between failures can be found in [GHHK10].

One of the challenges in this work is to develop models able to reproduce the obtained correlation. In principle, there are standard approaches in order to generate correlated random samples in the time horizon (see for instance [BCNN05]). However, in our case these approaches do not apply, as what is correlated is not the time

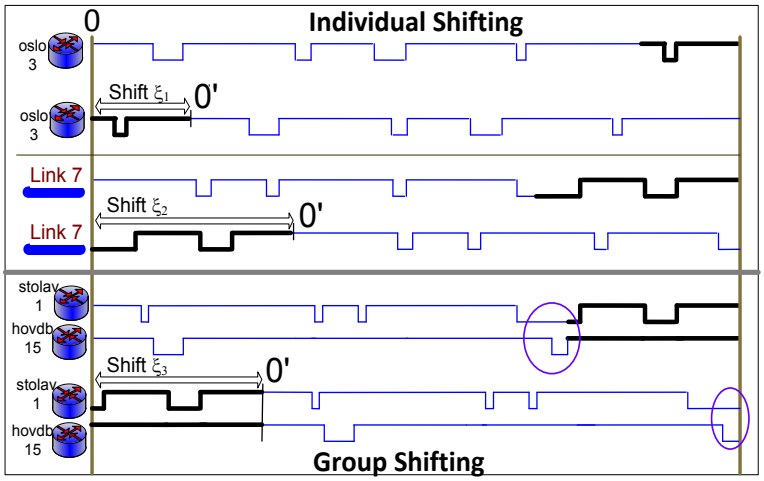


Figure 6. Circular shifting generation.

between events in the time horizon, but near simultaneous occurrences of events in parallel continuously running processes (time vertical simultaneous events). Hence, we propose two new alternative approaches that will be presented in Sections 6 and 7.

6. Study with Trace Driven Simulation

A simulation driven by historical input data is called a Trace Driven Simulation [Law07]. This technique captures directly the system properties but it requires the existence of large historical logs.

In order to generate enough number of observations that allow a proper study of the success probability in network connections using the measured ON/OFF processes, we use a methodology that combines trace driven simulation with circular shifting.

The idea is simulate R replications $(1,..r,..R)$, injecting failure/repair events on the platform described in Section 4, using circular shifting, i.e., in replication r we use the shifted trace $O_i^r(t) = O_i(t+\xi_{r,i})$, where $\xi_{r,i}$ is a random time sampled from an uniform distribution between 0 and τ.

With the use of replications, we may obtain enough number of scenarios to evaluate the success probability, keeping the properties of the $O_i(t)$ processes. According to the results obtained in Section 5, if two processes $(O_i(t), O_x(t))$ have a correlation coefficient bigger than 0.1, the circular shifting is implemented in blocks i.e., $\xi_{r,i} = \xi_{r,x}$. This mechanism preserves the estimated dependence as is illustrated in Fig. 6.

We evaluate the success probability of all the connections described in Section 4 using our TDS approach. For this, we setup α values shorter than A_n^S using equation (8) in order to observe $S(\tau,\alpha)$ when it converges to one. In the simulations, we obtain confidence intervals shorter than 0.005 after 50000 replications.

In Fig. 7(a), we provide an example that describes the behavior observed when the success probability is evaluated in the connection between Trondheim and Oslo where after applying the method described in Section 5, we found that the $O_i(t)$ processes in links 24 and 25 present dependencies. In this connection, the link 24 belongs to the working path and the link 25 to the backup path. Therefore, one can say that connection Oslo-Trondheim presents correlation in parallel. For the sake of comparison, Fig. 7(a) also includes the success probability when independence is assumed. We observe a significant difference between the two curves, showing the huge impact of correlation in parallel.

Fig. 7(b) shows the success probability of connection Tromso-Stavanger given that we found that the $O_i(t)$ processes in links 17 and 22 present dependencies. In this case study, both links belong to the working path, and for this reason, we may say that connection Tromso-Stavanger presents correlation in series. Like in Fig. 7(a), we also show the success probability in the case of independence obtaining a slightly better success probability. This result is interesting given that there is a generalized impression that correlated failures affect negatively the availability. However, in this specific case (correlation in series) the results show a positive effect in the availability. The reason of this result is that when independence is assumed, the failure on each of the links is considered individually, and hence the number of failures can be overestimated, i.e., when a simultaneous failure occurs, two or more failures are in fact summarized just in one. Finally, it is important to notice that the improvement in the success probability shown in Fig. 7(b) is not that considerable compared with the decrease generated by correlation in parallel.

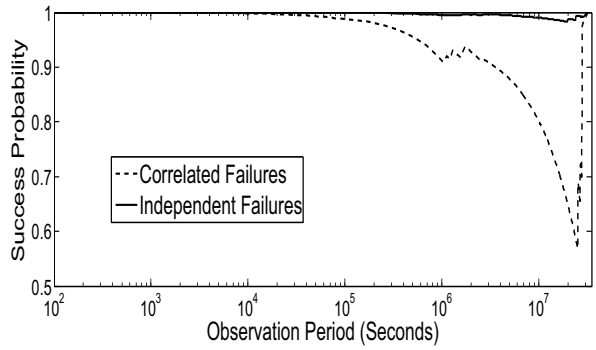
The results obtained through this approach are sufficient to identify potential mistakes in the selection of α if dependencies are not considered (e.g. connection Oslo-Trondheim). A general observation through all the simulations performed using this approach is that the success probability presents irregular tendencies for intermediate observation periods (e.g. times between 10^6 and 10^7 seconds in Fig. 7(b)). The reason is that this methodology recreates the stochastic interaction between network components, but just to some limited extent, the stochastic properties of UP/DOWN times, i.e., the up and down times in $O_i(t)$ are not continuous but selected from a finite set of variables that remain the same through all the replications.

7. Study with Monte Carlo Methods

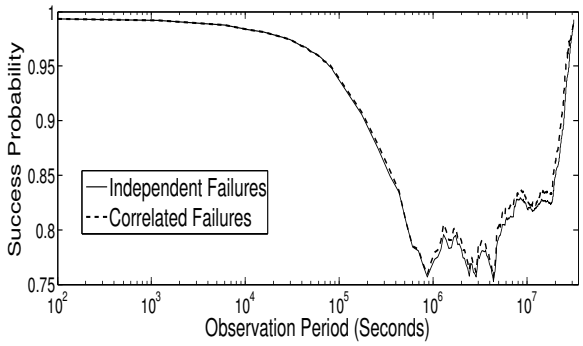
As an alternative approach, we apply a Monte Carlo simulation that allows a larger stochastic variations in the alternating up and down process. In this section, we present a model that integrates the fitted distributions in up and down times with the estimated correlation.

7.1 Uptime and Downtime Distributions

We have estimated UP/DOWN time distributions of network components using maximum likelihood estimation. We have found parameters that fit a hypothesized the-



(a) Correlation in parallel.



(b) Correlation in series.

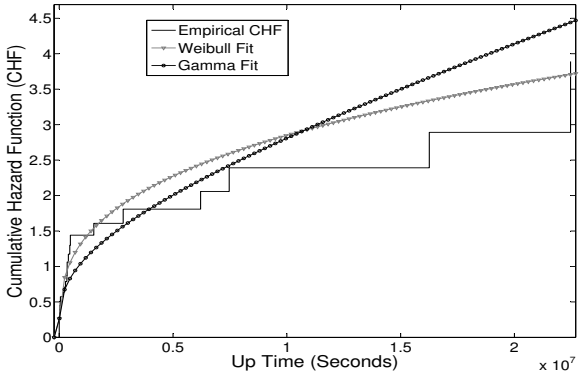
Figure 7. TDS Success probability.

oretical distribution and test them through the use of goodness-of-fit tests (Kolmogorov-Smirnov and Camer-von Mises).

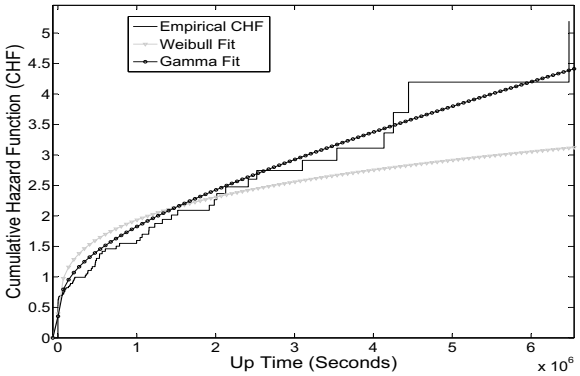
For the case of routers and short distance links, the finding of previous works where the Weibull distribution appears as a good option in order to model failure processes is confirmed. Nevertheless, this assumption seems to be not valid for long distance links (fibers between different cities), where the gamma distribution offers a better fit.

A clear way to distinguish the Weibull and the gamma distribution is through their hazard functions. Therefore, Fig. 8(a) describes a typical fit of the Cumulative Hazard Function (CHF) for uptimes of routers and short distance links. On the other hand, Fig. 8(b) shows the respective fit for the case of long distance links.

For the case of down time distributions, we observe different stochastic behaviors depending on the kind of device and its geographical location. Common to all of them is that they can not be fitted to any simple type of parameterized distribution. Fig. 9 illustrates this issue for down times on link 21. To simulate repair times (Fig. 4(b)), we use the inverse-transformation technique [BCNN05] to generate random samples from the empirical distribution.



(a) Cumulative Hazard Function (CHF) fit for uptimes in routers and short links.



(b) Cumulative Hazard Function (CHF) fit for uptimes in links that connect different cities.

Figure 8. CHF in the failure processes.

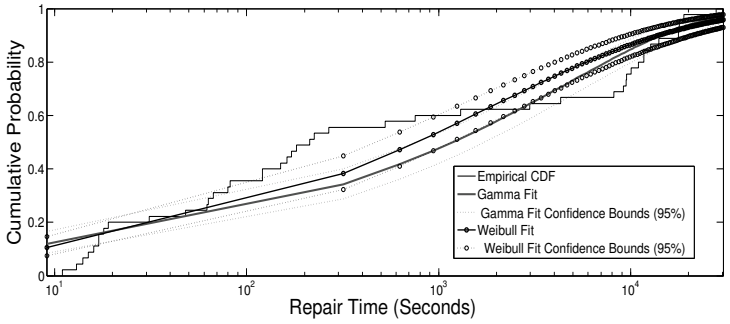


Figure 9. CDF in the repair processes.

Full details about the fitting procedures and specific distributions parameters per each device can be found in [GH11a].

7.2 Correlation Model

The challenge of modeling correlated failures is to generate simultaneous events without affecting the original characteristics of the measured marginal distributions (U_i), obtained through fitting techniques as explained in Section 7.1.

We will use the survival copula proposed by Marshall and Olkin. They assume that *two components (i, x) may fail due to three different events that affect each (λ_i, λ_x) or both (λ_{ix}) of the components* [MO67]. They develop a multivariate distribution that provides numerical solutions under exponential distributions. We adapt this concept from the simulation point of view where the processes may present more general distributions.

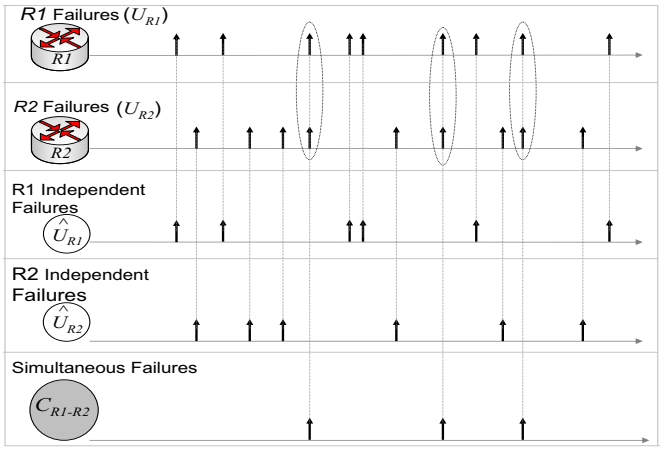
Our hypothesis is that correlation is mainly caused by common cause failures, e.g. electrical power or maintenance actions that affect multiple units. Therefore, a network component may be affected by two different kind of failures: Independent and non-independent. For this, we will regard two point processes $\hat{U}_i(t)$ and $C_{i,x}(t)$. $\hat{U}_i(t)$ represents the points in time when only device i is affected by a failure, and $C_{i,x}(t)$ represents the points in time when devices i and x are affected simultaneously. By using superposition of point processes, we may define the failure process on device i (U_i^*) as follows:

$$U_i^*(t) = \hat{U}_i(t) + \sum_{\forall x} C_{i,x}(t) \tag{11}$$

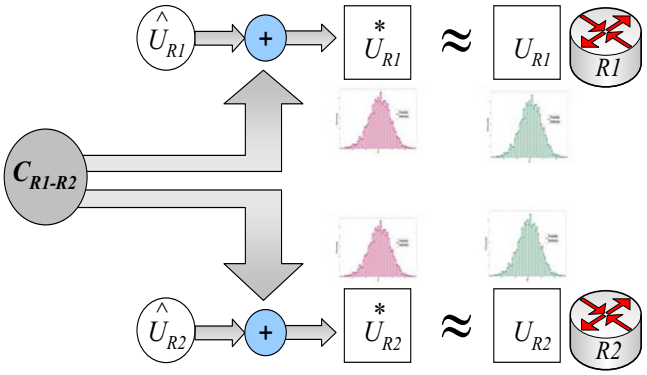
For the implementation of our proposed model, the following steps have to be followed.

- The devices that present correlation are identified using the method presented in Section 5.
- The failure events of a device i that presents correlation are divided in two: All simultaneous events are grouped in the $C_{i,x}$ group, and the non-simultaneous events in the $\hat{U}_i$ group. Fig. 10(a) explains how to implement this procedure.
- The distributions of the new processes $\hat{U}_i$ and $C_{i,x}$ are obtained through fitting techniques as explained in Section 7.1.
- In the simulation, when the failures on device i are injected (see Figure 4(b)), if correlation exists, device i may be affected either by a sample from $\hat{U}_i$ or by a sample from $C_{i,x}$ as Fig. 10(b) illustrates.

Our model should generate samples U_i^* with the same distribution of the measured marginal distribution U_i. In order to verify the accuracy of the proposed model, we use visualization tools and formal statistical test. In the visual approach, we plot the CDF of the original uptimes U_i obtained directly form the UNINETT measurements, and the CDF of the U_i^* uptimes generated with the correlation model. The idea is to



(a) Separation of failure processes.



(b) Generation of correlated processes.

Figure 10. Model to generate correlated failures.

verify that the shapes of the two CDFs do not differ. For instance, in Fig. 11 we show the original uptimes distribution U_{l25} from the measurement performed on link 25 and compare it with the generated uptimes U^*_{l25} obtained from the correlation model.

The results obtained with the CDF visualization may be very illustrative. Nevertheless, we use a Two-Sample Kolmogorov-Smirnov test to confirm that U_i and U^*_i come from the same continuous distribution on all case studies[1].

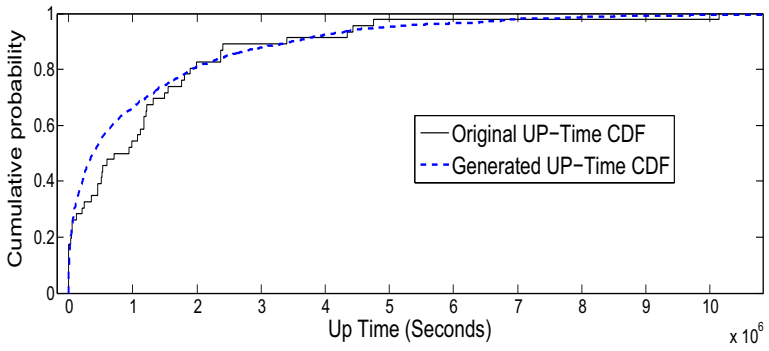


Figure 11. CDF test for the correlation model.

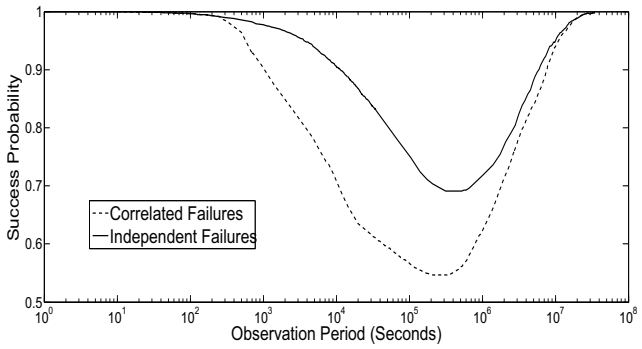


Figure 12. Obtained success probability using Monte Carlo techniques (correlation in parallel).

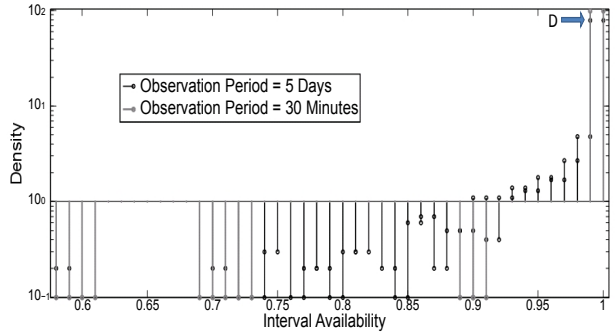
7.3 Success Probability Evaluation

Using the model previously described, we will evaluate the success probability on the connections described in Section 4. We will assume SLA contracts durations (observation period) of one year, and using equation (8) we will setup SLA availability guarantees (α) shorter than the lower bound of the steady state availability (A_n^S) in order to assure that our success probability converges to one after drooping below for a certain period.

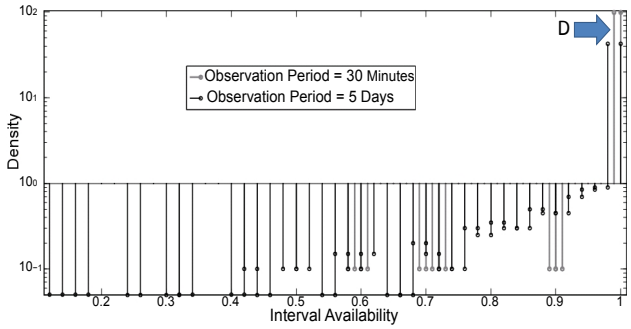
Fig. 12 shows the success probability of connection Oslo-Trondheim (same case study shown in Fig. 7(a)) where we know the existence of correlation in parallel. The success probability when links 24 and 25 are correlated (real case) presents a considerable difference with the behavior observed when independence is assumed. We observe that the minimum success probability obtained with independent assumptions

[1]The Kolmogorov-Smirnov test was performed with levels of significance of 5%.

is approximately 0.75 while for the real case (correlation between link 24 and 25) $S(\tau, \alpha)$ decrease approximately up to 0.55.



(a) $\hat{A}(\tau)$ Distribution with independence assumptions.



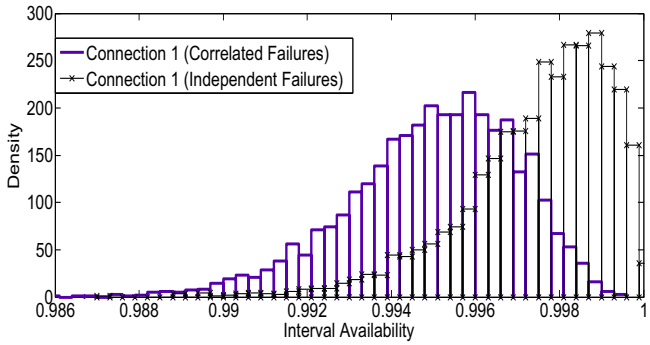
(b) $\hat{A}(\tau)$ Distribution with failure dependencies.

Figure 13. $\hat{A}(\tau)$ with correlation in parallel (Oslo-Trondheim). Short observation period.

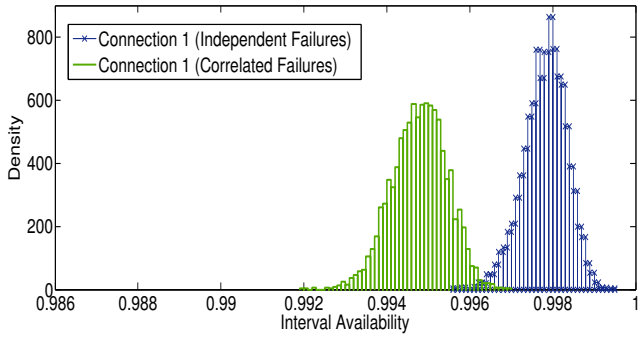
Network designers always have to follow relevant policies in order to provide reliable connections. The effort needed in this provision phase were described in Section 4. Some of the most relevant are: To avoid the existence of single point of failure, and the provision of backup paths according to the the *SRLG* rule wich states that the connections affected by one failure in their working paths can not share any backup resource.

Our results show that these policies may become useless if the effects of correlated processes such as $C_{i,x}$ are not considered. Therefore, the design of reliable networks considering dependencies becomes a third fundamental policy, and the methods described in this paper a useful tool to detect, verify and assess the consequences of correlated failures.

In order to have a better understanding of the results shown in Fig. 12 we studied the PDF of $\hat{A}(\tau)$ for correlated and independent failures in the same case study. We start observing $\hat{A}(\tau)$ for short observation periods i.e., $\tau = 30$ *min* and $\tau = 5$ *days*. Fig. 13(a)



(a) $\hat{A}(\tau)$ Distribution (τ = 1 Year).



(b) $\hat{A}(\tau)$ Distribution (τ = 10 Year).

Figure 14. $\hat{A}(\tau)$ with correlation in parallel (Oslo-Trondheim). Long observation period.

shows the case when independence is assumed. We observe high density when the interval availability is approximately equal to one, representing the case of perfect connectivity (no disconnection) during the observation period. Nevertheless there is a slight decrease "*D*" in the perfect connectivity's density when the observation period increases from 30 minutes to 5 days. On the other hand, Fig. 13(b) also shows the PDF of $\hat{A}(\tau)$ under the real scenario (correlated failures). The general behavior is similar to the one shown in Fig. 13(a). However, in this case the decrease "*D*" in the density of the perfect connectivity when the observation period increases from 30 minutes to 5 days is much more considerable. This shows that the probability of having perfect connectivity is quickly affected with the observation period when correlation in parallel exists, and explains the difference observed in Fig. 12.

We complement this study by analyzing the evolution of the interval availability's PDF, using two additional values of τ. First, we select $\tau = 1$ *year* in order to observe the distribution behind the success observed at the right hand side of Fig. 12. Then, we choose an observation period of 10 years in order to observe the behavior of the interval availability when it is closer to the steady state.

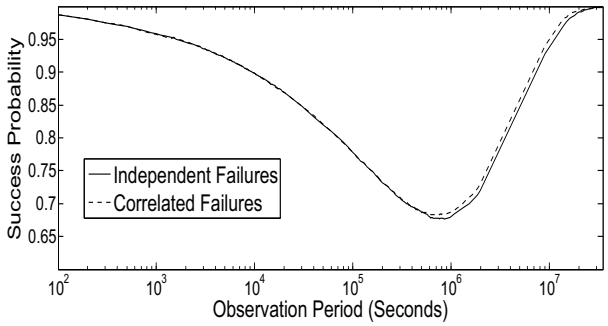
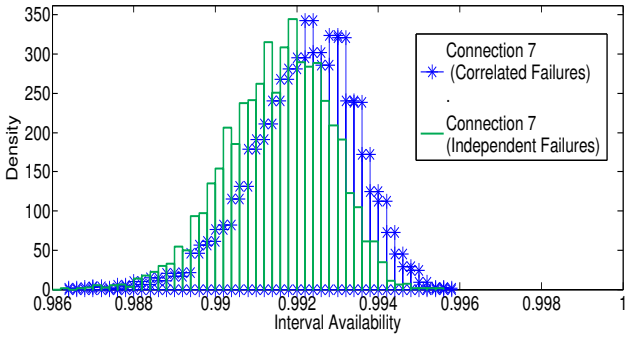
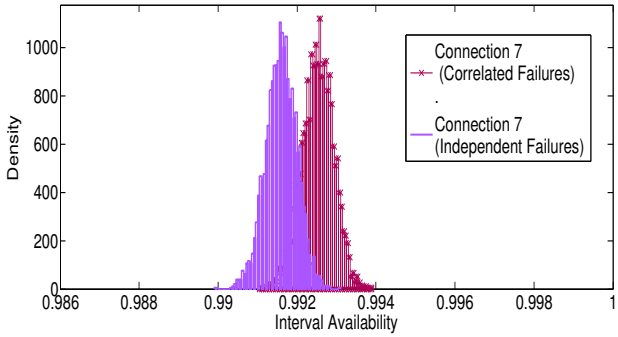


Figure 15. Obtained success probability using Monte Carlo techniques (correlation in series).



(a) $\hat{A}(\tau)$ Distribution (τ = 1 Year).



(b) $\hat{A}(\tau)$ Distribution (τ = 10 Year).

Figure 16. Interval availability with correlation in series.

The effect of correlation in parallel can be observed in Fig. 14. In this case for both observations periods (τ = 1 *year* and τ= 10 *years*) the PDFs appear shifted to the left, indicating lower levels of availability. In addition, the variance is shorter when

τ is equal to ten years and independence is assumed, indicating that the existence of dependencies delays the time to converge to the steady state.

Another case study is shown in Fig. 15. The success probability behavior of the connection between Tromso and Stavanger is evaluated (same case study shown in Fig. 7(b)), where we know the existence of correlation in series.

For short observation periods in Fig. 15, both curves present similar behavior. We observe a short improvement in the success probability for observation periods larger than two months when correlation is taken into account. Nevertheless, the gap between the two curves is small, and hence it may not represent any significant advantage for the network operator.

The analysis of the probability density function of the interval availability at specific observation periods offers a better understanding of the success probability (as was shown in Fig.13 for connection Oslo-Trondheim).

In Fig. 16, we observe that the PDFs in the correlation in series case appear shifted to the right when correlation is considered. A shift to the right represents a better level of availability and a smaller SLA risk, given that the integral over this curve from 0 to α is drastically reduced. As was explained at the end of Section 6, this result differs from the generalized impression that correlated failures affect negatively the availability. The reason is that the assessment under independent assumptions overestimates the number of failures in the working path. When correlation in series exists, two or more simultaneous failures may be summarized just in one. However, the magnitude of the shift in this case is shorter than the one observed in Fig. 14. Additionally, we obtain similar values of variance with correlated or independent failures, showing that when correlation in series exists the time to converge to the steady state is not affected considerably.

The presented PDFs provide an improved insight on the problem of guaranteeing availability in SLAs and offer a wider overview of the success probability's behavior. For instance, in any of those figures, α can be set at any point in the x axis, being the risk the probability mass for values shorter than the set value.

Finally, normality tests with significance level of 5% on the PDFs of $\hat{A}(\tau)$ when τ= 10 *years* show that their behavior approaches a normal distribution with expected value A. This empirical result for compound systems under non-Markovian and non-independent assumptions agrees with the original result obtained by Takacs in [Tak57] for a single-component system. Unfortunately, SLAs are usually defined for shorter periods than ten years, and hence, the distributions to deal with are much more complex.

8. Concluding Remarks

This paper yields an improved insight into the failure characteristics at a real operational core network and demonstrates the importance of taking into account correlation between failure processes for defining SLA's availability parameters.

In this study, we present two approaches that can be used to assess the SLA risk on network connections. First, using trace driven simulation combined with circular

shifting. Second, using a discrete event simulation that generates correlated samples according to marginal distributions.

The trace driven simulation approach showed to be effective in the detection of potential threats to the SLA fulfillment caused by correlation. The implementation is made directly from operational data, but irregular tendencies for intermediate observation periods were observed. On the other hand, with our Monte Carlo approach, it was possible to obtain the detailed behavior of the SLA success probability. The implementation of this approach in a general scenario requires the verification phase where statistical tests are performed in order to verify that the generated uptime samples belong to the same continuous marginal distribution originally measured.

Salient effects of the influence of dependencies on SLA success probability were observed when correlation is present in parallel and in series. For the case of correlation in series we observe a small reduction in the risk. On the other hand, the effect of correlation in parallel produces a negative and stronger impact on the SLA risk.

We suggest to take into account the effects of dependencies in the provision of reliable networks connections, given that they may become as important as well known policies such as avoid single point of failure, and assign backup paths considering the Shared Risk Link Group rule (SRLG). The methods described in this paper can be used to detect dangerous correlations and asses their respective consequences.

The study of the interval availability's PDF complements the observations obtained from the SLA success probability. It shows the variations that correlation produces in the whole distribution, generating shifts to the right (better availability level) for the case of correlation in series and shifts to the left (worse availability level) and variance increase (slower steady state convergence) for the case of correlation in parallel. Finally, the evolution of the PDF of the interval availability starts with a dominant density in values equal to one (perfect connectivity) for short observation periods, then that density starts to be distributed across other values ($0 \leq \hat{A}(\tau) < 1$) for intermediate observation periods, and finally the interval availability starts to approach a normal distribution with expected value A for long observation periods.

References

[BCNN05] Jerry Banks, John S. Carson, Barry L. Nelson, and David M. Nicol. *Discrete-Event System Simulation (4rd Edition)*. Prentice Hall, 4 edition, 2005.

[Bir03] Graham M. Birtwistle. *DEMOS - a system for Discrete Event Modelling on Simula*. University of Leeds, 2003.

[GH09] Andres J. Gonzalez and Bjarne E. Helvik. Guaranteeing service availability in SLAs; a study of the risk associated with contract period and failure process. *Proceeding of the IEEE Latin-American Conference on Communications (LATINCOM)*, pages 1 –6, Sep. 2009.

[GH10] Andres J. Gonzalez and Bjarne E. Helvik. Dynamic Sharing Mechanism for Guaranteed Availability in MPLS Based Networks. *Proceeding of the IEEE International Communications Quality and Reliability (CQR)*, Jun. 2010.

[GH11] Andres J. Gonzalez and Bjarne E. Helvik. Analysis of failures characteristics in the UNINETT IP backbone network. *Proceeding of the IEEE 7th International Symposium on Frontiers in Networking with Applications (FINA)*, Mar. 2011.

[GHHK10] Andres J. Gonzalez, Bjarne E. Helvik, Jon K. Hellan, and Pirkko Kuusela. Analysis of dependencies between failures in the UNINETT IP backbone network. *Proceeding of the IEEE International Symposium on Pacific Rim Dependable Computing (PRDC)*, 2010.

[GT88] Ambuj Goyal and Asser Tantawi. A measure of guaranteed availability and its numerical evaluation. *IEEE Transactions on Computers*, Volume 37, Issue 1:25 – 32, 1988.

[InCM+02] Gianluca Iannaccone, Chen nee Chuah, Richard Mortier, Supratik Bhattacharyya, and Christophe Diot. Analysis of link failures in an IP backbone. *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurment (IMW)*, pages 237–242, 2002.

[Law07] Averill M. Law. *Simulation Modeling and Analysis*. McGraw-Hill Education, New York, fourth edition, Apr. 2007.

[MIB+08] Athina Markopoulou, Gianluca Iannaccone, Supratik Bhattacharyya, Chen-Nee Chuah, Yashar Ganjali, and Christophe Diot. Characterization of Failures in an Operational IP Backbone Network. *IEEE/ACM Transactions on Networking*, 16(4):749–762, Aug. 2008.

[MO67] Albert W. Marshall and Ingram Olkin. A multivariate exponential distribution. *Journal of the American Statistical Association*, 62(317):pp. 30–44, 1967.

[RS95] Gerard0 Rubino and Bruno Sericola. Interval availability analysis using denumerable markov processes: application to multiprocessor subject to breakdowns and repair. *IEEE Transactions on Computers*, Volume 44, Issue 2:286 – 291, Feb. 1995.

[Tak57] Lajos Takacs. On certain sojourn time problems in the theory of stochastic processes. *Acta Mathematica Hungarica*, 8:169–191, 1957.

[UNI12] UNINETT. The Norwegian Research Network. Network Topology. [online]. Available at: http://drift.uninett.no/stat-q/load-map/uninett,,traffic,peak. 2012.

[XTMM11] Ming Xia, Massimo Tornatore, Charles U. Martel, and Biswanath Mukherjee. Risk-Aware Provisioning for Optical WDM Mesh Networks. *IEEE/ACM Transactions on Networking*, 19(3):921 –931, Jun. 2011.

PAPER F

Dynamic Sharing Mechanism for Guaranteed Availability in MPLS Based Networks

Andres J. Gonzalez and Bjarne E. Helvik

Proceedings of the IEEE 2010 International Communications Quality and Reliability Workshop CQR

Vancouver, Canada, June, 2010

DYNAMIC SHARING MECHANISM FOR GUARANTEED AVAILABILITY IN MPLS BASED NETWORKS

Andres J. Gonzalez, Bjarne E. Helvik
Centre for Quantifiable Quality of Service in Communication Systems
Norwegian University of Science and Technology,
Trondheim, Norway
{andresgm,bjarne}@q2s.ntnu.no

Abstract This paper proposes an algorithm to allocate connections with bandwidth and availability requirements in a telecommunication network, where the connections may share resources in their backup paths. The allocation is dynamic in the sense that incoming connections have random arrival times, source and destinations, and they stay in the network for a finite period. The algorithm has been designed for core backbone networks with continuous bandwidth distribution, like for example MPLS networks. An efficient bandwidth utilization was obtained by an intelligent sharing mechanism that takes into account the properties of networks with continuous bandwidth allocation. The problem may be formulated as an NIP (Nonlinear Integer Programming) problem. However, due to the well known complexity and scalability limitations in solving this kind of problems, the solution is based on heuristic procedures. A performance comparison with previously published algorithms is carried out for some "reference networks", demonstrating a substantially better resource utilization.

1. Introduction

Offer quality of service (QoS) to the connections that use a network is a matter of major impact. Availability is a significant element on providing a good QoS. It may be clearly defined in a Service Level Agreement as a parameter to be guaranteed by the network operator. Violation of the agreed value may have large consequences. For this reason is very relevant to develop techniques that help to fulfill the specific availability. At the same time another important concern is the efficient utilization of the network resources.

This paper proposes a mechanism that allocates connections in a network that may be affected by multiple failures, fulfilling specific availability requirements. Compared with previous related works, this mechanism reduces considerably the bandwidth reserved for backups. This is made through the use of a novel technique called *dynamic sharing*.

Protection is one of the most common strategies used to meet dependability requirements with a high probability. In this approach, the network resources that provide connection between two points are planned when the connection arrives, and before any failure affect it. For this, the use of predefined backups is a common policy. Depending on the characteristic of the network elements and on the implemented backup scheme, different dependability degrees may be obtained.

The backup scheme may be dedicated where the bandwidth of the working and the backup path is reserved exclusively for one connection. Works such as [MH08a] and [FT06] have proposed relevant ideas in these cases. On the other hand, a scheme known as Shared Backup Path Protection (SBPP) may be implemented, where the network resources may be used more efficiently. Under this scheme, a connection may share bandwidth in its backup path with other connections. For this reason the offered availability will depend not only on the network resources assigned to one connection, but also on the behavior of the others. There are some approaches that try to model this problem assuming that maximum two failures may occur in the network [MSW05]. In [HTH08], the use of SBPP is studied including partial restoration. A novel concept that reduces the connection blocking probability in networks that use SBPP is proposed in [MH08b], where the idea is to implement connection priorities that are proportional to the time that a connection has been allocated in the network, in this way, the availability of an existing connection will be unaffected by the establishment of new connections.

Finally, when SBPP is used, the concept of Shared Risk Link Group (SRLG) has to be considered in order to have robustness under single link failures. That means that the connections affected by one failure in the working paths can not share any backup resource [RBS$^+$01].

The problem of allocating connections in a network fulfilling bandwidth and availability requirements has been studied through the use of Integer Linear Programming (ILP). Formal optimization techniques have been successfully applied for the case of connections without protection and connection with dedicated path protection [ZZZ$^+$07]. For the case of shared path protection, the problem can not be efficiently solved due to the non linear constraints in the shared availability inequalities [LLY09]. Therefore, in further SBPP works the use of heuristics algorithms is common. An interesting proposal is found in [KL00], where by solving an ILP, a working and a shared backup path are obtained. However, this solution uses Shared Risk Link Group (SRLG) as the only availability constraint, and hence, it can not be used to fulfill specific availability values. That means that through the use of ILP may be obtained efficiently a working path that is protected by a disjoint backup path, but without achieving specific availability targets.

In our work, the obtained constraints becomes non linear, since specific availability values have to be fulfilled. For this reason, the proposed solution uses heuristic techniques.

In MPLS networks, the link capacities are split for allocate multiple connections, and the bandwidth may be assigned in any proportion as long as the capacity limit is not exceeded. This is known as continuous bandwidth assignment. A recent work

[HLS07] deals with the problem of SBPP considering the properties of continuous bandwidth. This approach can be used to allocate connections in a network with guaranteed availability, and additionally it scales very well. However, the sharing scheme used has some restrictions and the bandwidth utilization is therefore not completely efficient. Our proposal optimizes the amount of bandwidth reserved for backup paths through the use of dynamic sharing. This idea basically gives more flexibility on the selection of the connections that will share resources.

This paper is organized as follows. In Section 2, the problem is defined. Section 3 explains the core idea of *dynamic sharing*. In Section 4, the optimization problem that minimizes the backup bandwidth utilization is formulated. Section 5 shows the heuristic procedure that uses dynamic sharing to allocate connections, fulfilling bandwidth and availability requirements. Section 6 shows some case studies that illustrate the performance of the proposed mechanism. Finally, Section 7 concludes the paper.

2. Problem Definition

The resilience of a network may be defined as the ability of a network to automatically react to failures through the use of alternative failure-free paths. Planning redundancies and make use of them to deal with failure situations is one of the keys to provide dependable services.

The mechanism developed in this paper is oriented to core backbone networks defined under the standard notation $G(V,E)$, where V represents a set of routers, and E a set of links that interconnect those routers. The links l are characterized by a capacity B_l and a steady state availability ρ_l.

Other important elements for our mechanism are the connections that arrive and depart randomly, and that have to be allocated and removed from the network according to the stipulated contract period. An arriving request C_n is characterized by a quadruple (s, d, b_n, a_n), where s and d are the source and destination routers respectively, b_n is the capacity requirement, and a_n is the availability requirement . Through the use of routing algorithms, a path in $G(V,E)$ to allocate C_n may be found. In MPLS networks, this path is defined as an LSP (Label Switched Path). Once the path is known, the RSVP signaling protocol will perform the resource reservation in a continuous way. That means that the reserved resources fit exactly with the requested amount (this condition differentiate this work from the approaches where the link capacities are distributed in discrete amounts e.g., complete wavelengths). Section 3 explains how to take advantage of the continuous bandwidth allocation in order to optimize resource utilization.

In our work the protection mechanism will be implemented through the use backup paths that are used when the main connections fail. The working capacities are dedicated and the backup resources may be shared by several connections as long as the availability requirements are not violated.

Given that the bandwidth used on a link for working connections (B_l^W) is dedicated, the design of sharing strategies that optimize the bandwidth utilization is not possible. On the other hand, the bandwidth reserved for backup paths on an individual link (B_l^K)

is shared. This property allows the implementation of techniques able to optimize the total backup capacity in the network through the selection of an appropriate sharing scheme.

The proposed mechanism tries to fit as best as possible the bandwidth reserved for a connection in the shared areas.

Finally, all the allocated connections have to fulfill the bandwidth constrain defined by:

$$B_l \geq B_l^W + B_l^K \; \forall l \in G(V,E) \tag{1}$$

2.1 Availability Calculations

When a connection request C_n arrives with requirements b_n and a_n to the network, first a working path W_n is found. This path is composed by several links l with respective values ρ_l. The asymptotic availability of W_n is calculated using the following equation:

$$A_n^W = \prod_{\forall l \in W_n} \rho_l \tag{2}$$

Additionally, the unavailability of the working path of C_n may be defined as $U_n^W = 1 - A_n^W$. If A_n^W is smaller than a_n, a backup path K_n has to be found. The resources may be used more efficiently if K_n is shared with a group of other connections (G_n^s).

When any link that belongs to the working path of C_n is affected by a failure, the found backup path will be used in order to keep the connection operational. Nevertheless, the backup resources are available for connection C_n if the path itself is operational, and additionally if: The backup resources are not being used by another G_n^s connection, or if the connection that is using the backup has lower priority than C_n, and hence it can be preempted.

In the literature is well known that if the priority of C_n is assumed equal to zero, the total availability of a connection with sharing backup (A_n^S) may be defined by a lower bound as follows:

$$A_n^S \geq 1 - \left(1 - \prod_{\forall l \in W_n} \rho_l\right) \cdot \left(1 - \prod_{\forall l \in K_n} \rho_l \cdot \prod_{\forall C_x \in G_n^s} A_x^W\right) \tag{3}$$

If $A_n^S \geq a_n$, the C_n requirements are fulfilled.

During the allocation of C_n, the offered availability to any connection C_x previously established must be kept over a_x. Equation (3) shows that the availability of an already established connection x is affected by the availability of the working path of the new arriving connection (A_n^W). Therefore, an additional admission policy that restricts the sharing of connection n with x depending on A_n^W is needed. The minimum possible value of the availability of the working path of the incoming connection n can be expressed as follows:

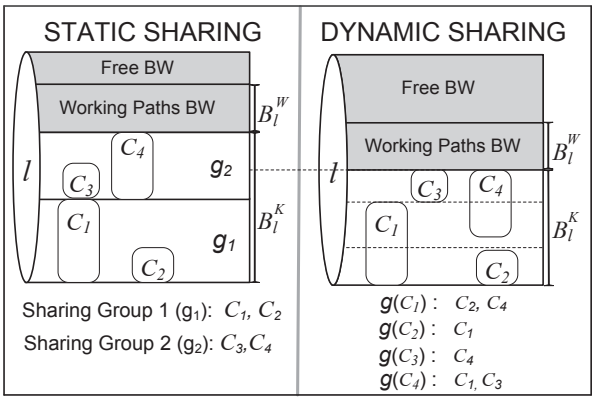


Figure 1. Static and Dynamic Sharing Schemes.

$$A_n^W \geq \frac{1 - \frac{1 - a_x}{[U_x^W]}}{(\prod_{\forall l \in K_x} \rho_l) \cdot (\prod_{\forall C_y \in G_x^s} A_y^W)} \tag{4}$$

Equations (3) and (4) will be used in this paper to verify the fulfillment of the availability requirements.

3. Dynamic Sharing

In MPLS networks, the link capacities are split in many parts, and the bandwidth may be assigned to several connections in any proportion, as long as the capacity limit is not exceeded. This is a concept known as continuous bandwidth assignment.

Under this scheme, a link capacity B_l may be split in three parts. The first part is the capacity that has not been reserved by any connection and therefore is free to be used by any future request. The second part is the bandwidth used by working resources B_l^W. Finally, there is a bandwidth reserved for backup paths B_l^K which may be shared by several connections. Figure 1 shows this concept.

Previous works have defined bandwidth sharing mechanisms, where the connections are grouped by a fixed scheme. If the availability requirements are not fulfilled for all of the connections inside a given group, new resources have to be used i.e., new groups are established. We define these traditional mechanisms as *static sharing*. An example of one of the latest static sharing approaches [HLS07] is shown in Figure 1.

This paper proposes a new sharing strategy defined as *dynamic sharing*, where the possible sharing group of an incoming connection C_n is not restricted by a fixed scheme. For this reason, any possible combination of the connections previously allocated on K_n may belong to the sharing group of C_n. Therefore, the concept of a group containing a fixed number of connections sharing resources does not exist. Instead each connection has its own sharing group. Given this flexibility, the solution

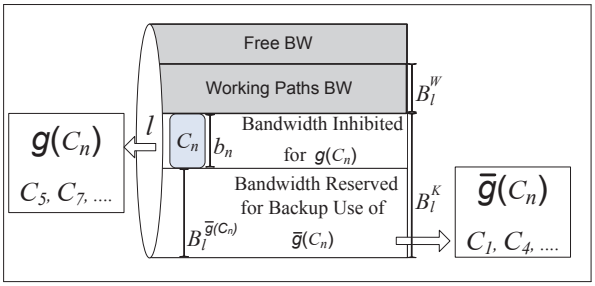


Figure 2. Example that describes the use of backup capacity.

may be manipulated in order to optimize the resource utilization. For example, in the dynamic sharing example shown in Figure 1, C_1 shares the backup with C_2 and C_4, but it does not mean that C_2 has to share with C_4, which would be the case in a static sharing group scheme. To the authors' knowledge, the problem of sharing resources in a complete dynamic way under bandwidth and availability constrains has not been considered previously.

3.1 Sharing Groups

The core idea of this paper is that each connection C_n has its own sharing group $g(C_n)$, independently of the sharing groups of the other connections.

On any link l that belongs to K_n, B_l^K may be used by a group of connections $G_l^{K_n}$.

The non sharing group of C_n ($\overline{g}(C_n)$) will be defined as $G_l^{K_n} - g(C_n)$, and it represents the connections that use any link $l \in K_n$ as a backup path, but that can not share resources with C_n due to the violation of the availability constraint.

The selection of $g(C_n)$ is made in such a way that the availability requirements posed by Equations 3 and 4 are kept for all the implied connections. Therefore, if C_n is using the backup, a policy that inhibits all connections in $g(C_n)$ from utilizing B_l^K is introduced. Figure 2 shows an example.

When C_n uses part of B_l^K, there is a remaining bandwidth available $B_l^{\overline{g}(C_n)}$, which is determined by $B_l^K - b_n$ (Figure 2). Given that C_n is not really sharing any capacity on l with $\overline{g}(C_n)$, a mechanism that assures that $B_l^{\overline{g}(C_n)}$ will be sufficient to accommodate the $\overline{g}(C_n)$ connections has to be implemented.

4. Minimizing the Shared Bandwidth Reserved for Backup Connections

The objective as stated earlier is to minimize the total amount of bandwidth reserved for backup paths. This can be achieved if for each incoming connection C_n, B_l^K is

minimized on all the links l that belongs to the backup path K_n by the appropriate selection of the C_n's sharing group $g(C_n)$. This can be expressed by the following equation.

$$min \sum_{l \in K_n} \theta_l \tag{5}$$

Where θ_l represents the increase in the bandwidth reserved for backup paths on link l.

The possible solutions to this problem are given by the feasible combination of connections in $G_l^{K_n}$ grouped in specific $g(C_n)$ and $\overline{g}(C_n)$. For one specific combination is possible to find the values θ_l. The next equation shows how to make this.

$$\theta_l = \max\{\ 0\ ,\ b_n + B_l^{\overline{g}(C_n)} - B_l^K\ \} \tag{6}$$

In (6), B_l^K and b_n are known values. Obtaining $B_l^{\overline{g}(C_n)}$ is hard as it depends on the sharing and non sharing groups of the connections previously established. Assuming that $G_l^{K_n}$ contains a number of m connections, this sharing information may be captured by a *Sharing Matrix* **S** which is an m x m matrix defined as follows:

$$S_{i,j} = \begin{cases} 1 & \text{if } i{=}j \\ 1 & \text{if } C_i \text{ NOT share with } C_j \\ 0 & \text{if } C_i \text{ share with } C_j \end{cases} \tag{7}$$

In Figure 3 is observed an example of how to obtain θ_l and $B_l^{\overline{g}(C_n)}$ under an specific $g(C_n)$ and $\overline{g}(C_n)$.

In order to obtain $B_l^{\overline{g}(C_n)}$, we evaluate one by one the bandwidth contribution T_j that each connection C_j that belongs to $G_l^{K_n}$ may add to $B_l^{\overline{g}(C_n)}$. Nevertheless, we only need to consider the connections that belong to $\overline{g}(C_n)$ given that they can not be inhibit by C_n. Therefore, the following variable is defined.

$$I_j = \begin{cases} 1 & \text{if } C_j \in \overline{g}(C_n) \\ 0 & \text{Otherwise} \end{cases} \tag{8}$$

On each evaluation all the pervious calculations have to be kept, using a cumulative scheme that may be represented by the matrix **I** as follows:

$$\mathbf{I} = \begin{pmatrix} I_0 & 0 & 0 & \cdots & 0 \\ I_0 & I_1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \\ I_0 & I_1 & I_2 & \cdots & I_m \end{pmatrix} \tag{9}$$

The one by one evaluation may be summarized by an element matrix multiplication ($\bullet$) between **I** and **S**. We define this new matrix as **A** which is given as:

$$\mathbf{A} = \mathbf{I} \bullet \mathbf{S} \tag{10}$$

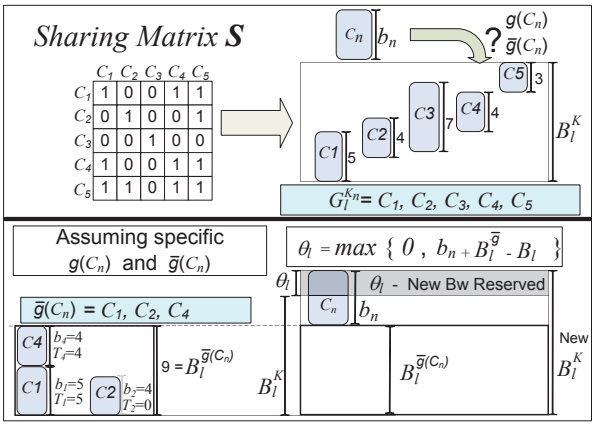


Figure 3. Example that illustrates how to obtain θ_l.

$$\mathbf{A} = \begin{pmatrix} I_0 \cdot S_{1,1} & 0 & \cdots & 0 \\ I_0 \cdot S_{2,1} & I_1 \cdot S_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ I_0 \cdot S_{m,1} & I_1 \cdot S_{m,2} & \cdots & I_m \cdot S_{m,n} \end{pmatrix} \tag{11}$$

Each row in matrix $\mathbf{A}$ contains a set of binary values that indicate which of the connections that belong to $G_l^{K_n}$ will contribute to the size of $B_l^{\overline{g}(C_n)}$ on each evaluation.

Finally, when a contribution T_j is considered, this amount does not need to be taken into account again on future contributions T_h for all C_h that belong to $g(C_j)$. Therefore T_j may be modeled by the following equation:

$$T_j = \max\{\ 0\ ,\ \min(b_j - T_h)\ \}\ \forall\ C_h \in g(C_j) \tag{12}$$

Where T_h is at the beginning equal to zero for all non yet evaluated C_h.

When all the m connections that belong to $G_l^{K_n}$ are considered, an iteration process is generated. This iteration may be captured by the following equation.

$$\mathbf{V_m} = \sum_{j=1}^{m} \mathbf{V_{j-1}} + \mathbf{A_j} \cdot I_j \cdot T_j \tag{13}$$

Where $\mathbf{A_j}$ is a 1 x m vector that represents the row j of the matrix $\mathbf{A}$. $\mathbf{V_m}$ is a vector of size m with initial value $\mathbf{V_0}$ defined as:

$$\mathbf{V_0} = [0,0,0,....0] \tag{14}$$

At the end of the process, the maximum value of $\mathbf{V_m}$ contains $B_l^{\overline{g}(C_n)}$. Therefore, it will be defined as:

$$B_l^{\overline{g}(C_n)} = \max(\mathbf{V_m}) \tag{15}$$

The objective function is now fully defined by (5), (6) and (15). On the other hand, the availability constrains are defined by (3) and bandwidth constrains are defined by (1). It can be observed that this formulation lies into the nonlinear integer programming. Given that the NIP solution of this problem may take huge time, this paper proposes an heuristic solution.

5. Connection Allocation Mechanism Using Dynamic Sharing

In this section, we explain the mechanism implemented to allocate an incoming connection C_n fulfilling its bandwidth need b_n and availability requirement a_n, using dynamic sharing.

First, through the use of conventional routing algorithms like minimum cost path, a working path between the source and the destination is found. The availability of this path A_n^W is calculated using equation (2). When $A_n^W > a_n$, the connection may be provided without protection, i.e., the connection may be unprotected.

On the other hand, when $A_n^W < a_n$, a backup path is needed. There are two basic rules known in the literature for the backup path selection. The first is that the working and the backup path have to be link disjoint. The second rule is known as Shared Risk Link Group (SRLG), and it basically states that given any link failure, the affected working connections can not share any backup resource [KL00].

Once those two rules are fulfilled, the selected backup path contains a number of links with connections previously allocated. In principle, C_n may potentially share resources with all of them. This initial group of possible sharing connections will be called *potential sharing group* (PSG), and it may be easily obtained using a link by link search through the backup path.

The next step is to guarantee that the availability offered to the PSG connections remains bigger than the respective availability requirement (Checking 1 Fig 4). Through the use of Equation (4), we can easily verify if the availability of the working path of C_n is sufficient in order to share with each of the different connections that belong to PSG. The above procedure reduces the size of PSG obtaining a *Reduced Potential Shared Group* (RPSG). It is also important to highlight that the connections that do not allow sharing with C_n will belong to the *Non Sharing Group* ($\overline{g}(C_n)$). In [MH08b] has been shown that the resources can be more efficiently used and recalculations are avoided if the availability of an existing connection is unaffected by the establishment of new connections, using a priorities scheme. Our mechanism is adaptable to this approach by including or removing the *Obtain RPSG* block (Figure 5).

The next step is to find which connections from RPSG should be included in $g(C_n)$ and which in $\overline{g}(C_n)$ as is illustrated in the *Checking 2* (Figure 4). To perform this step, all the connections from RPSG are evaluated one by one using (3). After being added into $g(C_n)$, the individual connections that generate a violation in the availability requirement a_n will be deleted from RPSG and added to $\overline{g}(C_n)$. This is represented in Figure 5 as *Update* $\overline{g}(C_n)$ step.

The next step is to find the RPSG connection combination that minimizes the total bandwidth reserved for backup capacities. This combination is obtained, evaluating

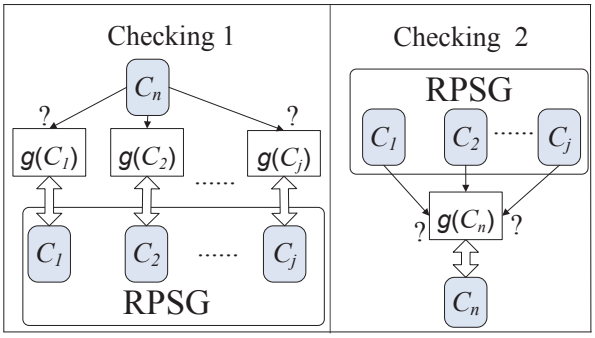


Figure 4. Checking Procedures to find the sharing group of an incoming connection C_n.

one by one the connections that currently belong to RPSG. On each iteration, the connection under evaluation is temporarily added to $\overline{g}(C_n)$. Then C_n is temporarily allocated in the network under that conditions, and using Equation (5) an increase in the total bandwidth used for backup resources may be obtained. After performing those calculations, the temporal changes are reversed and the next connection in RPSG is evaluated. When all the connections are analyzed, noticing which of them generates the minimum increase in backup bandwidth is easy. This connection will be deleted from RPSG and added to $g(C_n)$ (New $g(C_n)$ Figure 5).

After this step, there is a new scenario where RPSG has one element less and $g(C_n)$ one more. That means a change in A_n^S, and hence the repetition of the *Update* $\overline{g}(C_n)$ step is needed. On the other hand, if RPSG is not empty, there are some connections that still need to be evaluated. Therefore, the explained procedure will be performed until RPSG becomes empty. When this happens, the algorithm stops and C_n can be allocated through the backup path with the found sharing schema. The complexity of the algorithm described in this section is $O(E.\delta^3)$, and it can be deduced from the loop shown in Figure 5 and equation (13) that is implied inside this process. δ represent the connections that are in RPSG, which in the worst case, it may be equal to the number of allocated connections in G when C_n arrives.

6. Case Studies

In order to verify the performance of our mechanism, and compare the behavior of dynamic and static sharing schemes, some simulations were performed. The experiment setup was implemented in C++, using the topology COST266 network which is an European backbone with 37 nodes and 112 links (Figure 6) [SND]. It was assumed that each link has $\rho_l = 0.99$ and a capacity of 100Gb/s, which is big enough compared with the connection bandwidth request assumed (1-150Mb/s), in order to avoid bandwidth bottlenecks.

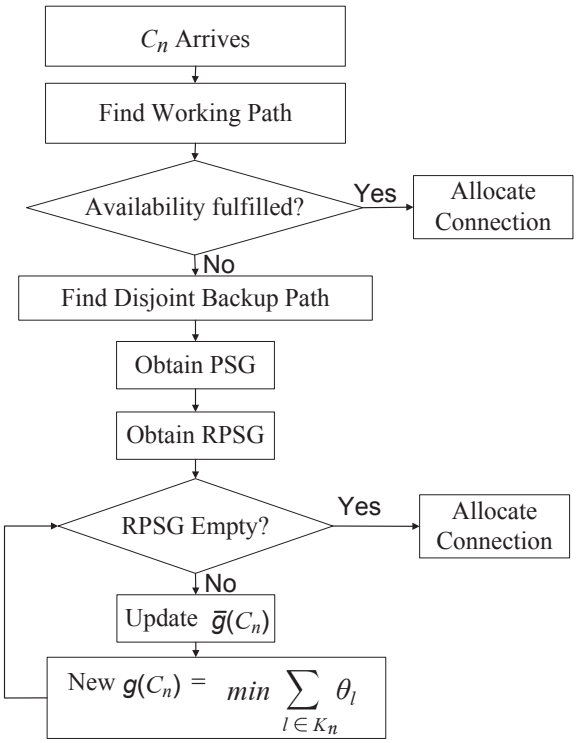


Figure 5. Flow Chart of the Allocation Mechanism.

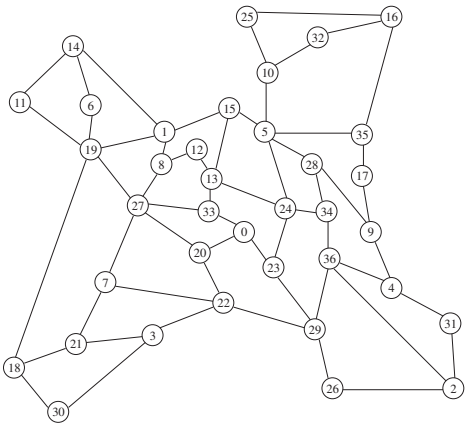


Figure 6. The COST266 Network Topology.

In the first experiment, we assumed that requests with random source and destination nodes arrive and depart at any time. Nevertheless, the *"number of allocated connections in the networks"* (load) is kept always constant at each point. Those allocated connections have bandwidth and availability requirement uniformly distributed

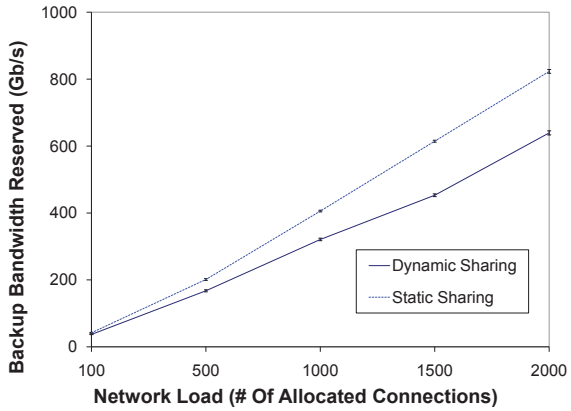


Figure 7. Bandwidth Utilization Under Different Network Loads.

with values among 1Mb/s - 150Mb/s and 0.991 - 0.999 respectively. Figure 7 shows the result of this experiment, where there is a clear more efficient resources utilization for all the possible number of allocated connections. Nevertheless, for bigger network loads the saved bandwidth becomes considerably bigger.

A second experiment studies the behavior of dynamic and static sharing schemes under different availability requirements. For each of the points shown in Figure 8, the load is 1000 connections with the same availability requirement (a_n) but random source, destination and bandwidth (b_n). In this experiment, we observe that for low availability requirements, the utilization on both schemes is similar, due to the connections may not need backup. The same tendency is observed for high availability requirements, where the connections need dedicated path protection instead of shared. That means that for all the areas where sharing is really needed, the performance of our mechanism is better.

7. Conclusions

This paper shows an improvement in the utilization of the network resources in a SBPP scheme, using the properties of continuous bandwidth assignment.

Considerable bandwidth amounts are saved through the use of our proposed mechanism. This advantage is valid for a wide range of connection availability requirements and network loads. Nevertheless, the complexity of our mechanism is approximately one order of magnitude bigger than the complexity obtained by the use of static sharing.

Finally, if the availability requirements are considerably high compared with the average link availability, both schemes will behave very similar given that under these conditions they will use a dedicated path protection scheme.

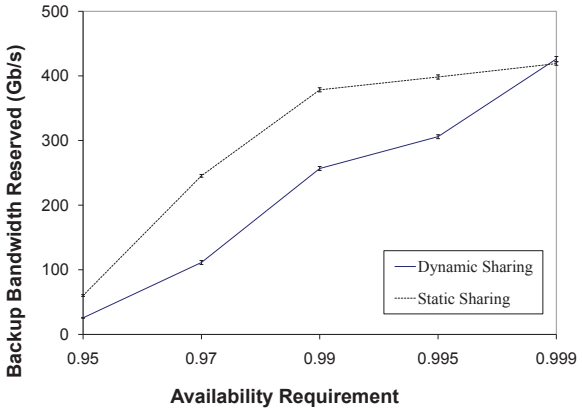


Figure 8. Bandwidth Utilization Under Different Availability Requirements.

References

[FT06] Andrea Fumagalli and Marco Tacca. Differentiated reliability (DiR) in wavelength division multiplexing rings. *IEEE/ACM Transactions on Networking*, 14(1):159–168, Feb. 2006.

[HLS07] Changcheng Huang, Minzhe Li, and A. Srinivasan. A scalable path protection mechanism for guaranteed network reliability under multiple failures. *IEEE Transactions on Reliability*, 56(2):254–267, Jun. 2007.

[HTH08] Pin-Han Ho, J. Tapolcai, and A. Haque. Spare capacity reprovisioning for shared backup path protection in dynamic generalized multi-protocol label switched networks. *IEEE Transactions on Reliability*, 57(4):551–563, Dec. 2008.

[KL00] Murali Kodialam and T. Lakshman. Dynamic routing of bandwidth guaranteed tunnels with restoration. In *Proceeding of the IEEE Nineteenth Annual Joint Conference of the Computer and Communications Societies (INFOCOM)*, volume 2, pages 902–911, 26–30 Mar. 2000.

[LLY09] Hongbin Luo, Lemin Li, and Hongfang Yu. Routing connections with differentiated reliability requirements in WDM mesh networks. *IEEE/ACM Transactions on Networking*, 17(1):253–266, Feb. 2009.

[MH08a] Anders Mykkeltveit and Bjarne E. Helvik. Comparison of schemes for provision of differentiated availability-guaranteed services using dedicated protection. In *Proceeding of the IEEE Seventh International Conference on Networking (ICN)*, Apr. 2008.

[MH08b] Anders Mykkeltveit and Bjarne E. Helvik. On provision of availability guarantees using shared protection. In *Proceeding of the IEEE/IFIP 12th Conference on Optical Network Design and Modelling (ONDM)*, Mar. 2008.

[MSW05] Darli A. Mello, Dominic A. Schupke, and Helio Waldman. A matrix-based analytical approach to connection unavailability estimation in shared backup path protection. *IEEE Communications Letters*, 9(9):844–846, Sep. 2005.

[RBS+01] Ramu Ramamurthy, Zbigniew Bogdanowicz, Shahrokh Samieian, Debanjan Saha, Bala Rajagopalan, Sudipta Sengupta, Sid Chaudhuri, and Krishna Bala. Capacity performance of dynamic provisioning in optical networks. *Journal of lightwave technology*, 19(1):40–48, Jan. 2001.

[SND] SNDlib. Library of test instances for Survivable fixed telecommunication Network Design. [Online]. Available: http://sndlib.zib.de.

[ZZZ+07] Jing Zhang, Keyao Zhu, Hui Zang, N. S. Matloff, and B. Mukherjee. Availability-Aware Provisioning Strategies for Differentiated Protection Services in Wavelength-Convertible WDM Mesh Networks. *IEEE/ACM Transactions on Networking*, 15(5):1177–1190, Oct. 2007.

PAPER G

Guaranteeing SLA Availability in Telecommunications Networks

Andres J. Gonzalez and Bjarne E. Helvik

Proceedings of the IEEE 2012 International Telecommunications Network Strategy and Planning Symposium NETWORKS

Rome, Italy, October, 2012

GUARANTEEING SLA AVAILABILITY IN TELECOMMUNICATIONS NETWORKS

Andres J. Gonzalez, Bjarne E. Helvik
Centre for Quantifiable Quality of Service in Communication Systems
Norwegian University of Science and Technology,
Trondheim, Norway
{andresgm,bjarne}@q2s.ntnu.no

Abstract

Network operators have to allocate connections fulfilling availability requirements stipulated in SLAs for a finite interval. However, modeling accurately the transient solution of repairable systems is still an open challenge. We study the SLA penalty scheme and propose a model to allocate connections with SLA requirements, maximizing the operator profit through a two-stage stochastic program. Our model considers the stochastic behavior of network components, correlation between failure/repair processes, the SLA finite duration, and the flexibility to allocate or reject a connection based on its commercial convenience. The model is designed for three different protection schemes: unprotected, dedicated backup and shared backup.

1. Introduction

Real world networks are not fault free. A single failure has negative impact on the providers reputation and profit. Therefore, the availability to be guaranteed has to be stipulated in Service Level Agreements known as SLA. This paper proposes a model to allocate connections, maximizing the operator profit and giving the flexibility to decide whether a connection should be accepted or not, i.e., it is commercially viable or not. The probability that a network operator meets the contracted interval availability α after the SLA period τ will be referred as *SLA success probability*. Many studies compute the connection availability based on steady state assumptions. However, this policy is inappropriate since it does not consider the probability density of the interval availability and the respective risk that it represents. This concept was first raised for general systems in [GT88] and verified for unprotected and protected connections in [GH09]. Failure correlation is a fact in telecommunications networks [GHHK10]. Correlation between failures may produce considerable availability degradation [GH11b]. These effects have to be avoided when reliable connections are provided.

In this paper, we formulate an stochastic optimization that targets directly the company revenue in the objective function. In addition, instead of dealing with steady

state values, we map all availability variables into the time domain. We formulate a two-stage stochastic program to handle this issue. First, we study the penalty function in an SLA. Second, we formulate the first-stage of our stochastic program as a conventional multi-commodity flow problem. Finally, we propose the entire stochastic optimization model that maximizes the operator profit, considering the interval availability. Since the cumulative downtime distribution is continuous in time, we choose the Sample Average Approximation [MMW99] to solve our model. We address three different kinds of connections: unprotected, shared backup protected and dedicated backup protected. We propose a method to compute the entire connection downtime in terms of individual link's downtime for each of these three cases. For this, we analyze the recursive nature of the end-to-end path downtime, and we propose an expansion method to perform this computation.

There is a rich amount of network optimization models in the literature. However, to our knowledge this is the first time that a model able to consider all the following details is proposed: (1) Interval availability, (2) Failure Correlation, (3) Realistic SLA penalty schemes, (4) Target directly the operator profit. The proposed model is solvable using CPLEX. It can be used either to directly optimize the network operator profit, or it can be used as reference to verify the accuracy of any future heuristic procedure when a given scenario demands a solution in short time. We present some procedures to be considered in the implementation of the proposed stochastic program, based on the solution of several case studies

This paper is organized as follows. In Section 2, we describe the proposed stochastic optimization model . Section 3 shows the method to calculate the end-to-end connection downtime under three different protection schemes. Section 4 presents some considerations on the implementation of the optimization models. Finally, Section 5 concludes the paper.

2. Stochastic Optimization Model

2.1 Problem Formulation

In this paper, we investigate the problem of allocating multiple requests in a network, considering SLA availability constraints. This may be modeled for deterministic scenarios as a *multi-commodity flow problem*. A variety of solutions of related deterministic problems can be found in [PM04].

We consider: i) A network graph $G = (V, L)$ where V is a set of nodes and L is a set of links (i, j) $i, j \in V$, with utilization cost per bandwidth unit $c_{i,j}$, and maximum capacity $W_{i,j}$. ii) A set of requests $(1, ..k, ..K)$ with defined source s_k, destination d_k, and bandwidth needed b_k. Our approach differs from the conventional multi-commodity flow, since we consider the stochastic penalty cost that depends on the stochastic behavior of network links. Instead of dealing directly with availability variables, the key idea of this paper is to map the requirements into the time domain. We define the maximum allowed cumulated downtime before penalty as $T = (1 - \alpha)\tau$. In addition to the penalty, any single failure before T affects the reputation of the network provider. For this reason, our model considers two different kinds of cost: the

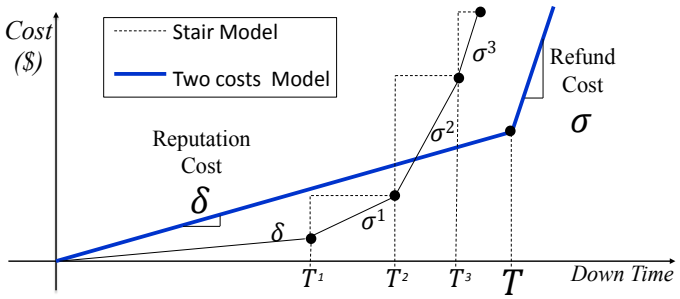


Figure 1. Cost of the cumulative down time of a network connection.

reputation cost δ and the refund cost σ. δ may be estimated using marketing models. The refund cost may be modeled with a constant slope or using a stair function, where after each step $r = \{1,...,R\}$, the refund slope σ^r is incremented. Fig. 1 shows the two schemes used to model penalty costs.

We define our problem as a two-stage stochastic program. First, we explain the deterministic approach. Then, we extend our model by formulating our stochastic optimization program. For the sake of simplicity, the problem is initially formulated as a minimization of the penalty cost. Then, it is extended to a maximization of the operator profit, where some connections may be rejected if they are not commercially viable.

2.2 First-Stage Problem

The multi-commodity flow formulation in order to provide a single path to each request without traffic splitting is:

$$\min \sum_{k=1}^{K} \sum_{(i,j)\in L} c_{i,j} x_{i,j}^k b_k \tag{1}$$

subject to constraints (s.t.)

$$\sum_{(j)\in V} x_{i,j}^k - \sum_{(j)\in V} x_{j,i}^k = \begin{cases} 1 & i = s \\ -1 & i = d \\ 0 & i \neq s,d \end{cases} \quad k = \{1,..,K\} \tag{2}$$

$$\sum_{k=1}^{K} b_k x_{i,j}^k \leq W_{i,j} \qquad \forall (i,j) \in L \tag{3}$$

$$x_{i,j}^k \in \{0,1\} \tag{4}$$

Constraint (2) controls the flow balance using the Kirchhoff law on each node. Constraint (3) keeps the solution inside the network capacity limits. Finally, constraint (4) defines $x_{i,j}^k$ as binary variables in order to avoid path bifurcation.

The multi-commodity flow that provides working path X and backup path Y is formulated as

$$\min \sum_{k=1}^{K} \sum_{(i,j)\in L} c_{i,j}(x_{i,j}^k + y_{i,j}^k) \tag{5}$$

$$\text{s.t. } x_{i,j}^k + y_{i,j}^k \leq 1 \tag{6}$$

Constraint (6) guarantees disjoint working and backup paths. In addition, the binary constraints, the network capacity limit, and the Kirchhoff equations have to be included.

Previous multi-commodity flow models do not take into account the penalty generated by downtime disconnections. They assume the knowledge of the steady state availability of network links and use integer linear programming to obtain the most reliable set of paths, fulfilling availability constraints (See for instance [CSK02] and [LML10]). Nevertheless, the solutions of these models are only valid in steady state and do not target company revenues. We propose a two-stage stochastic model that considers both, i) the transient nature of the problem posed by the temporal duration of the SLA, ii) the network operator profit.

2.3 Two-stage Stochastic Program Formulation

When the availability requirement and the failure/repair processes of a network component are mapped into the time domain, the formulation that takes into account the transient and stochastic nature of the problem becomes easier. We found that the task of allocating network connections under SLA constraints may be modeled as a two-stage Stochastic Program (In [BL97] a complete description of this kind of programs is presented). The models described in (1)-(5) may be interpreted as the first-stage problem, where paths for each request are obtained. The second-stage performs rerouting decisions, depending on the realization of ξ, which will be defined as a vector of random variables that contains the cumulative down time of each of the network links during a SLA period, i.e., $\xi = \langle \hat{t_1}(\tau), \hat{t_2}(\tau), ..., \hat{t_l}(\tau)..., \hat{t_L}(\tau) \rangle \ \ l \in L$. The cumulative downtime of a connection has a direct relation with the penalty to be paied. At the same time, it depends directly on the downtime of each of the individual links that belong to the paths assigned to that connection. Using this concept, we formulate a stochastic program that minimizes the operator penalty cost. In the formulation of the stochastic program that provides single path per connection, constraints such as flow conservation, capacity limit, and binary variables still apply. In addition, the rest of the problem is formulated as

MODEL I

$$\min \quad \sum_{k=1}^{K} \sum_{(i,j)\in L} c_{i,j} x_{i,j}^k + E_\xi \left(\sum_{k=1}^{K} \delta_k P1_k + \sigma_k P2_k \right) \tag{7}$$

$$\text{s.t.} \qquad P1_k + P2_k \geq \hat{\Psi}_\xi^k \ , \ k = \{1, ..K\} \tag{8}$$

$$P1_k \leq T_k \tag{9}$$

where $P1_k$ represents the downtime that connection k accumulates under the reputation cost and $P2_k$ the downtime that connection k accumulates under the refund cost. Constraints (8) and (9) model the penalty cost, where $\hat{\Psi}^k_\xi$ is the total cumulated downtime of path k which is a stochastic variable. If the realization of $\hat{\Psi}^k_\xi$ (Ψ^k) is less or equal than T_k, $P1_k = \Psi^k$ and $P2_k = 0$ otherwise $P1_k = T_k$ and $P2_k = \Psi^k - T_k$.

This model can be extended to the stair penalty function shown in Figure 1 as follows.

$$\min \quad \sum_{k=1}^{K} \sum_{(i,j)\in L} c_{i,j} x^k_{i,j} + E_\xi \Big(\sum_{k=1}^{K} \delta_k P1_k + \sum_{r=1}^{R} \sigma^r_k P^r_k \Big) \tag{10}$$

$$\text{s.t.} \qquad P1_k + \textstyle\sum_{r=1}^{R} \sigma^r_k P^r_k \geq \hat{\Psi}^k_\xi \tag{11}$$

$$P^r_k \leq T^r_k - T^{r-1}_k \tag{12}$$

The definition of $\hat{\Psi}^k_\xi$ is one of the key concepts of the paper and it will be explained in detail in the next section.

A different model has to be developed when the stochastic program provides working and backup paths. If constraints (2), (3), (4) and (6) are kept, the model may be formulated as

MODEL II

$$\min \sum_{k=1}^{K} \sum_{i,j\in L} c_{i,j}(x^k_{i,j} + y^k_{i,j}) + E_\xi \sum_{k=1}^{K} \delta_k P1_k + \sigma_k P2_k$$

$$\text{s.t. } P1_k + P2_k \geq \hat{\Phi}^k_\xi \tag{13}$$

$$P1_k \leq T_k \tag{14}$$

The above structure is similar to the one used for single path connections. However, the cumulative down time of connection k $\hat{\Phi}^k_\xi$ is a random variable that depends on two disjoint paths, making its formulation more complex. The definition of $\hat{\Phi}^k_\xi$ will be also explained in detail in the next section.

The random processes that affect our stochastic program can be measured and characterized from an operational network (see for instance [GH11a]). We choose the Sample Average Approximation (SAA) method in order to solve the formulated problem for two reasons. First, it makes finite the original infinite number of scenarios posed by the continuous PDFs. Second, it provides a way to control the solution quality by defining lower and upper bounds that monotonically improves as the size of the sample increases [MMW99]. Using the procedures defined by SAA, we define the specific lower bound of our problem as follows:

MODEL I LB

$$\min \sum_{k=1}^{K} \sum_{(i,j)\in L} c_{i,j} x_{i,j}^{k} + \frac{1}{S}\left(\sum_{s=1}^{S} \sum_{k=1}^{K} \delta_k P1_k^s + \sigma_k P2_k^s\right)$$

$$\text{s.t.} \qquad P1_k^s + P2_k^s \geq \Psi_s^k \tag{15}$$

$$P1_k^s \leq T_k \ , \ s = \{1,..S\} \ , \ k = \{1,..K\} \tag{16}$$

where S represents a group of independent samples of vector ξ, and constraints (15) and (16) are the extension of constraints (8) and (9) to each individual scenario s.

MODEL I LB shows in *extensive form* the stochastic optimization used to obtain the lower bound. However, depending on the bound's gap, the size of S may be very large (In SAA the lower and the upper bound gap monotonically improves as the size of the sample increases). Therefore, alternative procedures such as the Branch-and-Cut method [VAK$^+$03] or the Integer L-Shaped Method [BL97] may be used in order to solve MODEL I LB more efficiently.

By solving MODEL I LB, we obtain a solution X^* that provides the paths to be assigned. By definition, the objective function evaluated in X^* is smaller than or equal to the objective function evaluated in the real optimal solution. We use the method suggested in [VAK$^+$03] to improve the quality of solution X^*. We solve MODEL I LB in N independent replications, where each replication contains S independent samples of ξ. After this, we select the solution X_n^* with the smallest objective value. Once a potential solution X_n^* is selected (X^*=X_n^*), the upper bound is calculated as follows:

MODEL I UB

$$\sum_{k=1}^{K} \sum_{(i,j)\in L} c_{i,j} x_{i,j}^{*k} + \frac{1}{M} \sum_{s=1}^{M} \sum_{k=1}^{K} \delta_k P1_k^s + \sigma_k P2_k^s \tag{17}$$

$$P1_k^s + P2_k^s \geq \Psi_s^k \tag{18}$$

$$P1_k^s \leq T_k \ , \ s = \{1,..M\} \ , \ k = \{1,..K\} \tag{19}$$

where M represents the number of independent samples of ξ. Given that MODEL I UB does not imply any minimization process, but only sampling and direct calculations, the size of M can be much larger than S.

The final extension of our model includes the selling price F_k for allocating connection k, and the possibility to reject the request given that it may not be convenient in terms of the network operator profit. For this, additional binary variables f_k are included. $f_k = 1$ if the network operator decides to sell and allocate connection k, or zero otherwise. The model is defined as

$$\max \sum_{k=1}^{K} F_k f_k - \sum_{k=1}^{K} \sum_{(i,j)\in L} c_{i,j} x_{i,j}^{k}$$

$$-E_\xi\left(\sum_{k=1}^{K} \delta_k P1_k + \sigma_k P2_k\right) \tag{20}$$

s.t.

$$\sum_{(j)\in V} x_{i,j}^k - \sum_{(j)\in V} x_{j,i}^k = \begin{cases} f_k & i = s \\ -f_k & i = d \\ 0 & i \neq s,d \end{cases} \quad ,k = \{1,..K\} \tag{21}$$

Constraints (3), (4), (8) and (9) must be included.

3. Connection Downtime Under Different Protection Schemes

The proposed optimization model is viable only if we are able to characterize $\hat{\Psi}_{\xi}^{k}$ and $\hat{\Phi}_{\xi}^{k}$. In this section, we define the realization of the cumulative down times Ψ^k and Φ^k for unprotected and protected connections, respectively, based on samples of $\hat{f}(\tau)$ on each of the links. The proposed solution not only allows the complete definition of MODEL I and II, but it also keeps the problem linear.

3.1 Unprotected Connections

For unprotected connections, Ψ^k is equal to the single path downtime. The path is formed by a series of interconnected links. A failure in any of them affects the connectivity of the whole path. We found that the cumulative downtime of a path with n links $\Psi(n)$ can be expressed in terms of the downtime of link n t_n and $\Psi(n-1)$, describing a recursive nature. This is illustrated in Figure 2. Using this principle and the binomial theorem, we obtain an expression for Ψ^k as follows:

$$\Psi^k = \sum_{\zeta=1}^{L-1} (-1)^{(\zeta+1)} \left[\sum_{u=\zeta}^{L} \sum_{v=1}^{\binom{u-1}{\zeta-1}} t_u P(\bigcap\nolimits^v | u) \right] \tag{22}$$

where ζ represents the order of the expression, u is an index that represents the link under evaluation, v is an index used to keep track of the possible link combinations depending on the current value of ζ, $\bigcap^v$ is the downtime intersection of all links that belong to a specific combination $\binom{u-1}{\zeta-1}$, and $P(\bigcap^v | u)$ is the probability of $\bigcap^v$ given that link u is down .

Expression (22) may become very large with the increase of the number of links in the path. In addition, the higher the order ζ, the smaller the influence in the path downtime. We will expand (22) giving a physical interpretation to each of the terms in a given order ζ.

When $\zeta = 1$, the downtime of each of the links that belongs to the path is considered separately, without taking into account the overlapping that may exist between two downtimes. In this case, the path downtime including only first order terms Ψ_1^k is the sum of each individual t_l ($l = \{1,2,..L\}$). Ψ_1^k is an upper bound ($\Psi_1^k \geq \Psi^k$) of the real path downtime given that each possible downtime intersection reduces this value. Using the first order approximation Ψ_1^k, constraint (8) may be redefined as

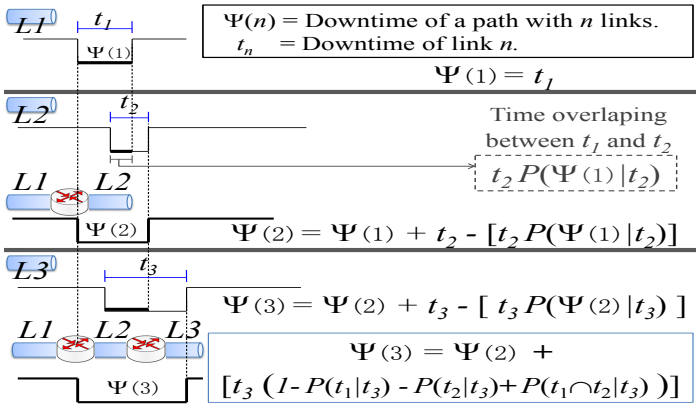


Figure 2. Recursive nature of the path cumulative downtime.

$$P1_k + P2_k \geq \sum_{(i,j)\in L} t_{i,j} x^k_{i,j} \tag{23}$$

Second order terms ($\zeta = 2$) take into account the overlapping downtime between two links. When first and second order terms are considered (Ψ^k_2), Ψ^k is approximated better. Ψ^k_2 is a lower bound ($\Psi^k_2 \leq \Psi^k$) of the real path downtime given that all pair-intersection are subtracted without considering possible double counting produced by the overlapping downtime between three or more links. Constraint (8) may be redefined using first and second order terms as

$$P1_k + P2_k \geq \sum_{(i,j)\in L} t_{i,j} x^k_{i,j} - \sum_{p=1}^{L-1} \sum_{q=p+1}^{L} t_p P(q|p) z^k_{p,q} \tag{24}$$

where p and q are indexes that represent down states of network links (a,b) and (c,d) respectively, and $z^k_{p,q}$ is a matching binary variable equal to one only if both links ((a,b) and (c,d)) belong to the path assigned to request k, i.e., $x^k_{a,b} = 1$ and $x^k_{c,d} = 1$. In order to fulfill the last condition we add the following constraints to the model

$$\begin{gathered} z^k_{p,q} \geq x^k_p + x^k_q - 1 \quad \forall \ \{p,q\}\ (p \neq q) \in L^2 \\ z^k_{p,q} \leq x^k_p \quad , \quad z^k_{p,q} \leq x^k_q \end{gathered} \tag{25}$$

Higher order terms can be included. This can be made by expanding (22) and by using additional matching binary variables $z^k_{p_1,p_1,..,p_\zeta}$ for each new order ζ included. $z^k_{p_1,p_1,..,p_\zeta}$ must be equal to one only when all links of a specific combination $\binom{L}{\zeta}$ belong to the path assigned to request k.

3.2 Dedicated Path Protection

Protection is one of the most common strategies to fulfill dependability requirements. Dedicated path protection is a scheme where the capacity used by the working and the backup path is reserved exclusively for one connection. In this section, we develop the term Φ^k in (13), in order to define explicitly MODEL II. The methodology that we use to define Φ^k follows the same expansion principle used for the definition of Ψ^k. A network connection that uses dedicated backup goes down when any of the links of the working path goes down and the backup path is not available. We use indexes p and q to represent down states of links (i,j) and (a,b) in the working and backup path, respectively. First order terms are derived by evaluating one by one the down time t_p on each of the links that belongs to the working path X ($x_{i,j}=1$). During each interval t_p, we evaluate if any of the links that belongs to the backup path ($y_{a,b}=1$) is down . We redefine Φ^k in constraint (13) using first order terms as follows

$$P1_k + P2_k \geq \sum_{p=1}^{L}\sum_{q=1}^{L} t_p P(q|p) z_{p,q}^k \tag{26}$$

where $z_{p,q}^k$ is a binary variable used to capture the match when link (i,j) belongs to the working path and link (a,b) belongs to the backup path. This is made by adding the following constraints to the model

$$z_{p,q}^k \geq x_p^k + y_q^k - 1 \;\;,\;\; z_{p,q}^k \leq x_p^k \;\;,\;\; z_{p,q}^k \leq y_q^k \tag{27}$$

Φ^k can also be defined using first and second order terms. The second order terms are obtained by subtracting possible overlapping between down states f and g of links that belong to the working path, but keeping track of the down state h of links that belong to the backup path. Constraint (13) may be redefined using first and second order terms as

$$\begin{aligned} P1_k + P2_k \geq & \sum_{p=1}^{L}\sum_{q=1}^{L} t_p P(q|p) z_{p,q}^k - \\ & \sum_{f=1}^{L-1}\sum_{g=f+1}^{L}\sum_{h=1}^{L} t_f P(g\bigcap h|f) z_{f,g,h}^k \\ \text{s.t.} \quad & z_{f,g,h}^k \geq x_f^k + x_g^k + y_h^k - 2 \\ & z_{f,g,h}^k \leq x_f^k \;,\; z_{f,g,h}^k \leq x_g^k \;,\; z_{f,g,h}^k \leq y_h^k \end{aligned} \tag{28}$$

Third order terms consider the time overlapping between two links of the backup path. These terms have to be also subtracted from (26), given that $P(\text{Backup Path}|p) < \sum_{q=1}^{L} P(q|p)$. The expression for third order terms has similar structure to the second order terms, but in the additional part, indexes g and h would refer links in the backup path, i.e., $z_{f,g,h}^k \geq x_f^k + y_g^k + y_h^k - 2$.

3.3 Shared Path Protection

In Shared Backup Path Protection (SBPP), a connection may share bandwidth in its backup path with other connections. In this case, the offered availability not only depends on the network resources assigned to one connection, but also on the state of other connections. The way to access the backup resources may be using FIFO scheduling. Under this scheme, a connection goes down when at least one link p that belongs to the working path goes down and the backup path is unavailable. The backup path is unavailable either (i) if any link q in the backup is down, or (ii) if the backup is already in use, which means that any of the links g of the working path of another request is down. We may redefine constraint (13) for a SBPP scheme using first order terms as follows

$$P1_k + P2_k \geq \sum_{p=1}^{L} \sum_{q=1}^{L} t_p P(q|p) z_{p,q}^k + \sum_{p=1}^{L} \sum_{g=1(g\neq p)}^{L} t_p P(g|p) z_{p,g}^k \tag{29}$$

The first part of expression (28) is similar to (26). The last part of the equation contains an additional component that corresponds to the down time added by sharing. In this way, constraint (13) may also be redefined with first and second order terms for shared path protection, using a similar structure to (27) and adding the downtime caused by sharing.

4. Implementation Considerations

In order to have a tighter estimation of the expected operator profit, we implement a Monte Carlo simulation that computes precisely the expected operator profit based on an obtained solution X^*. We develop a platform in DEMOS [Bir03], where failure/repair events in network links are injected. With the knowledge of the connection-path, the simulation can recreate the end-to-end path downtime and the expected profit.

We will present a case study that illustrates the importance of considering the entire PDF of downtimes instead of only the steady state link availability A_l. We implement our model in CPLEX to allocate an unprotected connection between Stavanger and Tromso, using the current topology of the operational UNINETT network [UNI12b]. We assume a selling price $F = \$100000$, the reputation cost is \$600 per disconnection hour, the refund penalty is \$8000 per disconnection hour, and $T = 8$ hours. Considering only steady state values, we obtain that the optimal path is Stavanger-Kristiansand-Oslo-Trondheim-Tromso (Path1) with 0.99862 asymptotic path availability. On the other hand, considering the PDF, we obtain that the optimal path is Stavanger-Bergen-HoytekB-Trondheim-Tromso (Path2) with 0.99853 asymptotic path availability.

Keeping the same asymptotic values in all cases, we will evaluate the operator profit when: i) Path 2 is assigned (Optimal solution considering the PDF). The links

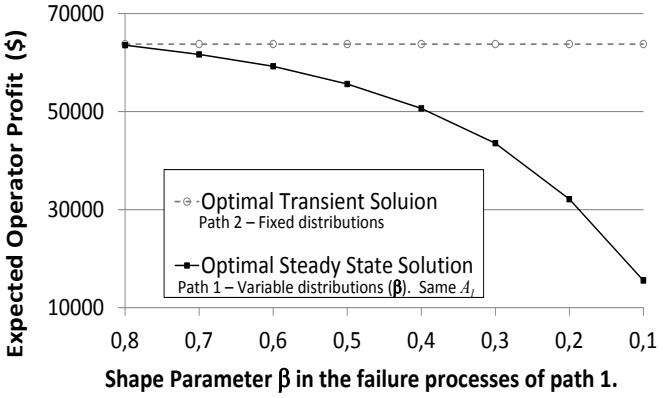
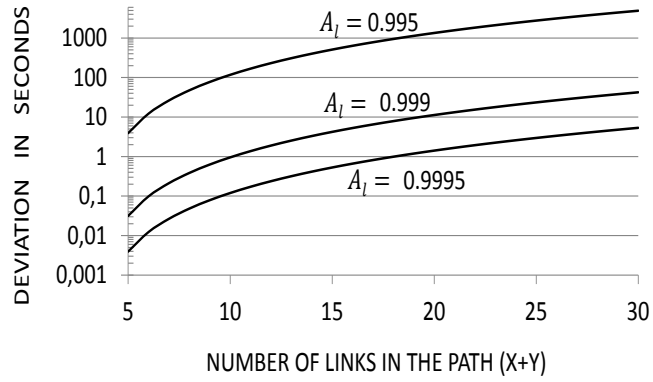


Figure 3. Expected operator profit using: i) the optimal path in steady state with different $\Omega(\tau,t)$ but same A_l. ii) the optimal path in the transient solution with fixed up and down time distributions.

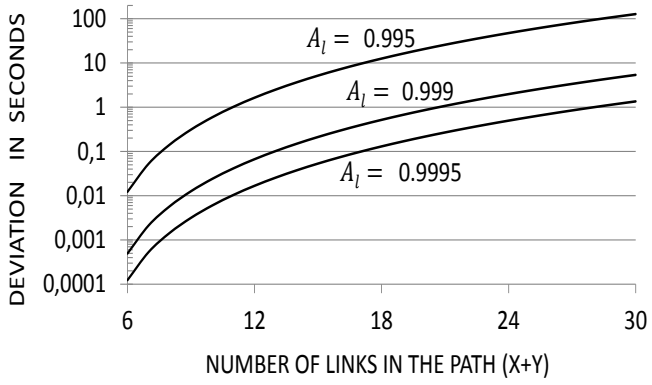
of this path have fixed gamma distributed uptimes and downtimes with parameters $\beta = 0.8$, $\theta = 74$ days, and $\beta = 1$, $\theta = 1$ hour, respectively. ii) Path 1 is assigned (Optimal solution in steady state). The links of this path have gamma distributed downtimes with parameters $\beta = 1$, $\theta = 1$ hour. In order to evaluate the effect of different cumulated downtime PDFs with the same steady state, we variate the shape parameter of the uptime distribution of path 1 from 0.8 to 0.1, tuning θ in order to have the same asymptotic path availability (0.99862) in all cases. Figure 3 presents the expected operator profit using fixed parameters for path 2 and the described variable shape parameter for path 1. This case study shows the huge economical impact if network operators do not have at hand a tool able to consider the entire probability distribution of the failure/repair processes.

Using CPLEX, we obtain path protected solutions using first and second order terms. Figure 4 illustrates the average time deviation produced by the approximation during one year (τ) by computing $[(\frac{\Phi^k}{\Phi_1^k})(t(\tau))]$ and $[(\frac{\Phi_2^k}{\Phi^k})(t(\tau))]$. Usually, links in operational networks present A_l values higher than 0.999 ([MIB$^+$08] and [GH11a]). In addition, SLA time-requirements have a granularity in the order of minutes or hours. Under these conditions (high reliable links and SLA granularity in minutes), the first order provides a good approximation, as Figure 4 illustrates. For more extreme cases, the second order approximation is enough. However, our model can be scaled in order to deal with any kind of scenario. Finally, given that Φ_1^k and Φ_2^k define an upper and lower bound respectively, a way to verify the optimal solution is by comparing the solutions obtained using first and second order approximation. If the solution is the same, further orders are not needed.

When correlation exists, $P(\bigcap^v|u)$ could be big even for high orders ζ. In this case not all the combinations $\{p_1,..p_\zeta\} \in \binom{L}{\zeta}$ should be considered (i.e., be expanded), but only those values that due to correlation are relevant. This makes the expansion method more efficient without loosing accuracy.



(a) Time deviation of Φ_1^k.



(b) Time deviation of Φ_2^k.

Figure 4. Error (in seconds) posed by our approximation of the path cumulative down time for a protected connection.

We implement our optimization model in protected connections, using different conditional probabilities, obtaining similar profit deviations to the one illustrated in Figure 3. Our stochastic program avoids the assignation of main and backup paths with correlated failures, but goes beyond previous approaches by allowing this kind of scenarios only if they are economically viable, e.g. if there is not enough resources to provide uncorrelated paths, but the profit (including penalties) is better than the one received by providing nothing or unprotected paths.

5. Conclusion

We introduced a resource allocation model that can be used when SLA requirements have to be fulfilled. Important factors such as the temporal duration of an SLA, the stochastic behavior of network components, the correlation between failure/repair events and the cost generated by resources utilization, refund penalties, and bad repu-

tation have to be considered to allocate connections in a network. To our knowledge, there is not a previous optimization model in the literature able to integrate all these factors, being our work the first in this direction.

The proposed stochastic programs are solvable in CPLEX. Therefore, it can be used either to directly optimize network operator profit, or it can be used as a reference to verify the accuracy of any future heuristic procedure, when given the dynamic of a certain scenario, the solution is required in short time. In addition, we proposed the use of a Monte Carlo simulation that complement the optimization results, obtaining a tighter estimation of the expected operator profit, offering to network providers the possibility to make better finance planning.

The proposed model avoids the assignation of main and backup paths with correlated failures, but goes beyond previous approaches by allowing this kind of scenarios only if they are economically viable.

References

[Bir03] Graham M. Birtwistle. *DEMOS - a system for Discrete Event Modelling on Simula*. University of Leeds, 2003.

[BL97] John R. Birge and Francois Louveaux. *Introduction to Stochastic Programming*. Springer Series in Operations Research and Financial Engineering. Springer, 1997.

[CSK02] Weidong Cui, Ion Stoica, and Randy H. Katz. Backup path allocation based on a correlated link failure probability model in overlay networks. In *Proceeding of the IEEE International Conference on Network Protocols (ICNP)*, 2002.

[GH09] Andres J. Gonzalez and Bjarne E. Helvik. Guaranteeing service availability in SLAs; a study of the risk associated with contract period and failure process. *Proceeding of the IEEE Latin-American Conference on Communications (LATINCOM)*, pages 1 –6, Sep. 2009.

[GH11a] Andres J. Gonzalez and Bjarne E. Helvik. Analysis of failures characteristics in the UNINETT IP backbone network. *Proceeding of the IEEE 7th International Symposium on Frontiers in Networking with Applications (FINA)*, Mar. 2011.

[GH11b] Andres J. Gonzalez and Bjarne E. Helvik. Guaranteeing Service Availability in SLAs on Networks with Non Independent Failures. *IEEE-IFIP International Workshop on Design of Reliable Communication Networks (DRCN)*, Oct. 2011.

[GHHK10] Andres J. Gonzalez, Bjarne E. Helvik, Jon K. Hellan, and Pirkko Kuusela. Analysis of dependencies between failures in the UNINETT IP backbone network. *Proceeding of the IEEE International Symposium on Pacific Rim Dependable Computing (PRDC)*, 2010.

[GT88] Ambuj Goyal and Asser Tantawi. A measure of guaranteed availability and its numerical evaluation. *IEEE Transactions on Computers*, Volume 37, Issue 1:25 – 32, 1988.

[LML10] Hyang-Won Lee, E. Modiano, and Kayi Lee. Diverse routing in networks with probabilistic failures. *IEEE/ACM Transactions on Networking*, 18(6):1895 –1907, Dec. 2010.

[MIB+08] Athina Markopoulou, Gianluca Iannaccone, Supratik Bhattacharyya, Chen-Nee Chuah, Yashar Ganjali, and Christophe Diot. Characterization of Failures in an Operational IP Backbone Network. *IEEE/ACM Transactions on Networking*, 16(4):749–762, Aug. 2008.

[MMW99] Wai-Kei Mak, David P. Morton, and Kevin R. Wood. Monte Carlo bounding techniques for determining solution quality in stochastic programs. *ELSEVIER Operations Research Letters*, 24:47–56, 1999.

[PM04] Michal Pióro and Deepankar Medhi. *Routing, Flow, and Capacity Design in Communication and Computer Networks*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2004.

[UNI12] UNINETT. The Norwegian Research Network. Network Topology. [online]. Available at: http://drift.uninett.no/stat-q/load-map/uninett,,traffic,peak. 2012.

[VAK+03] Bram Verweij, Shabbir Ahmed, Anton J. Kleywegt, George Nemhauser, and Alexander Shapiro. The Sample Average Approximation Method Applied to Stochastic Routing Problems: A Computational Study. *Springer Computational Optimization and Applications*, 24(2):289–333–333, Feb. 2003.

PAPER H

System Management to Comply with SLA Availability Guarantees in Cloud Computing

Andres J. Gonzalez and Bjarne E. Helvik

Proceedings of the IEEE International Conference on Cloud Computing Technology and Science CloudCom, Taiwan, December, 2012

Taipei, Taiwan, December, 2012

SYSTEM MANAGEMENT TO COMPLY WITH SLA AVAILABILITY GUARANTEES IN CLOUD COMPUTING

Andres J. Gonzalez, Bjarne E. Helvik
Centre for Quantifiable Quality of Service in Communication Systems
Norwegian University of Science and Technology,
Trondheim, Norway
{andresgm,bjarne}@q2s.ntnu.no

Abstract

SLAs are common means to define specifications and requirements of cloud computing services in business relationships. The terms that define the guaranteed availability for a given period are fundamental to these contracts. In this context, a natural question for cloud providers is: How to guarantee the availability promised? This paper studies the level of availability offered to a virtual machine during an SLA period in clouds with different: size, redundancy, and fault tolerance techniques. Finally, this paper proposes the use of the *SLA-budget* for the implementation of smart policies in: **i)** the assignment of spare servers when virtual machines are restored. **ii)** the dynamic use of different fault tolerance licenses. Using such policies results in a considerable reduction of the probability of breaching the SLA guarantee, by making an efficient use of the cloud resources available. This paper is a first step in the design of SLA-aware cloud architectures.

1. Introduction

This paper is focused on Infrastructure as a Service (IaaS), where operators provide virtual machines with a given memory, processor speed, and storage capacity [MG11], [SMLF09]. Cloud computing environments are not fault free. Failures are unavoidable events that occur according to stochastic processes that are usually difficult to model. Any single failure impacts the cloud provider reputation and finances, and may have significant consequences for the customers through the affected applications. A common policy to handle this issue is the stipulation of the availability to be guaranteed in a business contract known as Service Level Agreement SLA (see for instance [Ama08]). The contracted availability must be commercially competitive, and it must fit the customer needs. However, provisioning highly reliable services requires the use of expensive infrastructure and resources.

In this context, an important question that cloud providers must answer is: *How to asses the amount and kind of resources that need to be provided to a customer in order to guarantee the availability promised?* In this paper, we study three relevant aspects

to keep the availability offered according to the SLA. **i)** Fault tolerance features of the virtualization platform. **ii)** Redundancy and size of the computing center providing cloud services. **iii)** Policies for the efficient use of the cloud resources.

Virtual machines (VM) are deployed in physical data centers composed by network infrastructure, computing facilities and storage systems, using a virtualization platform. Virtualization is a technology that started in the 1970's [Gol74], [PG73], and that has acquired a significant importance nowadays with the development of cloud computing. It allows the emulation of independent logical computer facilities, with memory, storage and computational capacity, where operating systems, applications and services can be executed. A mayor advantage is the independence of the physical hardware, since any affected virtual machine may be restored on any another working server [BDF$^+$03], [MUKX06]. The mechanisms used to restore virtual machines are hot research topics, where several approaches are proposed, e.g., [DHJ$^+$07], [CLM$^+$08], [LJL$^+$09], [SFK$^+$09]. In this paper, we select virtualization platforms commercially available, according to their market share and their diversity in the use of fault tolerance techniques. The first is vSphere Fault Tolerance (FT), which uses active replication of virtual machines [VMw09]. The second is vSphere High Availability (HA) [VMW07]. Finally, vSphere Data Recovery is considered [VM-09].

Cloud computing environments may have large scale failures, where several servers may be affected at the same time. See for instance, [FLP$^+$10] and [GJN11]. When simultaneous or correlated failures occur, the number of spare servers in the data center may not be sufficient to restore all the affected virtual machines. This raises two important issues that cloud providers should take into account. First, planning for an adequate spare capacity. Second, the priority scheme that the virtualization platform should follow in case of insufficient spare capacity.

We simulate several cloud computing scenarios, in order to study the effects of redundancy, cloud size, and virtualization platform, in the availability offered to a virtual machine. In the first simulation setting, the virtualization platform is fixed, and the cloud size and redundancy level are variable. In this setting, independent and simultaneous failures arrive according to randomly distributed times, affecting one or several servers at the same time. In the second simulation setting, the cloud size is fixed, and the redundancy level and virtualization platform are variable. The results of these simulations are the different expected availabilities offered to virtual machines allocated on each of the studied scenarios.

The expected availability does not give enough information to assess *the probability of breaching the availability promised in an SLA* (*risk*). For instance, [GT88] and [GH09] show that the risk may be considerable, even in scenarios where the expected availability of the system is larger than the availability promised. In addition, this paper will show that when two cloud computing environments have identical infrastructure and failure/repair processes, but different restoration policies, the expected availability of both systems is identical, but the SLA-risk is very different. In conclusion, the study of the distribution of the interval availability is needed.

In [MH09], a new model known as *Adaptive Provisioning with Preemption* was proposed. In that work, the remaining time that a connection has before failing

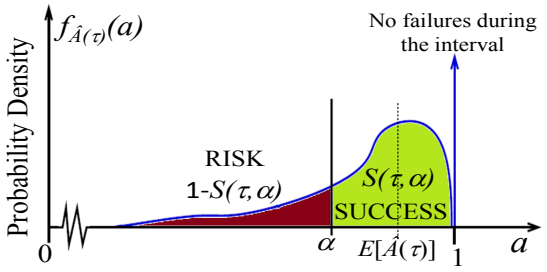


Figure 1. General distribution of the interval availability and its implications in an SLA context

the contracted availability is evaluated, in order to define priorities for accessing shared resources in MPLS networks. The implementation of this mechanisms in real networks is hard, mainly due to the complexity posed by the definition of uniform sharing groups, the number of network components involved in a network connection, and the dynamism of the network. On the other hand, in cloud computing scenarios, virtual machines may be deployed on any hardware, and hence, shared resources can be uniformly and easily defined.

Based on [MH09], we propose two policies to efficiently use cloud resources and reduce the risk of failing the SLA guarantee. The first policy proposes a priority scheme for the assignment of spare servers to the virtual machines that need to be restored. The second policy proposes a priority scheme for the utilization of different fault tolerance licences, according to the current needs of the system.

This paper is organized as follows. In Section 2, we present issues related to the distribution of the interval availability and SLA risk. Section 3 describes relevant dependability features in a cloud computing environment, and it analyzes the impact of redundancy, cloud size and fault tolerance technique in the availability offered to a virtual machine. In Section 4, two policies to reduce the SLA risk are proposed. They are studied in single class and different classes environments. Finally, Section 5 concludes the paper.

2. Interval Availability and SLAs

The interval availability becomes fundamental in business scenarios where an availability guarantee of α for a contract period τ is defined in an SLA.

The state of a virtual machine can be modeled as a function of time with a random process $O(t)$ defined as follows:

$$O(t) = \begin{cases} 1 & \text{If the virtual machine is working.} \\ 0 & \text{Otherwise.} \end{cases} \tag{1}$$

The interval availability $\hat{A}(\tau)$ is an stochastic variable that measures the percentage of time that a cloud service has been working during the defined period τ, i.e.:

$$\hat{A}(\tau) = \frac{1}{\tau}\int_0^{\tau} O(t)\,dt. \tag{2}$$

When the transient behavior of a virtual machine is studied, the expected interval availability $A(\tau) = E[\hat{A}(\tau)]$ is usually evaluated, given that computing the first moment of a random variable is simpler than obtaining the entire probability distribution. When an SLA is defined, the probability that the availability after τ will be larger than or equal to α becomes a parameter of especial interest. The evaluation of the expected interval availability is not enough for the estimation of such probability, and hence, the entire probability distribution has to be considered.

Figure 1 shows the general shape of the Probability Density Function (PDF) of the interval availability. It is defined as $f_{\hat{A}(\tau)}(a)$, and as Figure 1 illustrates, it may be used to calculate the SLA risk and success.

The *SLA Success* Probability is defined as:

$$S(\tau,\alpha) = \int_{\alpha}^{1} f_{\hat{A}(\tau)}(a)\,da. \tag{3}$$

The *SLA risk* is defined in this paper as the probability that the specified availability α will not be met, i.e., $1 - S(\tau,\alpha)$.

In addition, Figure 1 illustrates why the SLA risk may be considerable even if $E[\hat{A}(\tau)]$ is larger than the SLA requirement α. One of the most important considerations for the SLA risk assessment is the shape of $f_{\hat{A}(\tau)}(a)$. In this paper, we propose two mechanisms cable of modifying that shape, in order to reduce the SLA risk, without spending money in additional infrastructure.

3. Dependability in Cloud Computing Environments

3.1 Failure and Repair Processes in Cloud Computing Environments

Best practices suggest the design of robust cloud computing environments with enough redundancy in order to avoid service downtime. However, real clouds are not failure free. In this paper, we consider two failure models: 1) Single server downtime generated by hardware or software problems. 2) Multiple servers downtime, where more than one server or even an entire rack may experience downtime due to correlation between failure processes. This second case may be generated by a failure in the TOR (top of rack), or in the network, and it is very common as is shown in [FLP+10], [GJN11] and [GHHK10].

Based on the dynamics posed by the iteration of failure and repair processes in a cloud computing environment, the servers will be classified in three different groups: *active* servers, *spare* servers, and *on-repair* servers. The active server group contains the servers that are currently running virtual machines. The spare servers are working devices waiting for virtual machines that eventually may need them. Finally, *on-repair* servers are servers that were affected by any failure, and hence, they are not operational, neither able to allocate virtual machines.

The server state changes according to the following dynamics: An active server that is hit by a failure becomes immediately part of the on-repair group. After the server got repaired, it becomes part of the spare pool. Finally, a spare server becomes

an active server when a virtual machine request it and the VM-image is successfully restored on it, through the use of the fault tolerance technique of the virtualization platform.

Current cloud computing technologies make use of different fault tolerance techniques in order to keep high levels of availability. In this paper, we analyze three different commercially available fault tolerance techniques: vSphere Fault Tolerance (FT), vSphere High Availability (HA) and vSphere Data Recovery.

One of the most advanced fault tolerance techniques is vSphere Fault Tolerance [VMw09]. It uses identical/parallel VM-images running. Under this technique, a VM runs on two different physical servers. Every input is duplicated and received by both servers. In addition, both servers are constantly synchronized in order to keep the coherence of the virtual images. The idea is that in case of server failure, the virtualization platform is able to keep the VM running without generating any downtime from the user point of view. According to the vendors, this solution*"provides continuous availability, without any data loss or downtime, to any application"*. However, empirical applications may experience downtime of few seconds. Based on [VMw09], we assume that vSphere-FT has a restoration time of two seconds plus a negatively exponentially distributed (*n.e.d*) time with expected value of 3 seconds. In addition, this solution may not work if both images are hit at the same time due to the mentioned correlated failures. Finally, when a failure affects a virtual machine, and the backup image becomes the running image, the virtualization platform requests a server from the spare group in order to restore the image, and return to the original state when both images are running and synchronized.

vSphere High Availability [VMW07] is an alternative fault tolerance technique that does not use identical/parallel VM-images running. Instead, a smart storage of the VM image is performed, and it is constantly updated, keeping an active running image and a synchronized passive stored image. This technology requires high quality storage systems [TLM06]. The VM image is constantly synchronized and monitored. When a failure occurs, the cloud manager requests a spare server in order to restore the affected VM image. When the spare is obtained, the image is restored and the VM is up and operational again. *"It provides cost effective, automated restart within minutes for all applications in the event of hardware or operating system failures"*. Based on [VMW07], we assume that vSphere-HA has a restoration time of 1 minute plus a *n.e.d* time with expected value of 1 minute.

Finally, the last fault tolerance technique considered is vSphere Data Recovery [VM-09]. It uses the same logistics than vSphere High Availability, but here, the failure detection and restore process require some degree of human intervention, making the entire VM-recovery process longer. However, it *"provides simple, cost effective, agentless backup and recovery for virtual machines"*. Based on [VM-09], we assume that vSphere-Data-Recovery has a restoration time of 10 minutes plus a *n.e.d* time with expected value of 10 minutes.

3.2 Redundancy and Fault Tolerance

In this section, we will study the behavior of the expected interval availability $E[\hat{A}(\tau)]$ under different fault tolerance techniques, redundancy levels, and cloud sizes.

A cloud computing provider who offers IaaS receives requests from costumers with specific needs specified in terms of processor speed, memory and storage capacity. Based on this information, the minimum amount of servers needed to supply the customers demands may be calculated straightforward. This amount of servers will be called "*Basic Cloud size*".

The efficient application of a fault tolerance mechanism depends on the existence of spare servers available. Planning in advance the number of spare computers per active servers is a relevant task with huge influence on the availability offered to any virtual machine allocated in the cloud. At the same time, it represents an additional investment for the provider that may be unnecessary due to server overprovisioning. The redundancy percentage (%) will be defined as follows:

$$\text{redundancy } (\%) = \left(\frac{\text{Number of spare servers}}{\text{Basic cloud size}} \right) 100\% \tag{4}$$

We study the impact of redundancy on the availability offered to any virtual machine allocated in one of the two following scenarios: 1) Fixed fault tolerance technique, and variable *basic cloud size* and redundancy. 2) Fixed *basic cloud size* and variable fault tolerance technique and redundancy.

Using the first scenario, Figure 2 shows the effects of redundancy in the unavailability (1-availability) when VMware-HA is combined with three different *basic cloud sizes* (10, 100 and 1000). The results are obtained from a discrete-event simulation made in DEMOS [Bir03]. We assume that each rack contains 10 servers. The servers failure and repair times are *n.e.d* with expected time to failure $1/\lambda_s$ equal to one month and expected time to repair $1/\mu_s$ equal to 6 hours (the server steady state availability is 99%). In addition, an entire rack may be affected by a failure. In this case, failure and repair times are also *n.e.d* with expected time to failure $1/\lambda_r$ equal to three months and expected time to repair $1/\mu_r$ equal to 4 hours (the rack steady state availability is 99.8%).

Figure 2 shows unavailability values in the x axis, presented in a logarithmic scale in order to offer a better visualization of the effects of redundancy on the availability (1-unavailability). Form this figure, we observe three important behaviors: 1) The unavailability level offered to any VM allocated in the described scenario has the same minimum value (left limit in the figure), independently of the *basic cloud size*. This limit (best possible availability) is posed by the expected recovery time of VMware-HA. It is reached when there is server overprovisioning that makes very high the probability of having a spare server available (insignificant availability improvement to any redundancy increase). 2) There is an area, before reaching the unavailability limit, where any small increase in redundancy represents a considerable improvement in availability. This observation highlights the importance of planning the redundancy

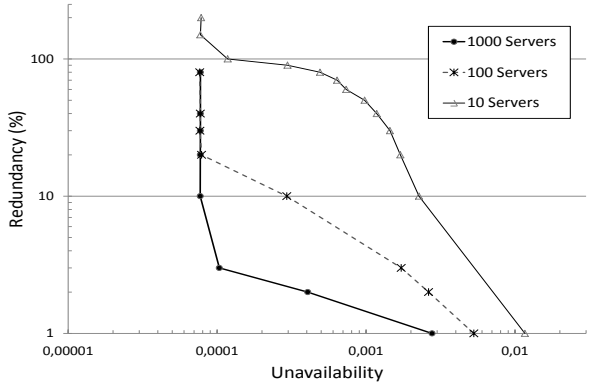


Figure 2. VM unavailability using VMware-HA with different *Basic-Cloud-Size* and redundancy

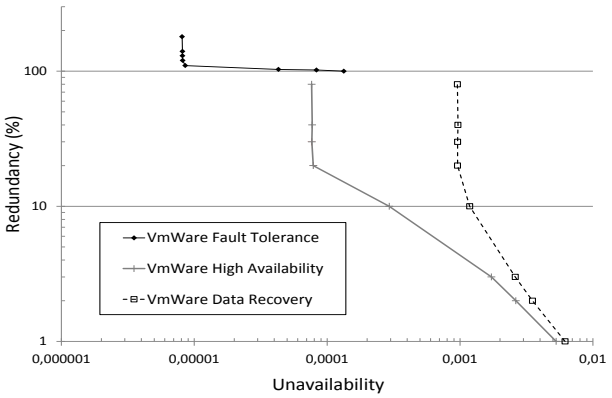


Figure 3. VM unavailability using a *Basic-Cloud-Size* of 100 servers with different fault tolerance techniques and redundancy

of a cloud. 3) The bigger the *basic cloud size* is, the smaller the redundancy percentage needed to reach the best possible availability.

In the second scenario, the fixed *basic cloud size* is 100 servers. The failure and repair processes have identical features to the one described in the first scenario. In Figure 3, different maximum limits (best possible availability) per each fault tolerance technique can be observed. As in the first scenario, these limits occur when the redundancy increase does not play a relevant role in the availability, i.e., the level of redundancy is large enough, making the probability of having available spares very high. Finally, the availability gap between technologies is of more than one order of magnitude.

Figure 3 may also be used to highlight one of the advantages obtained by making the virtual machines independent from the hardware. In this example, the original availability of a server is 99%. In addition, a rack (may hit 10 servers at the same time) has an availability of 99.8%, making the total server availability less than 99%.

However, by the use of cloud computing technologies, the virtual machines actually may have an availability around 99.9% with vSphere Data Recovery, more than 99.99% with vSphere High Availability, and more than 99.999% with vSphere Fault Tolerance.

From the provider point of view, the expenses of VMware-FT include: software licence, storage devices, and at least the double number of *basic cloud size* servers (more than 100% redundancy). For the case of VMware-HA, the software licence is less expensive, and the optimal redundancy level is smaller than 100%. Finally, the solution with vSphere Data Recovery does not require expensive storage solutions, and the software licence is much cheaper. The difference in licence prices can be appreciated in [VMW12].

With this information, the investment that a provider needs to make in order to provide cloud services to a set of virtual machines may be calculated. In cloud businesses, customers are willing to pay more to obtain a higher level of availability, and when the availability offered decreases, the customer expects to pay less. When the selling and the investment cost are estimated, they can be compared in order to guide cloud providers to sign profitable business, by selecting the technology that best fits the customer needs and the provider finances.

4. Policies to Reduce the SLA Risk

Figures 2 and 3 show the expected interval availability offered to a virtual machine under different scenarios. However, it can not be used to estimate $S(\tau, \alpha)$ accurately. In this case, the study of the entire distribution of the interval availability is needed. In this section, we will propose novel failure management policies that manipulate the distribution of $\hat{A}(\tau)$ in order to reduce the SLA risk.

The two most relevant factors that affect the availability offered to a virtual machine are: redundancy and fault tolerance technique. Regarding redundancy, few spare servers may be requested by many virtual machines at the same time. Therefore, a mechanism that helps to distribute the spares in an smart way has to be implemented. Regarding fault tolerance techniques, they are usually included in commercial licensed products [VMW12], limited by the total amount of memory and computational processes that they are able to handle, which depends directly on the price that the cloud provider is paying. A cloud provider may have different licences of different fault tolerance techniques. In this case, the provider needs to distribute in the best possible way the virtual machines requested by customers among the different licences available. This process will be defined as assignment of fault tolerance licences.

4.1 SLA-Budget

The concept of SLA-Budget was proposed by Mykkeltveit and Helvik in [MH09], in order to handle with failures in MPLS network connections, meeting availability constraints defined in an SLA. Our paper is an extension of that work. We use the SLA-Budget concept in cloud computing environments, finding that its implementation is easier and more flexible.

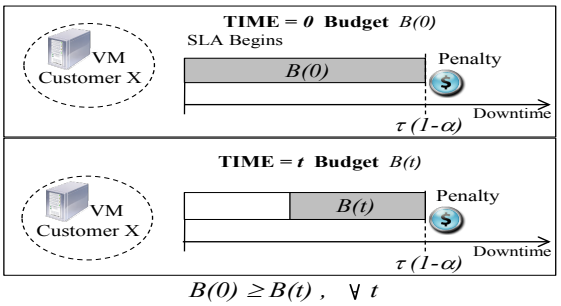


Figure 4. SLA Budget of a Virtual Machine.

Figure 4 illustrates the SLA-Budget concept. It is defined as *the current reaming downtime allowed for a virtual machine before the provider has to pay a penalty*. As explained in Section 2, when an SLA is negotiated, the availability promise α is stipulated for a defined time interval τ. Therefore, at the beginning of the contract the initial budget $B(0)$ is $\tau(1-\alpha)$.

At any moment t during the SLA period, the virtual machine may be affected by a failure that will reduce the remaining budget that a given virtual machine has before breaching the contract. Therefore, the current budget $B(t)$ is a dynamic value that may change at any time, depending on the stochastic features of the failure and repair processes of the cloud. In the best case, the virtual machine will never fail. Therefore, $B(0) \geq B(t)$.

4.2 Assignment of Spare Servers

When a virtual machine is affected by a failure, it requests a spare server to restore the affected image. At the same time t, other virtual machines may be waiting for spare servers. When the number of spares is not sufficient to restore all the waiting virtual machines, a policy must be implemented. A conventional policy that is currently used is to assign an spare server to the virtual machine that is waiting for longer time in the queue, using a classical FIFO approach. We will refer this policy as *Non - Adaptive*.

In this paper, we propose a novel approach that uses the SLA-budget in order to give priorities in the assignment of spare servers. In this way, at time t, virtual machines with smaller budget $B(t)$ will get the spare servers first. The highest priority will be given to the virtual machine with *min* $B(t)$ and the lowest to the virtual machine with *max* $B(t)$. With this policy, the cloud manager provides the available resources to the customers that need them more at time t. This policy is dynamic and it is always aware of the current needs of the system. Given that the budget of a virtual machine in the queue is constantly reduced with t, it is very unlikely that a virtual machine stay for a very long time in the queue. We will refer this policy as *Adaptive*.

A very interesting observation is that $E[\hat{A}(\tau)]$ is the same using both policies (Adaptive and Non-Adaptive). However, the difference is observed when the distribution of $\hat{A}(\tau)$ is considered. In order to have a better understanding of the effects of the

assignment policies, we simulate the operation of a cloud computing center, and we evaluate the probability distribution of the interval availability of a virtual machine on that cloud, using a discrete-event simulation in DEMOS [Bir03]. The simulation setting considers a cloud with 105 servers located in 11 racks, a *basic cloud size* of 100 servers (5% redundancy), and the SLA duration (τ) is one year. The fault tolerance technique used is vSphere High Availability, where the duration of the restoration time from the moment when the spare server is available is 1 minute, plus a *n.e.d* time with expected value of 1 minute. The servers failure and repair times are *n.e.d* with respective expected time to failure $1/\lambda_s$ equal to one month and expected time to repair $1/\mu_s$ equal to 6 hours. Rack's failure and repair times are also *n.e.d* with expected time to failure $1/\lambda_r$ equal to three months and expected time to repair $1/\mu_r$ equal to 4 hours.

Figure 5(a) shows the distribution of the interval availability for a virtual machine allocated in the scenario previously described, using the two different spare assignment policies mentioned (*Adaptive* and *Non-Adaptive*). The proposed adaptive technique reduces considerably the variance of the interval availability, making it *"more predictable"* and easier for the definition and fulfillment of the availability guarantee α. One of the most desired goals in SLA scenarios is the manipulation of the interval availability distribution in a way that benefits the could provider interests. This policy is a good alternative to make such manipulation.

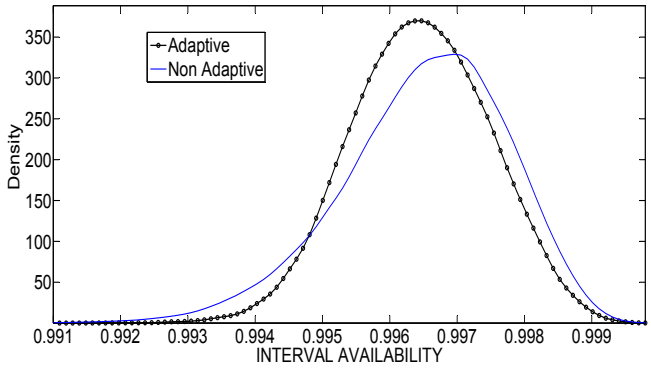
As was illustrated in Figure 1, when α is defined in an SLA, there is a probability to fulfill the promise ($S(\tau,\alpha)$) and a probability to fail the SLA (1-$S(\tau,\alpha)$). In Figure 5(b), we show explicitly the magnitude of 1-$S(\tau,\alpha)$ under both policies, which in fact is the area from 0 to α in Figure 5(a). Figure 5(b) shows that the SLA-risk may be ten times smaller when our assignment policy is used.

If the availability promised in an SLA is very high, the risk increases considerably, independently of the spare assignment policy used. In this case (e.g., SLA promised availability of 0.9955 in Figure 5(b)), the risk converges to similar values in both policies. However, this kind of scenarios are not commercially common given that cloud providers always try to operate with small risk values.

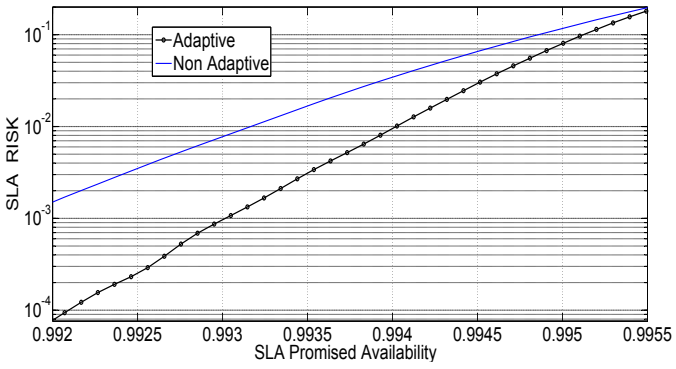
4.3 Assignment of Different Fault Tolerance Licences

Commercially available virtualization platforms such as vSphere Fault Tolerance, vSphere High Availability and vSphere Data Recovery are licensed products with a defined capacity. The price of the licence depends on the features of the manager and on the *maximum load* (capacity) that it is able to administrate (see [VMW12] for a price reference). For this reason, cloud providers may have different licences with different capacities. Let us assume one scenario where the provider has two fault tolerant licences H and S. H offers higher availability than S, e.g., H is vSphere High Availability and S is vSphere Data Recovery, with respective licence capacity C_H and C_S. In this case, the total capacity C that can be administrated by the cloud is $C = C_H + C_S$.

We assume a scenario where the provider needs to allocate a number of virtual machines with the same availability requirements and with a *basic cloud size* equal to



(a) $\hat{A}(\tau)$ of a Virtual Machine, using different policies for the assignment of spare servers

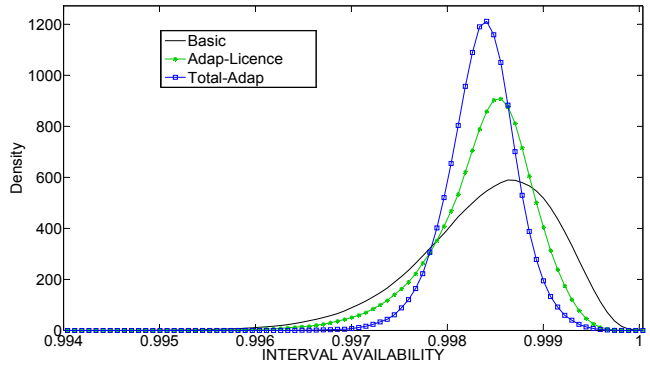


(b) SLA Risk

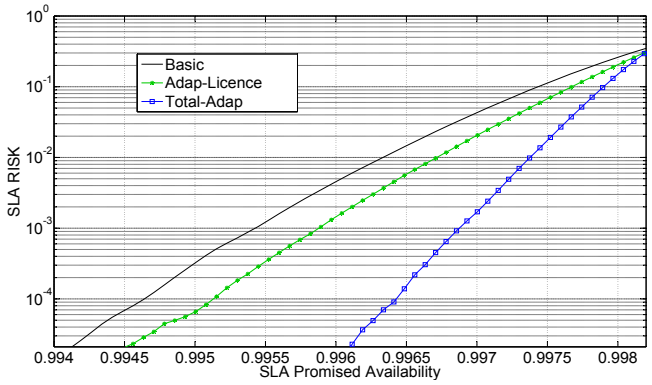
Figure 5. The impact of the smart assignment of spare servers.

C. A trivial solution in this situation could be to distribute the virtual machines among the different technologies by allocating each virtual machine half of the SLA period in solution H and the other half in solution S, an vice versa. This solution will be called "*HALF*". However, this basic solution is an intuitive and raw approach, and hence, better criteria have to be taken in order to design a better license assignment policy.

We propose an smarter mechanism to implement the licenses distribution using the SLA budget concept. In our mechanism, when a virtual machine s from S is affected by a failure, it looks for an spare to restore its backup VM-image. When an spare server is assigned, the restore process may consider the use of a different fault tolerance license, i.e., H. This may be done without violating the licence limit, if one of the virtual machines in H is downgrade to S. This process does not represent downtime for the implied virtual machines, since what is switched is the way to storage the backup virtual image, and not the running image itself. In the studied scenario, all the availability requirements are the same, e.g., $B(0)$ is equal on all virtual machines. In that way, when the virtual machine s is restored, it looks for the



(a) $\hat{A}(\tau)$ of a Virtual Machine, using different policies for the assignment of spare servers and fault tolerance licenses



(b) SLA Risk

Figure 6. The impact of the smart assignment of fault tolerance licences.

maximum current budget of all the virtual machines using license H $max\ \{B_H(t)\}$. If $max\ \{B_H(t)\} > B_s(t)$, the virtual machine with $max\ \{B_H(t)\}$ is downgraded to use license S, and the virtual machine under the restoration process s is restored using license H. We will refer this policy as *Adaptive* license assignment.

In summary, a cloud provider may have the ability to implement the following two different mechanism in order to reduce the SLA risk: 1) Spare Servers Assignment 2) Licences Assignment. Table 1 summarizes the possible settings that a cloud provider may implement, according to the policies used.

In order to show the effects of our policies in the distribution of $\hat{A}(\tau)$, we use the same simulation scenario described in Section 4.2 (*basic cloud size* 100 servers with 5% of redundancy and $\tau =$ one year). We will assume that the provider has a vSphere High Availability license with capacity equivalent to 50 servers and a vSphere Data Recovery license with capacity equivalent to 50 servers. In addition, we assume that the provider has to allocate 100 virtual machines with the same SLA requirements that demand the capacity of one server per virtual machine. From the simulation

		Spare Servers Assignment	
		FIFO$_1$ / FIXED$_2$	ADAPTIVE
Licences Assignment	HALF$_1$ / FIXED$_2$	*Basic*$_1$ / *Trivial*$_2$	*Adap-Queue*
	ADAPTIVE	*Adap-Licence*	*Total-Adap*

1. It applies for same VM class (Section 4.3)
2. It applies for different VM classes (Section 4.4)

Table 1. Policies used for VM restoration

results, we observe that the expected availability $E[\hat{A}(\tau)]$ is the same for the four policies combinations. However, the distribution of the interval availability differs considerably depending on the assignment policies selected.

Figure 6(a) shows the distributions of the interval availability for any virtual machine allocated in the scenario previously described, using the different assignment policies mentioned in table 1. We observe even a more pronounced reduction of the variance of the interval availability when both policies are adaptive (Total-Adap). As explained before, the use of the proposed policies makes $\hat{A}(\tau)$ *"more predictable"*, which means that our policy manipulates the probability distribution in a way that reduces the SLA risk.

In Figure 6(b), we show explicitly the magnitude of the SLA-risk under different SLA availability promises (this is the area from 0 to α in Figure 6(a)). In this case, the SLA-risk may be even one hundred times smaller, by using the proposed assignment mechanisms.

4.4 Virtual Machines with Different Classes

Previous sections propose smart assignment of spare servers and fault tolerance licences in order to reduce the SLA risk of virtual machines of the same class, i.e., the SLA availability requirements are the same. This section extends this concept by studying the application of these policies in scenarios where virtual machines belong to different classes.

In order to illustrate how to deal with this kind of situations, we assume the following scenario: A cloud provider has two different fault tolerance licenses. e.g., vSphere-HA and vSphere Data Recovery, and two different classes of customers (class 1 and 2). The availability requirement of class 1 customers is higher than the requirements of class 2 customers. A *trivial* solution to this scenario is to fix class 1 customers to the vSphere-HA solution and class 2 customers to the vSphere Data Recovery solution. In addition, the trivial solution will give always higher priority to class 1 customers in the assignment of spare servers.

We propose the use of a better alternative called *adaptive queue*. Here, the spare servers are assigned according to the SLA budget criteria. However, this solution still operates with fixed licenses assignment.

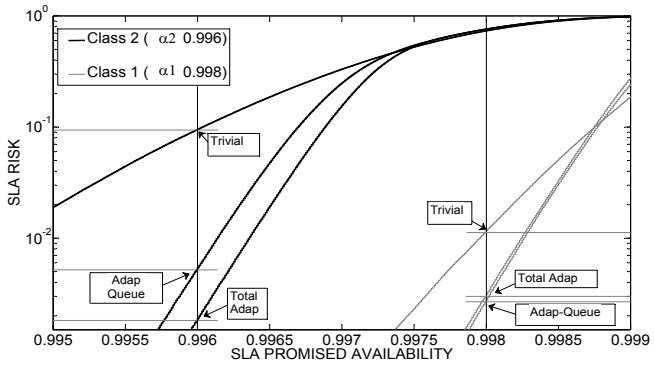


Figure 7. SLA Risk of virtual machines with different α (0.998 - 0.996)

Our last proposed solution is called *total adaptive* and it considers the assignment of spares and licenses according to the SLA budget. In this case, a virtual machine from class 2 may use the vSphere-HA license if the current conditions of the cloud demand it.

Table 1 also summarizes the possible policies combination that a cloud provider may implement under the current scenario.

Figure 7 shows the SLA risk after one year, for any virtual machine allocated in a scenario with $\alpha_1 = 0.998$ for class 1 customers that demand the capacity of 50 servers, and $\alpha_2 = 0.996$ for class 2 customers that demand the capacity of 50 servers. The cloud setting is the same that the one described in Section 4.3. The risk curves under the three policies (*trivial, adap-queue, total-adap*) show first of all, the huge improvement obtained by using *adap-queue* in both classes. On the other hand, when *total-adap* is implemented, the risk is considerably reduced for class 2 customers but the trade-off is the small increase in the risk of the class 1 customers. In all the simulations performed, the risk reduction in class 2 is much bigger than the increase in class 1. However, the cloud provider should evaluate if this policy is convenient, based on the service price and penalty scheme used on each of the specific signed SLAs.

We are interested in observing the effect of our adaptive policies in an scenario when there is a bigger gap between α_1 and α_2. We use the same simulation setting, but this time the values of α are 0.999 and 0.993 for class 1 and class 2 respectively. Figure 8 illustrates the most relevant information obtained from this simulation. First, the smart assignment of spares (*adap-queue*) is beneficial for both classes. This results is the same that the one obtained in the previous simulation, and it means that the smart assignment of spares becomes a very important policy independently of the level of availability promised and the respective gap between classes. For the class 2 customers, we observe that the risk obtained with the total adaptive policy is shorter that the one obtained using only adaptive queue. This is expected given that the total adaptive policy allows to class 2 customers the access to the best fault tolerant license, when it is very necessary. However, the most interesting observation is that this policy

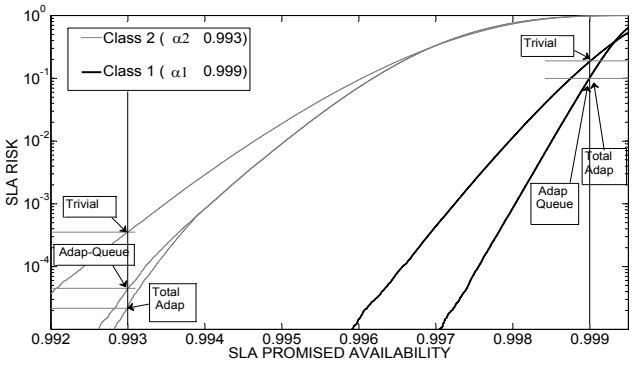


Figure 8. SLA Risk of virtual machines with different α (0.999 - 0.993)

does not represent any considerable increase in the risk of the class 1 customers. We may conclude that the total adaptive policy is self-regulated, which means that the increase in the risk experienced by high class customers becomes negligible with the increase of the gap between availability requirements of different classes.

5. Concluding Remarks

The main contribution of this paper is the proposal of two smart assignment policies that reduce considerably the SLA risk without investing in additional resources. This is also a first step in the design of SLA-aware cloud architectures.

The smart assignment of spare servers is a fundamental policy that can be easily implemented, and that is beneficial in all kind of situations, i.e., different class of customers, fault tolerance licences, cloud size, etc. The smart assignment of licenses is an effective way to use the resources available in a cloud. For same-class customers, this policy is always beneficial. For different class of customers, there is a trade-off between the risk reduction in the lower class customers and the risk increase in the higher class customers. However, there are two advantages in this situation: First, the risk-reduction is much bigger than the risk increase. Second, the policy is self-regulated and it does not have any strong effect when the difference in availability requirement between classes is big.

The results presented in the paper show that the risk reduction is valid only if the availability α stipulated in the SLA is at least smaller than $E[\hat{A}(\tau)]$. If this is not the case, the SLA-risk is considerably high, independently of the assignment policies used.

Finally, this paper also shows that redundancy is a very important factor that cloud providers should consider. The policies proposed in this paper, allow cloud providers to control the SLA risk without overprovisioning spare servers, opening an interesting future work, where savings in energy consumption can be analyzed, contributing to the growing needs in green technologies.

References

[Ama08] Amazon. Amazon EC2 service level agreement. [Online]. Available: http://aws.amazon.com/ec2-sla/. 2008.

[BDF+03] Paul Barham, Boris Dragovic, Keir Fraser, Steven Hand, Tim Harris, Alex Ho, Rolf Neugebauer, Ian Pratt, and Andrew Warfield. Xen and the art of virtualization. *SIGOPS Oper. Syst. Rev.*, 37(5):164–177, Oct. 2003.

[Bir03] Graham M. Birtwistle. *DEMOS - a system for Discrete Event Modelling on Simula*. University of Leeds, 2003.

[CLM+08] Brendan Cully, Geoffrey Lefebvre, Dutch Meyer, Mike Feeley, Norm Hutchinson, and Andrew Warfield. Remus: high availability via asynchronous virtual machine replication. In *Proceeding of the USENIX 5th Symposium on Networked Systems Design and Implementation (NSDI)*, pages 161–174, Berkeley, CA, USA, 2008.

[DHJ+07] Giuseppe DeCandia, Deniz Hastorun, Madan Jampani, Gunavardhan Kakulapati, Avinash Lakshman, Alex Pilchin, Swaminathan Sivasubramanian, Peter Vosshall, and Werner Vogels. Dynamo: amazon's highly available key-value store. *SIGOPS Oper. Syst. Rev.*, 41(6):205–220, Oct. 2007.

[FLP+10] Daniel Ford, François Labelle, Florentina I. Popovici, Murray Stokely, Van-Anh Truong, Luiz Barroso, Carrie Grimes, and Sean Quinlan. Availability in globally distributed storage systems. In *Proceeding of the USENIX 9th conference on Operating systems design and implementation (OSDI)*, pages 1–7, Berkeley, CA, USA, 2010.

[GH09] Andres J. Gonzalez and Bjarne E. Helvik. Guaranteeing service availability in SLAs; a study of the risk associated with contract period and failure process. *Proceeding of the IEEE Latin-American Conference on Communications (LATINCOM)*, pages 1 –6, Sep. 2009.

[GHHK10] Andres J. Gonzalez, Bjarne E. Helvik, Jon K. Hellan, and Pirkko Kuusela. Analysis of dependencies between failures in the UNINETT IP backbone network. *Proceeding of the IEEE International Symposium on Pacific Rim Dependable Computing (PRDC)*, 2010.

[GJN11] Phillipa Gill, Navendu Jain, and Nachiappan Nagappan. Understanding network failures in data centers: measurement, analysis, and implications. In *Proceeding of the ACM SIGCOMM conference*, pages 350–361, New York, NY, USA, 2011.

[Gol74] Robert P. Goldberg. Survey of virtual machine research. *IEEE Computer Magazine*, 1974.

[GT88] Ambuj Goyal and Asser Tantawi. A measure of guaranteed availability and its numerical evaluation. *IEEE Transactions on Computers*, Volume 37, Issue 1:25 – 32, 1988.

[LJL+09] Haikun Liu, Hai Jin, Xiaofei Liao, Liting Hu, and Chen Yu. Live migration of virtual machine based on full system trace and replay. In *Proceeding of the ACM 18th International Symposium on high performance distributed computing (HPDC)*, pages 101–110, New York, NY, USA, 2009.

[MG11] Peter Mell and Timothy Grance. *The NIST Definition of Cloud Computing*. National Institute of Science and Technology NIST, Jul. 2011.

[MH09] Anders Mykkeltveit and Bjarne E. Helvik. Adaptive management of connections to meet availability guarantees in SLAs. In *Proceeding of the IFIP/IEEE International Symposium on Integrated Network Management, IM. Mini-Conference*, Jun. 2009.

[MUKX06] Mark F. Mergen, Volkmar Uhlig, Orran Krieger, and Jimi Xenidis. Virtualization for high-performance computing. *SIGOPS Oper. Syst. Rev.*, 40(2):8–11, Apr. 2006.

[PG73] Gerald J. Popek and Robert P. Goldberg. Formal requirements for virtualizable third generation architectures. *Proceeding of the ACM Fourth Symposium on Operating System Principles, (SOSP). Yorktown Heights, New York*, Oct. 1973.

[SFK+09] Atul Singh, Pedro Fonseca, Petr Kuznetsov, Rodrigo Rodrigues, and Petros Maniatis. Zeno: eventually consistent byzantine-fault tolerance. In *Proceeding of the USENIX 6th Symposium on Networked systems design and implementation (NSDI)*, pages 169–184, Berkeley, CA, USA, 2009.

[SMLF09] Borja Sotomayor, Ruben S. Montero, Ignacio M. Llorente, and Ian Foster. Virtual infrastructure management in private and hybrid clouds. *IEEE Internet Computing*, 13(5):14 –22, Sept.-Oct. 2009.

[TLM06] Jon Tate, Fabiano Lucchese, and Richard Moore. *Introduction to Storage Area Networks*. Vervante, 2006.

[VM-09] VM-Ware. VMware DataRecovery: Complete Data Protection For Virtual Machines. 2009.

[VMW07] VMWare. VMware High Availability: Concepts, Implementation and Best Practices. 2007.

[VMw09] VMware. Protecting Mission-Critical Workloads with VMware Fault Tolerance. 2009. White Paper.

[VMW12] VMWare. VMware Sphere Pricing. 2012.

Bibliography

[ALRL04] Algirdas Avizienis, Jean-Claude Laprie, Brian Randell, and Carl Landwehr. Basic concepts and taxonomy of dependable and secure computing. *IEEE Transactions on Dependable and Secure Computing*, 1(1):11–33, Mar. 2004.

[Ama08] Amazon. Amazon EC2 service level agreement. [Online]. Available: http://aws.amazon.com/ec2-sla/. 2008.

[AMZ87] Khalil Sheikh Anwar, Ahmad Munir, and Ali Zulfiqar. Some remarks on the hazard functions of the inverted distributions. *IEEE Transactions on Reliability*, 19(4):255 – 261, 1987.

[BAK04] Claudia Betous-Almeida and Karama Kanoun. Construction and stepwise refinement of dependability models. *Elsevier Science Journal of Performance Evaluation*, 56:277 – 306, 2004.

[BCNN05] Jerry Banks, John S. Carson, Barry L. Nelson, and David M. Nicol. *Discrete-Event System Simulation (4rd Edition)*. Prentice Hall, 4 edition, 2005.

[BDF+03] Paul Barham, Boris Dragovic, Keir Fraser, Steven Hand, Tim Harris, Alex Ho, Rolf Neugebauer, Ian Pratt, and Andrew Warfield. Xen and the art of virtualization. *SIGOPS Oper. Syst. Rev.*, 37(5):164–177, Oct. 2003.

[Bir03] Graham M. Birtwistle. *DEMOS - a system for Discrete Event Modelling on Simula*. University of Leeds, 2003.

[BL97] John R. Birge and Francois Louveaux. *Introduction to Stochastic Programming*. Springer Series in Operations Research and Financial Engineering. Springer, 1997.

[CDG+04] Pablo. Cirrone, Stefania Donadio, Susanna Guatelli, Alfonso Mantero, Barbara Mascialino, S. Parlati, Maria G. Pia, Andreas Pfeiffer, Alberto Ribon, and Paolo Viarengo. A goodness-of-fit statistical toolkit. *IEEE Transactions on Nuclear Science*, 51(5):2056 – 2063, Oct. 2004.

[CH01] Yinong Chen and Zhongshi He. Dependability modelling of homogeneous and heterogeneous distributed systems. In *Proceeding of the IEEE 5th International Symposium on Autonomous Decentralized Systems (ISADS)*, pages 176 –183, 2001.

[CLM+08] Brendan Cully, Geoffrey Lefebvre, Dutch Meyer, Mike Feeley, Norm Hutchinson, and Andrew Warfield. Remus: high availability via asynchronous virtual machine replication. In *Proceeding of the USENIX 5th Symposium on Networked Systems Design and Implementation (NSDI)*, pages 161–174, Berkeley, CA, USA, 2008.

[CMH+07] Piotr Cholda, Anders Mykkeltveit, Bjarne E. Helvik, Otto J. Wittner, and Andrezej Jajszczyk. A survey of resilience differentiation frameworks in communication networks. *IEEE Communications Surveys & Tutorials*, 9(4):32–55, Fourth Quarter 2007.

[Cox67] David R. Cox. *Renewal Throry*. Methuen, 1967.

[CSK02] Weidong Cui, Ion Stoica, and Randy H. Katz. Backup path allocation based on a correlated link failure probability model in overlay networks. In *Proceeding of the IEEE International Conference on Network Protocols (ICNP)*, 2002.

[CSKM07] Baek Young Choi, Sejun Song, George Koffler, and Deep Medhi. Outage analysis of a university campus network. *Proceedings of 16th IEEE International Conference on Computer Communications and Networks (ICCCN)*, pages 675 – 680, 13-16 Aug. 2007.

[Das00] Luiz A. Dasilva. Pricing for QoS-enabled networks: A survey. *IEEE Communications Surveys & Tutorials*, 3(2):2–8, 2000.

[DHJ+07] Giuseppe DeCandia, Deniz Hastorun, Madan Jampani, Gunavardhan Kakulapati, Avinash Lakshman, Alex Pilchin, Swaminathan Sivasubramanian, Peter Vosshall, and Werner Vogels. Dynamo: amazon's highly available key-value store. *SIGOPS Oper. Syst. Rev.*, 41(6):205–220, Oct. 2007.

[Dij59] Edsger W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1(1):269–271, Dec. 1959.

[EA02] Osborne Eric and Simhar Ajay. *Traffic Engineering with MPLS*. Cisco Press, Jul. 17, 2002.

[FDL00] Matthias Falkner, Michael Devetsikiotis, and Ioannis Lambadaris. An overview of pricing concepts for broadband IP networks. *IEEE Communications Surveys & Tutorials*, 3(2):2–13, 2000.

[FLP+10] Daniel Ford, François Labelle, Florentina I. Popovici, Murray Stokely, Van-Anh Truong, Luiz Barroso, Carrie Grimes, and Sean Quinlan. Availability in globally distributed storage systems. In *Proceeding of the USENIX 9th conference on Operating systems design and implementation (OSDI)*, pages 1–7, Berkeley, CA, USA, 2010.

[FP01] Sally Floyd and Vern Paxson. Difficulties in simulating the internet. *IEEE/ACM Transactions on Networking*, 9(4):392–403, 2001.

[FT06] Andrea Fumagalli and Marco Tacca. Differentiated reliability (DiR) in wavelength division multiplexing rings. *IEEE/ACM Transactions on Networking*, 14(1):159–168, Feb. 2006.

[FTUF02] Andrea Fumagalli, Marco Tacca, Ferenc Unghvary, and Andras Farago. Shared path protection with differentiated reliability. In *Proceeding of the IEEE International Conference on Communications (ICC)*, volume 4, pages 2157–2161, 28 Apr.–2 May. 2002.

[FY94] Kenich Funaki and Kazuho Yoshimoto. Distribution of total uptime during a given time interval. *IEEE Transactions on Reliability*, 43(3):489 –492, Sep. 1994.

[FZ02] Maxim S. Finkelstein and Vladimir I. Zarudnij. Laplace-transforms and fast-repair approximations for multiple availability and its generalizations. *IEEE Transactions on Reliability*, 51(2):168 –176, Jun. 2002.

[GH09] Andres J. Gonzalez and Bjarne E. Helvik. Guaranteeing service availability in SLAs; a study of the risk associated with contract period and failure process. *Proceeding of the IEEE Latin-American Conference on Communications (LATINCOM)*, pages 1 –6, Sep. 2009.

[GH10] Andres J. Gonzalez and Bjarne E. Helvik. Dynamic Sharing Mechanism for Guaranteed Availability in MPLS Based Networks. *Proceeding of the IEEE International Communications Quality and Reliability (CQR)*, Jun. 2010.

[GH11a] Andres J. Gonzalez and Bjarne E. Helvik. Analysis of failures characteristics in the UNINETT IP backbone network. *Proceeding of the IEEE 7th International Symposium on Frontiers in Networking with Applications (FINA)*, Mar. 2011.

[GH11b] Andres J. Gonzalez and Bjarne E. Helvik. Guaranteeing Service Availability in SLAs on Networks with Non Independent Failures. *IEEE-IFIP International Workshop on Design of Reliable Communication Networks (DRCN)*, Oct. 2011.

[GH12a] Andres J. Gonzalez and Bjarne E. Helvik. A Study of the Interval Availability and its Impact on SLAs Risk. *Proceeding of the Springer International Conference on Computer Science, Engineering and Applications (ICCSEA)*, 2012.

[GH12b] Andres J. Gonzalez and Bjarne E. Helvik. Guaranteeing SLA Availability in Telecommunications Networks. *IEEE International Telecommunications Network Strategy and Planning Symposium (NETWORKS)*, Oct. 2012.

[GHHK10] Andres J. Gonzalez, Bjarne E. Helvik, Jon K. Hellan, and Pirkko Kuusela. Analysis of dependencies between failures in the UNINETT IP backbone network. *Proceeding of the IEEE International Symposium on Pacific Rim Dependable Computing (PRDC)*, 2010.

[GJN11] Phillipa Gill, Navendu Jain, and Nachiappan Nagappan. Understanding network failures in data centers: measurement, analysis, and implications. In *Proceeding of the ACM SIGCOMM conference*, pages 350–361, New York, NY, USA, 2011.

[Gol74] Robert P. Goldberg. Survey of virtual machine research. *IEEE Computer Magazine*, 1974.

[Gro04] Wayne D. Grover. *Mesh-Based Survivable Networks. Options and Strategies for Optical, MPLS, SONET, and ATM Networks*. Prentice Hall PTR, Upper Saddle River, NJ, 2004.

[GT88] Ambuj Goyal and Asser Tantawi. A measure of guaranteed availability and its numerical evaluation. *IEEE Transactions on Computers*, Volume 37, Issue 1:25 – 32, 1988.

[HH99] S. Huzurbazar and Aparna V. Huzurbazar. Survival and hazard functions for progressive diseases using saddlepoint approximations. *International Biometric Society Journal of Biometrics*, 55(1):pp. 198–203, 1999.

[HLS07] Changcheng Huang, Minzhe Li, and A. Srinivasan. A scalable path protection mechanism for guaranteed network reliability under multiple failures. *IEEE Transactions on Reliability*, 56(2):254–267, Jun. 2007.

[HTH08] Pin-Han Ho, J. Tapolcai, and A. Haque. Spare capacity reprovisioning for shared backup path protection in dynamic generalized multi-protocol label switched networks. *IEEE Transactions on Reliability*, 57(4):551–563, Dec. 2008.

[InCM+02] Gianluca Iannaccone, Chen nee Chuah, Richard Mortier, Supratik Bhattacharyya, and Christophe Diot. Analysis of link failures in an IP backbone. *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurment (IMW)*, pages 237–242, 2002.

[ITU94] ITU-T. Terms and Definitions related to Quality of Service and Network Performance Including Dependability. ITU-T Rec. E.800, Aug. 1994.

[ITU00] ITU-T. B-ISDN Semi-permanent Connection Availability. ITU-T Rec. I.357, Nov. 2000.

[ITU01] ITU-T. Data Networks and Open Systems Communications, Open Systems Interconnection Service definitions. ITU-T Rec. X.213, Oct. 2001.

[ITU02a] ITU-T. Framework of a Service Level Agreement. ITU-T Rec. E.860, Jun. 2002.

[ITU02b] ITU-T. Internet Protocol Data Communication Service — IP Packet Transfer and Availability Performance Parameters. ITU-T Rec. Y.1540, Dec. 2002.

[KKM03] Mohamed Kaniche, Karama Kanoun, and Magnos Martinello. A user-perceived availability evaluation of a web based travel agency. In *Proceeding of the IEEE International Conference on Dependable Systems and Networks (DSN)*, pages 709 – 718, Jun. 2003.

[KL00] Murali Kodialam and T. Lakshman. Dynamic routing of bandwidth guaranteed tunnels with restoration. In *Proceeding of the IEEE Nineteenth Annual Joint Conference of the Computer and Communications Societies (INFOCOM)*, volume 2, pages 902–911, 26–30 Mar. 2000.

[KN10] Pirkko Kuusela and Illka Norros. On/Off process modeling of IP network failures. In *Proceeding of the IEEE/IFIP the 40th Annual International Conference on Dependable Systems and Networks (DSN)*, 2010.

[KNR09] Pirkko Kuusela, Ilkka Norros, and Pertti Raatikainen. Report on modelling the reliability of an ip-network and strategies for improving the reliability. Technical report, A report of the IPLU-II project, Jun. 2009. Available at http://iplu.vtt.fi.

[KV06] Bernhard Korte and Jens Vygen. *Combinatorial Optimization: Theory and Algorithms*. Springer, Germany, 3rd edition, 2006.

[Law07] Averill M. Law. *Simulation Modeling and Analysis*. McGraw-Hill Education, New York, fourth edition, Apr. 2007.

[LBCG02] Li Li, Milind M. Buddhikot, Chandra Chekuri, and Jinhong Katherine Guo. Routing bandwidth guaranteed paths with local restoration in label switched networks. In *Proceeding of the IEEE 10th International Conference on Network Protocols (ICNP)*, pages 110–121, Washington, DC, USA, 2002.

[LJL+09] Haikun Liu, Hai Jin, Xiaofei Liao, Liting Hu, and Chen Yu. Live migration of virtual machine based on full system trace and replay. In *Proceeding of the ACM 18th International Symposium on high performance distributed computing (HPDC)*, pages 101–110, New York, NY, USA, 2009.

[LLY09] Hongbin Luo, Lemin Li, and Hongfang Yu. Routing connections with differentiated reliability requirements in WDM mesh networks. *IEEE/ACM Transactions on Networking*, 17(1):253–266, Feb. 2009.

[LML10] Hyang-Won Lee, E. Modiano, and Kayi Lee. Diverse routing in networks with probabilistic failures. *IEEE/ACM Transactions on Networking*, 18(6):1895 –1907, Dec. 2010.

[Lom66] Z. A. Lomnicki. A note on the weibull renewal process. *Biometrika*, 53(3/4):pp. 375–381, 1966.

[MdSeSG89] Richard Robert Muntz, E. de Souza e Silva, and Ambuj Goyal. Bounding availability of repairable computer systems. *SIGMETRICS Perform. Eval. Rev.*, 17(1):29–38, Apr. 1989.

[Med10] Aref Meddeb. Internet QoS: Pieces of the puzzle. *IEEE Communications Magazine*, 48(1):86 –94, january 2010.

[MG11] Peter Mell and Timothy Grance. *The NIST Definition of Cloud Computing*. National Institute of Science and Technology NIST, Jul. 2011.

[MH07] Anders Mykkeltveit and Bjarne E. Helvik. Provision of connection-specific availability guarantees in communication networks. In *Proceedings of the IEEE 6th International Workshop on Design of Reliable Communication Networks (DRCN)*, Oct, 2007.

[MH08a] Anders Mykkeltveit and Bjarne E. Helvik. Comparison of schemes for provision of differentiated availability-guaranteed services using dedicated protection. In *Proceeding of the IEEE Seventh International Conference on Networking (ICN)*, Apr. 2008.

[MH08b] Anders Mykkeltveit and Bjarne E. Helvik. On provision of availability guarantees using shared protection. In *Proceeding of the IEEE/IFIP 12th Conference on Optical Network Design and Modelling (ONDM)*, Mar. 2008.

[MH09] Anders Mykkeltveit and Bjarne E. Helvik. Adaptive management of connections to meet availability guarantees in SLAs. In *Proceeding of the IFIP/IEEE International Symposium on Integrated Network Management, IM. Mini-Conference*, Jun. 2009.

[MIB+08] Athina Markopoulou, Gianluca Iannaccone, Supratik Bhattacharyya, Chen-Nee Chuah, Yashar Ganjali, and Christophe Diot. Characterization of Failures in an Operational IP Backbone Network. *IEEE/ACM Transactions on Networking*, 16(4):749–762, Aug. 2008.

[MMW99] Wai-Kei Mak, David P. Morton, and Kevin R. Wood. Monte Carlo bounding techniques for determining solution quality in stochastic programs. *ELSEVIER Operations Research Letters*, 24:47–56, 1999.

[MN11] Loretta Mastroeni and Maurizio Naldi. Violation of service availability targets in service level agreements. In *Proceeding of the Federated Conference on Computer Science and Information Systems (FedCSIS)*, pages 537 –540, Sep. 2011.

[MO67] Albert W. Marshall and Ingram Olkin. A multivariate exponential distribution. *Journal of the American Statistical Association*, 62(317):pp. 30–44, 1967.

[MPP+06] Barbara Mascialino, Andreas Pfeiffer, Maria Grazia Pia, Alberto Ribon, and Paolo Viarengo. New developments of the goodness-of-fit statistical toolkit. *IEEE Transactions on Nuclear Science*, 53(6):3834 –3841, Dec. 2006.

[MQWS06] Darli A. Mello, G. Quiterio, Helio Waldman, and Dominic A. Schupke. Specification of SLA survivability requirements for optical path protected connections. In *Proceeding of the National Fiber Optic Engineers Conference on Optical Fiber Communication Conference (OFC)*, page 3pp., 5–10 Mar. 2006.

[MSW05] Darli A. Mello, Dominic A. Schupke, and Helio Waldman. A matrix-based analytical approach to connection unavailability estimation in shared backup path protection. *IEEE Communications Letters*, 9(9):844–846, Sep. 2005.

[MUKX06] Mark F. Mergen, Volkmar Uhlig, Orran Krieger, and Jimi Xenidis. Virtualization for high-performance computing. *SIGOPS Oper. Syst. Rev.*, 40(2):8–11, Apr. 2006.

[Mut68] Eginhard J. Muth. A method for predicting system downtime. *IEEE Transactions on Reliability*, R-17(2):97 –102, Jun. 1968.

[MVM02] Steven M. Matz, Lawrence G. Votta, and Mohammad Malkawi. Analysis of failure and recovery rates in a wireless telecommunications system. *Proceedings of the IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, pages 687 – 693, 2002.

[ND08] Maurizio Naldi and Giuseppe D'Acquisto. A normal copula model for the economic risk analysis of correlated failures in communications networks. *Journal of Universal Computer Science*, 14(5):786–799, 2008.

[NIS11] NIST/SEMATECH. e-handbook of statistical methods, [online]. available at: http://www.itl.nist.gov/div898/handbook/. 2011.

[OPTW07] S. Orlowski, M. Pióro, A. Tomaszewski, and R. Wessäly. SNDlib 1.0–Survivable Network Design Library. In *Proceeding of the 3rd International Network Optimization Conference (INOC), Spa, Belgium*, Apr. 2007. `http://sndlib.zib.de`.

[PER] PERL. The Perl programming language. http://www.perl.org/.

[PG73] Gerald J. Popek and Robert P. Goldberg. Formal requirements for virtualizable third generation architectures. *Proceeding of the ACM Fourth Symposium on Operating System Principles, (SOSP). Yorktown Heights, New York*, Oct. 1973.

[PM04] Michal Pióro and Deepankar Medhi. *Routing, Flow, and Capacity Design in Communication and Computer Networks*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2004.

[RBS+01] Ramu Ramamurthy, Zbigniew Bogdanowicz, Shahrokh Samieian, Debanjan Saha, Bala Rajagopalan, Sudipta Sengupta, Sid Chaudhuri, and Krishna Bala. Capacity performance of dynamic provisioning in optical networks. *Journal of lightwave technology*, 19(1):40–48, Jan. 2001.

[RS95] Gerard0 Rubino and Bruno Sericola. Interval availability analysis using denumerable markov processes: application to multiprocessor subject to breakdowns and repair. *IEEE Transactions on Computers*, Volume 44, Issue 2:286 – 291, Feb. 1995.

[San08] William H. Sanders. *Mobius Manual. Version 2.2.1. [online]. Available at: https://www.mobius.illinois.edu/manual/MobiusManual.pdf*, Nov. 20, 2008.

[SEG86] De Souza E Silva and H.R. E. Gail. Calculating cumulative operational time distributions of repairable computer systems. *IEEE Transactions on Computers*, Volume: C-35, Issue: 4:322–332, Apr. 1986.

[SFK+09] Atul Singh, Pedro Fonseca, Petr Kuznetsov, Rodrigo Rodrigues, and Petros Maniatis. Zeno: eventually consistent byzantine-fault tolerance. In *Proceeding of the USENIX 6th Symposium on Networked systems design and implementation (NSDI)*, pages 169–184, Berkeley, CA, USA, 2009.

[Sho76] Martin L. Shooman. Structural models for software reliability prediction. In *Proceeding of the ACM/IEEE 2nd international conference on Software engineering (ICSE)*, pages 268–280, Los Alamitos, CA, USA, 1976.

[SK04] Nozer D. Singpurwalla and Chung-Wai Kong. Specifying interdependence in networked systems. *IEEE Transactions on Reliability*, 53(3):401–405, 2004.

[SL99] Toshiyuki Shimokawa and Min Liao. Goodness-of-fit tests for type-i extreme-value and 2-parameter weibull distributions. *IEEE Transactions on Reliability*, 48(1):79 –86, Mar. 1999.

[SMLF09] Borja Sotomayor, Ruben S. Montero, Ignacio M. Llorente, and Ian Foster. Virtual infrastructure management in private and hybrid clouds. *IEEE Internet Computing*, 13(5):14 –22, Sept.-Oct. 2009.

[SND] SNDlib. Library of test instances for Survivable fixed telecommunication Network Design. [Online]. Available: http://sndlib.zib.de.

[SO93] William H. Sanders and II Obal, Douglas. Dependability evaluation using UltraSAN. In *Proceeding of the Twenty-Third International Symposium on Fault-Tolerant Computing FTCS-23. Digest of Papers*, pages 674–679, 22–24 Jun. 1993.

[Sol03] Ixia Solutions. Multi-Protocol Label Switching (MPLS): Conformance and Performance Testing. *Whitepaper IXIA*, Dec. 2003.

[Sti11] Stine Barstad. Mobiltrøbbel for Telenor-kunder [online]. Available at: http://www.aftenposten.no/nyheter/iriks/article4144897.ece, 2011.

[SW07] Andrew P. Snow and Gary R. Weckman. What Are the Chances an Availability SLA will be Violated? In *Proceeding of the IEEE Sixth International Conference on Networking (ICN)*, page 35, Apr. 2007.

[SWG10] Andrew P. Snow, Gary R. Weckman, and Vivek Gupta. Meeting SLA Availability Guarantees through Engineering Margin. In *Proceeding of the IEEE Ninth International Conference on Networks (ICN)*, pages 331 –336, Apr. 2010.

[Tak57] Lajos Takacs. On certain sojourn time problems in the theory of stochastic processes. *Acta Mathematica Hungarica*, 8:169–191, 1957.

[Tan03] Andrew S. Tanenbaum. *Computer Networks*. Prentice-Hall, fourth edition, 2003.

[TLM06] Jon Tate, Fabiano Lucchese, and Richard Moore. *Introduction to Storage Area Networks*. Vervante, 2006.

[UNI12a] UNINETT. The Norwegian Research Network. Downtime Statistics. [online]. Available at: http://drift.uninett.no/downs/. 2012.

[UNI12b] UNINETT. The Norwegian Research Network. Network Topology. [online]. Available at: http://drift.uninett.no/stat-q/load-map/uninett,,traffic,peak. 2012.

[VAK+03] Bram Verweij, Shabbir Ahmed, Anton J. Kleywegt, George Nemhauser, and Alexander Shapiro. The Sample Average Approximation Method Applied to Stochastic Routing Problems: A Computational Study. *Springer Computational Optimization and Applications*, 24(2):289–333–333, Feb. 2003.

[VM-09] VM-Ware. VMware DataRecovery: Complete Data Protection For Virtual Machines. 2009.

[VMW07] VMWare. VMware High Availability: Concepts, Implementation and Best Practices. 2007.

[VMw09] VMware. Protecting Mission-Critical Workloads with VMware Fault Tolerance. 2009. White Paper.

[VMW12] VMWare. VMware Sphere Pricing. 2012.

[VPD04] Jean-Philippe Vasseur, Mario Pickavet, and Piet Demeester. *Network Recovery. Protection and Restoration of Optical, SONET-SDH, IP, and MPLS*. Morgan Kaufmann Publishers, San Francisco, CA, 2004.

[Win95] Rainer Winkelmann. Duration dependence and dispersion in count-data models. *Journal of Business & Economic Statistics*, 13(4):pp. 467–474, 1995.

[WM08] Helio Waldman and Darli A Mello. SLA-aware survivability. In *Proceeding of the 10th Anniversary International Conference on Transparent Optical Networks (ICTON)*, volume 3, pages 46 –49, Jun. 2008.

[WM09] Helio Waldman and Darli A. Mello. On the risk of non-compliance with some plausible SLA requirements. In *Proceeding of the 11th International Conference on Transparent Optical Networks (ICTON)*, pages 1 –4, Jul. 2009.

[WVMD84] Brian W. Woodruff, Philip J. Viviano, Albert H. Moore, and Edward J. Dunne. Modified goodness-of-fit tests for gamma distributions with unknown location and scale parameters. *IEEE Transactions on Reliability*, R-33(3):241 –245, Aug. 1984.

[WZY04] WeiWei, Qingji Zeng, and YunWang. Multi-layer differentiated integrated survivability for optical internet. *Photonic Network Communications*, 8(3):267–284, 2004.

[XTMM11] Ming Xia, Massimo Tornatore, Charles U. Martel, and Biswanath Mukherjee. Risk-Aware Provisioning for Optical WDM Mesh Networks. *IEEE/ACM Transactions on Networking*, 19(3):921 –931, Jun. 2011.

[ZZZ+07] Jing Zhang, Keyao Zhu, Hui Zang, N. S. Matloff, and B. Mukherjee. Availability-Aware Provisioning Strategies for Differentiated Protection Services in Wavelength-Convertible WDM Mesh Networks. *IEEE/ACM Transactions on Networking*, 15(5):1177–1190, Oct. 2007.