A CONTINUOUS OVERLAY PATH PROBING ALGORITHM FOR OVERLAY NETWORKS

By

MARYAM FEILY

Thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

UNIVERSITI SAINS MALAYSIA

JULY 2013

بسم الله الرحمن الرحيم

ACKNOWLEDGEMENTS

I would like to take this opportunity to convey my sincere thanks and deepest gratitude to my awesome supervisor **Dr. Andrew Meulenberg** for all the help and invaluable guidance provided to me during the preparation of this thesis. It has been truly a gift of grace to have the privilege to work under his direction.

I graciously acknowledge the generous and prestigious support from the **UNIVERSITI SAINS MALAYSIA (USM)** through the **Postgraduate Research Grant Scheme (PRGS)** and the **USM Fellowship** awarded to me. My sincere gratitude goes to **Professor Tan Sri Datuk Dzulkifli Abdul Razak**, **Professor Dato' Omar Osman**, **Professor Muhamad Jantan**, **Mr. Abd Hadi Ahmad**, **Professor Ahmad Shukri Mustapa Kamal**, **Professor Asma Ismail**, **Professor Roshada Hashim**, **Professor Othman Sulaiman**, **Professor Abdul Rahman Othman**, **Mr. Mohamad Azahar Mahmood**, **Mr. Mohd Jalaludin Bin Azizan**, and **Mr. Aizat Hisham Ahmad** for their enormous support during my journey at USM. Besides, I would like to extend my warm appreciations to each and everyone in the **Institute of Postgraduate Studies (IPS)** for their endless help.

Moreover, I would like to express my deepest appreciation and warmest gratefulness to my extraordinary partner **Mr. Alireza Shahrestani** for the enthusiastic cooperation and kind support in all stages of this thesis.

Best of best, my most sincere thanks and warmest gratitude belong to my dear parents for their encouragement and unending moral support. This would not happen without their sacrifices and patience.

The favor, beyond all, is entirely Allah's, to whom my never-ending thanks and praise are humbly due.

Thank You!

Maryam Feily

Penang Island, Malaysia, July 2013

DEDICATION

I am delighted to graciously dedicate this thesis to my lovely parents, who cultivated a success ground for me with passion to develop strong and healthy roots for a flourishing future. All my achievements are truly the smiling blossoms of their endless care, with the nice fragrance of their unconditional love towards me in every season of my life. I salute them and I humbly kiss their supporting hands.

Dear Mom & Dear Dad

Thank You!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACBP	Approximate Cluster-Based Policy
API	Application Programming Interface
ARP	Address Resolution Protocol
BARON	Bandwidth-Aware Routing in Overlay Networks
BGP	Border Gateway Protocol
BTC	Bulk Transport Capacity
CBP	Cluster-Based Policy
CBR	Constant Bit Rate
CDN	Content Distribution Network
CLVL	Controlled-Loss Virtual Link
COPPA	Continuous Overlay Path Probing Algorithm
DES	Discrete Event Simulation
DHT	Distributed Hash Tables
E2E	End-to-End
FCFS	First-Come, First-Served
FEC	Forward Error Correction
FIFO	First In, First Out
GCC	GNU Compiler Collection
GUI	Graphical User Interface
HD	High Definition
ICMP	Internet Control Message Protocol
IDE	Integrated Development Environment
IEEE	Institute of Electrical and Electronics Engineers
IP	Internet Protocol
ISP	Internet Service Provider
LRD	Long-Range Dependency

MAC	Media Access Control
MPLS	Multiprotocol Label Switching
NAPA-WINE	Network Aware Peer-to-Peer Application over WIse NEtwork
NAT	Network Address Translation
NED	Network Description
NIC	Network Interface Card
OMNeT++	Objective Modular Network Test-bed in C++
OSI	Open Systems Interconnection
OWD	One-Way Delay
P2P	Peer-to-Peer
P2P-HQTV	Peer-to-Peer High Quality Television
PGM	Probe Gap Model
PPP	Point-to-Point Protocol
QoS	Quality of Service
RON	Resilient Overlay Network
RTT	Round-Trip-Time
SCTP	Stream Control Transmission Protocol
SLoPS	Self-Loading Periodic Streams
SNMP	Simple Network Management Protocol
TCP	Transmission Control Protocol
TOPP	Train Of Packet Pairs
TTL	Time-To-Live
UDP	User Datagram Protocol
VDN	Video Distribution Network
VOD	Video-On-Demand
VPN	Virtual Private Network
VPS	Variable Packet Size

SATU ALGORITMA PEMANTAUAN LALUAN PENINDIHAN ATAS SECARA BERTERUSAN UNTUK RANGKAIAN PENINDIHAN ATAS

ABSTRAK

Lebar jalur (*Bandwidth*) adalah faktor utama dalam teknologi rangkaian dan telah menjadi keutamaan sepanjang sejarah rangkaian paket. Secara faktanya, penganggaran lebar jalur sangat bermanfaat bagi mengoptimumkan prestasi perhubungan hujung-ke-hujung secara keseluruhan dalam beberapa lapisan aplikasi seperti *Content Distribution Networks (CDNs)* (Rangkaian Pengedaran Kandungan), perkongsian fail rakan-ke-rakan (*P2P*) dan penghalaan lapisan secara dinamik. Kewujudan lebar jalur yang tersedia menentukan lebar jalur tambahan yang boleh disediakan sebagai lapisan lalu lintas. Pengetahuan mengenai keadaan lebar jalur dalam sesuatu lapisan membolehkan kadar keupayaan dinamik dan penggunaan lebar jalur yang lebih baik bagi pengedaran kandungan di dalam lapisan rangkaian. Walaubagaimanapun, isu yang paling penting ialah dengan bagaimana menentukan keadaan lebar jalur secara keseluruhan (*end-to-end*) dalam lapisan rangkaian tanpa pengetahuan awal mengenai keadaan fizikal rangkaian. Lebih dua dekad yang lalu, para penyelidik telah cuba untuk mencipta algoritma untuk mengukur keadaan lebar jalur secara keseluruhan dan perkara-perkara lain yang berkaitan dengan ketepatan metrik lebar jalur, secara pantas dan tidak menjejaskan lalu lintas rangkaian. Teknik pengukuran secara aktif dilakukan melalui titik lapisan rangkaian memberikan anggaran lebar jalur dalam sesuatu rangkaian secara keseluruhan. Tesis ini menjelaskan tentang satu algoritma baru yang dikenali sebagai *"COPPA"* iaitu satu algoritma untuk mengukur keadaan penindihan laluan jalur lebar di antara penghantar dan penerima secara tepat dan berterusan. Matlamatnya ialah untuk memberikan informasi lebar jalur yang terkini untuk proses pengedaran kandungan di dalam keseluruhan rangkaian. Idea utamanya ialah untuk menjalankan pengukuran secara aktif menggunakan paket aplikasi bukan dengan menggunakan lebihan paket siasatan. Penggunaan algoritma ini mengurangkan pengiraan kos melebihi dalam lapisan yang dipilih. Sebilangan eksperimen telah dijalankan menggunakan simulasi rangka *OMNeT++*. Data eksperimen telah disahkan menggunakan teori model

algoritma yang direka sedia. Keputusan yang diperolehi menunjukkan bahawa keseluruhan jalur dalam lapisan algoritma ini memberikan informasi terkini mengenai jalur lebar dengan mengurangkan kos dan impak dalam lalu lintas rangkaian.

A CONTINUOUS OVERLAY PATH PROBING ALGORITHM FOR OVERLAY NETWORKS

ABSTRACT

Bandwidth is a key factor in network technologies and it has been of major importance throughout the history of packet networks. In fact, bandwidth estimation is very beneficial to optimize the performance of end-to-end transport in several overlay applications such as Content Distribution Networks (CDNs), Peer-to-Peer (P2P) file sharing, and dynamic overlay routing. The end-to-end available bandwidth determines the extra bandwidth that can be provided to overlay traffic. Knowledge about the available bandwidth of an overlay path enables dynamic rate adoption and better bandwidth utilization by content distribution schemes in overlay networks. However, the important issue is how to measure the available bandwidth on an end-to-end overlay path without prior knowledge about the physical network. Over the last two decades, researchers have been trying to create algorithms to measure end-to-end available bandwidth and other bandwidth-related metrics accurately, quickly, and without affecting the traffic of the path. Active measurement techniques performed by overlay nodes can provide bandwidth estimations of an end-to-end overlay path. This thesis describes a new algorithm called *"COPPA,"* which is an in-band path probing algorithm for measuring the end-to-end available bandwidth of an overlay path accurately and continuously. The aim is to provide up-to-date bandwidth information for enhanced content distribution processes in overlay networks. The primary idea is to perform active measurements using the applications' packets instead of using extra probe packets. Such an in-band probing algorithm reduces measurement overhead on the selected overlay path. Several experiments were carried out using the *OMNeT++* simulation framework. The designed algorithm was evaluated using experimental data. The obtained results show that the continuous in-band overlay path probing algorithm (COPPA) provides up-to-date bandwidth information with reduced overhead and minimal impact on the traffic of the path.

CHAPTER ONE
INTRODUCTION

This chapter is a preface to the thesis and presents the motivation, the research problem, objectives, and key contributions of the research presented in this thesis. In addition, key research steps carried out in this thesis are explained briefly. The last section of this chapter provides the organization of the thesis.

1.1 Background

Overlay networking technologies have emerged as an active area of research and development in recent years, due to their capability to provide enhanced network functionalities not provided by the predominant IP networks. Several overlay networks have been proposed in academia. Moreover, Internet and Web companies have developed different types of overlay networks to offer enhanced communication services (Tarkoma, 2010).

Recent developments in network technologies have had a profound impact on network requirements and its performance. Current trends of communications over the Internet include Peer-to-Peer (P2P) file sharing, video streaming, Video-On-Demand (VOD), and High Definition (HD) content. These trends lead to an ever-increasing load on the network and require new approaches to keep the network reliable, cost-efficient and manageable. On the other hand, the end-to-end communication nature of the Internet, which places the intelligence at the edges of the Internet, provides a natural building ground for overlay technologies (Shen, Yu, Buford, & Akon, 2009; Tarkoma, 2010).

An overlay network is a virtual network of nodes and logical links that is built on top of an existing physical network. The overlay network thus relies on the underlying IP network for basic networking functions such as routing and forwarding. A logical link between two overlay nodes may consist of a sequence of hops in the physical network (Tarkoma, 2010). A general view of an overlay network is depicted in Figure 1.1.

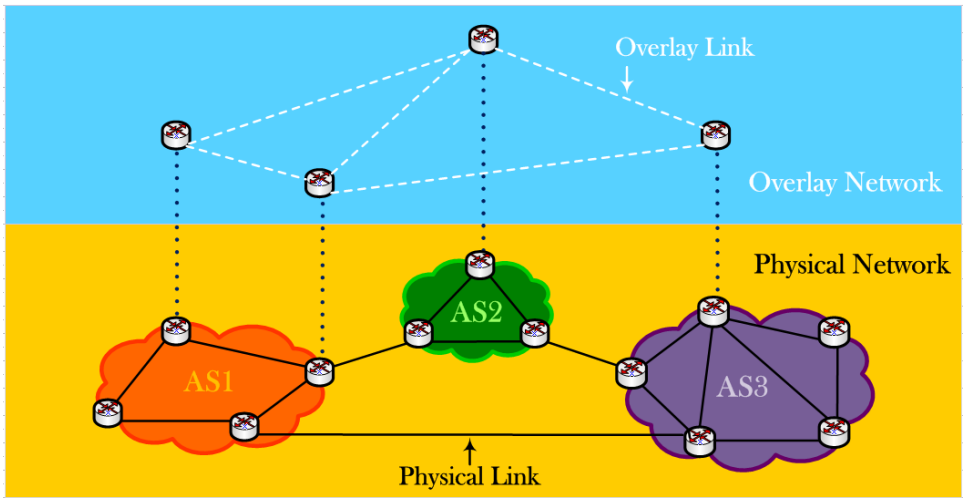

Figure 1.1: A general view of an overlay network.

The key aim of overlay technologies is to extend network features and functionality of regular IP networks in a low-cost and deployable fashion. Currently, most overlay networks are built in the application layer on top of the TCP/IP network protocol (Shen et al., 2009; Tarkoma, 2010). The enhanced services and functionalities are implemented in the application layer and provide an abstraction from the network layer. Thus, overlay network technologies can be used to overcome some of the limitations of native networks without changing the IP infrastructure. Specifically, overlay networks can solve many problems related to massive information distribution and processing tasks by enhanced functionalities provided in the application layer (Tarkoma, 2010).

1.2 Motivation

Bandwidth estimation is very beneficial in optimizing the performance of end-to-end transport in several overlay network applications. For instance, Content Distribution Networks (CDNs) and Peer-to-Peer (P2P) file sharing applications can utilize bandwidth estimates for the dynamic adoption of appropriate rates to provide a better utilization of bandwidth resources. Moreover, knowledge about the end-to-end available bandwidth of overlay paths enables simultaneous parallel downloads from multiple sources with multiple rates. This leads to speed-up of the file distribution process in a heterogeneous overlay network (Feily, Shahrestani, Noori Saleh, & Meulenberg, 2012).

Furthermore, routing overlays can benefit from timely and accurate bandwidth measurements to improve the performance of dynamic path selections, significantly (Jain & Dovrolis, 2008; Zhu, Dovrolis, & Ammar, 2006). A majority of prior research on dynamic overlay routing schemes has mainly considered delay, loss rate, and TCP throughput as the key performance metrics for path selection. They assumed that only information related to these performance metrics can be measured or inferred about the underlying network (Zhu et al., 2006). Hence, they have seldom considered bandwidth, which is a key factor in several emerging network technologies such as video streaming, multicasting, and multi-homing.

Only a few studies (Jain & Dovrolis, 2008; Ogasa et al., 2009; Zhu et al., 2006) have focused on overlay path probing approaches and there is potential for more research on this aspect of overlay networks as bandwidth is a key concept in many overlay applications. This thesis mainly focuses on an effective and efficient probing

algorithm for end-to-end bandwidth measurements along overlay paths. Such an algorithm provides the needed information continuously, with minimal probing overhead, and without affecting the current traffic in the network sharing the same physical links with overlay traffic. The primary set of target applications includes CNDs and P2P file sharing applications.

1.3 Research Problem

The available bandwidth metric is a direct indicator of the traffic load in a network path or a network link. Unlike, other network-level metrics such as delay, jitter, loss rate, or TCP throughput, the available bandwidth directly represents the extra traffic rate that a path or a link can carry, before it gets saturated (Zhu et al., 2006). However, the important issue is how to measure the available bandwidth on an end-to-end overlay path. It is relatively difficult to estimate bandwidth characteristics of links and paths due to the lack of explicit feedback related to this metric in regular IP networks (Prasad, Dovrolis, Murray, & Claffy, 2003). Therefore, overlay networks mostly rely on 'costly' active probing techniques, with high overhead, for this kind of estimations (Jain & Dovrolis, 2008; Zhu et al., 2006).

Due to scalability, it is neither efficient, nor effective for each overlay node to frequently measure the bandwidth to all other nodes (Zhu et al., 2006). There are also proposals that build an extra layer, called the social networking layer, dedicated to periodic measurements for providing network-level proximities to overlay nodes (Nakao, Peterson, & Bavier, 2003; Ogasa et al., 2009). However, this approach also causes inefficient resource utilization as bandwidth is utilized by the extra probing packets required. Moreover, the measurement results are neither accurate nor up-to-

date (see Section 2.6). Thus, an effective and efficient probing method is needed to provide relatively accurate and up-to-date bandwidth estimations to overlay nodes.

1.4 Research Objectives

The key aim of this thesis is to provide a means of up-to-date measurement of available bandwidth of overlay paths, for use in content distribution schemes for dynamic rate adoption in heterogeneous overlay networks. In this context, the main objectives of this thesis are:

- To propose and design an efficient and effective in-band path probing algorithm for content distribution networks that is capable of:

    - Performing an end-to-end measurement of the available bandwidth along an overlay path without a pre-knowledge of the link-layer network topology.

    - Probing the selected overlay path continuously while sending the actual content over TCP and thus, providing up-to-date estimates of the available bandwidth on that specific overlay path. Since the probing technique used is in-band, all measurements are performed only during the time that an overlay path is used. The bandwidth information is also up-to-date due to the in-band nature of the algorithm.

    - Improving bandwidth resource utilization by eliminating the extra social networking layer used for probing overlay networks.

- To evaluate the performance of the proposed algorithm using a theoretical model and simulations.

To achieve the objectives of this thesis, a *Continuous Overlay Path Probing Algorithm (COPPA)* is designed using an in-band probing approach, whereby application's packets are used for continuous overlay path probing.

1.5 Thesis Contributions

The key contributions of this thesis are:

- The design and development of a new algorithm (COPPA) for end-to-end available bandwidth measurement in an overlay network using an efficient in-band probing algorithm that eliminates any extra probing overhead caused by sending extra probing packets. The in-band probing algorithm exploits the applications' packets instead of sending dummy probing packets on the native network.

- Use of up-to-date bandwidth measurements in a pull-based chunk content distribution model to adjust the transmission rate of chunks' packets dynamically, according to the available bandwidth along the overlay path.

- Performance evaluation of the proposed Continuous Overlay Path Probing Algorithm (COPPA).

1.6 Key Research Steps

As mentioned earlier, the key aim of this thesis is to design a new algorithm for continuous path probing in an overlay network using an in-band probing method. To achieve this goal, a number of key research steps were carried out as demonstrated in Figure 1.2. The research was started with a comprehensive study of overlay networks and different bandwidth measurement techniques, with specific exploration of how other researchers have used active measurement techniques for probing overlay paths. Based on the findings of this study, the key requirements of an appropriate probing mechanism for overlay networks are defined. Then, a new in-band path probing algorithm is designed for overlay networks to fulfill those requirements. The algorithm design is based on a theoretical modeling framework developed for overlay path probing.

In order to examine the performance of the proposed in-band path probing algorithm in a controlled and repeatable manner, the *OMNeT++* simulation framework ("OMNeT++," 2012) was used. Simulations allow investigating the desired aspects of the algorithm, avoiding issues like route changes or multi-channel links that can distort measurement results. A small overlay network simulation model is created as a test-bed in OMNeT++, and simulation code for the in-band path probing algorithm is implemented in the *INET* framework ("INET Framework," 2012). Several simulations/experiments were then conducted using the created overlay network to collect experimental data for analysis. The designed algorithm is evaluated using the experimental data and the theoretical model of the algorithm. Finally, the performance of the proposed algorithm (COPPA) is compared with other existing overlay probing methods.

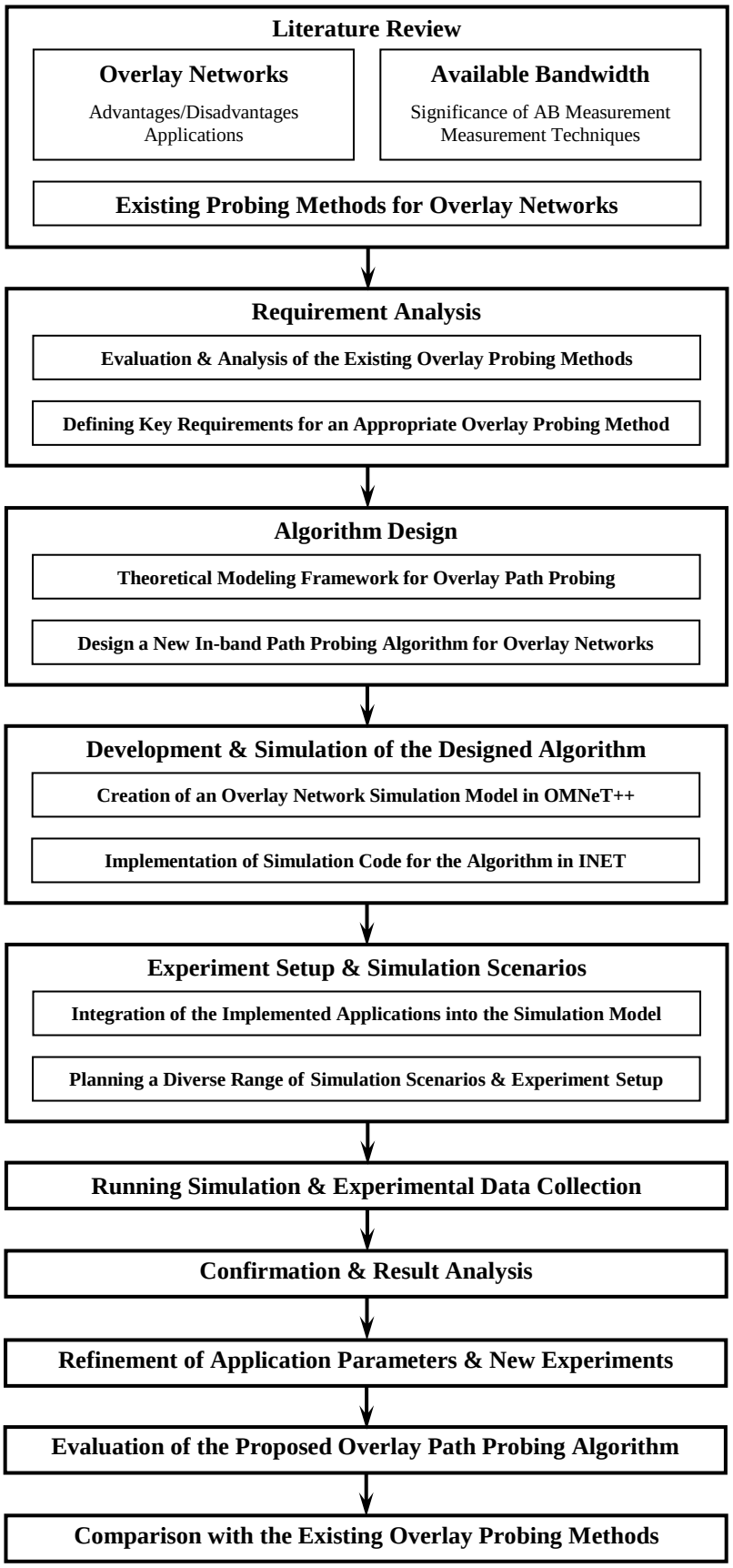

Figure 1.2: Key research steps.

1.7 Thesis Organization

This thesis is arranged into six chapters. The content is prepared in a way that each chapter provides a brief premise to the following chapter. The rest of this thesis is organized as follows.

Chapter 2 serves as the literature review, whereby fundamental concepts related to this research will be reviewed. In this chapter, different bandwidth measurement techniques will be studied thoroughly. The OMNeT++/INET simulation framework is also introduced in this chapter. Furthermore, existing approaches for overlay probing will be presented and analyzed.

Chapter 3 presents methodologies used in the research presented in this thesis. First, the theoretical modeling framework developed for overlay path probing is discussed. Then, creation of the overlay network simulation model, which is used as the test-bed in OMNeT++ environment, is described. In addition, simulation scenarios, experiment setup and data collection method are explained. Finally, key performance metrics and the methodology used for evaluation of the proposed algorithm are defined in this chapter.

Chapter 4 mainly covers the design and development of the proposed Continuous Overlay Path Probing Algorithm (COPPA). In this chapter, the design of COPPA will be discussed based on a theoretical modeling framework developed for overlay paths and traffic. In addition, the essential details about implementation of COPPA are provided in this chapter.

Chapter 5 provides an in-depth analysis of the results obtained from experiments carried out using the proposed in-band probing method. This chapter also covers the performance evaluation of COPPA and provides a comparison of the existing overlay probing approaches with the proposed in-band probing algorithm.

The thesis is concluded in **Chapter 6**, where a summary of findings and research is presented and directions for further research in this area are recommended accordingly.

CHAPTER TWO
LITERATURE REVIEW

This chapter provides the background knowledge required for better understanding of the whole thesis and reviews related work to the research presented in this thesis. The literature review in this thesis is organized into seven parts (see Figure 2.1). The first part (Section 2.1) introduces overlay networks and describes advantages, limitations, and applications of this emerging technology. In this context, content delivery in overlay networks and overlay path probing are explained. In Section 2.2, significance of bandwidth estimation is discussed. Accordingly, bandwidth measurement methods will be studied in Section 2.3. The OMNeT++/INET framework, used for simulation of the proposed overlay path probing algorithm will be introduced in Section 2.4. Then, in Section 2.5, the most related work to this research will be reviewed. Existing approaches for overlay probing will be discussed and analyzed in Section 2.6. This section will provide a summary of the literature review conducted for the research presented in this thesis. Finally, the last part (Section 2.7) summarizes the whole chapter.

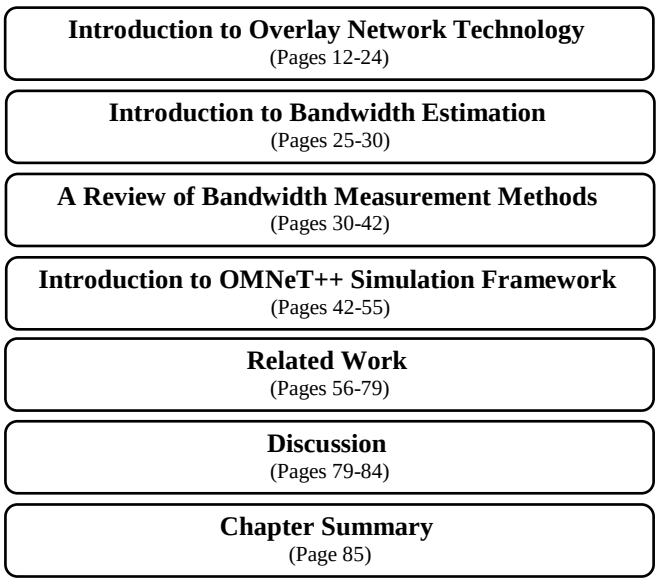

Figure 2.1: Chapter two organization.

2.1 Introduction to Overlay Network Technology

Overlay networks and Peer-to-Peer networking (P2P) have emerged as alternative solutions to solve many problems related to massive information distribution and processing tasks by providing enhanced services in the application layer. For instance, overlay networks can improve data dissemination in P2P file sharing applications and Content Distribution Networks (CDNs). Overlay technology aims to extend network features and functionality of the current IP networks in a low-cost and deployable fashion. This technology enables the introduction of enhanced networking functionalities on top of the regular IP routing mechanism. For example, overlay networks can offer new routing and forwarding features without changing the network level routers and IP infrastructures. Dynamic routing, onion routing, filter-based routing, Distributed Hash Tables (DHTs), and trigger-based forwarding are examples of new kinds of communication models enabled by overlay technologies. The key aim of many of these technologies is to provide deployable solutions for processing and distribution of vast amounts of data and at the same time reducing the scaling costs (Shen et al., 2009; Tarkoma, 2010).

Several large-scale distributed applications can take advantage of the promising characteristics of overlay networks. Peer-to-Peer networks and Virtual Private Networks (VPNs) are two typical overlay networks used for constructing large-scale distributed applications. Both P2P networks and VPNs leverage basic services provided by the communication layers beneath them to extend network features and functionality of current regular IP networks in a low cost and deployable fashion (Galan-Jimenez & Gazo-Cervero, 2011).

Although, P2P networks and VPNs have some common characteristics, they also have some differences in their design, applications, and management. Mainly, these two overlays differ in their purposes and their communications technology. While VPNs use tunneling protocols to extend an enterprise network over public networks, P2P networks create an overlay network on top of the physical IP network topology. In other words, P2P networks are designed in the application layer over the IP network, whereas VPNs can be designed over layer 1, 2, or 3 (Galan-Jimenez & Gazo-Cervero, 2011). Figure 2.2 compares a VPN with a P2P overlay network.

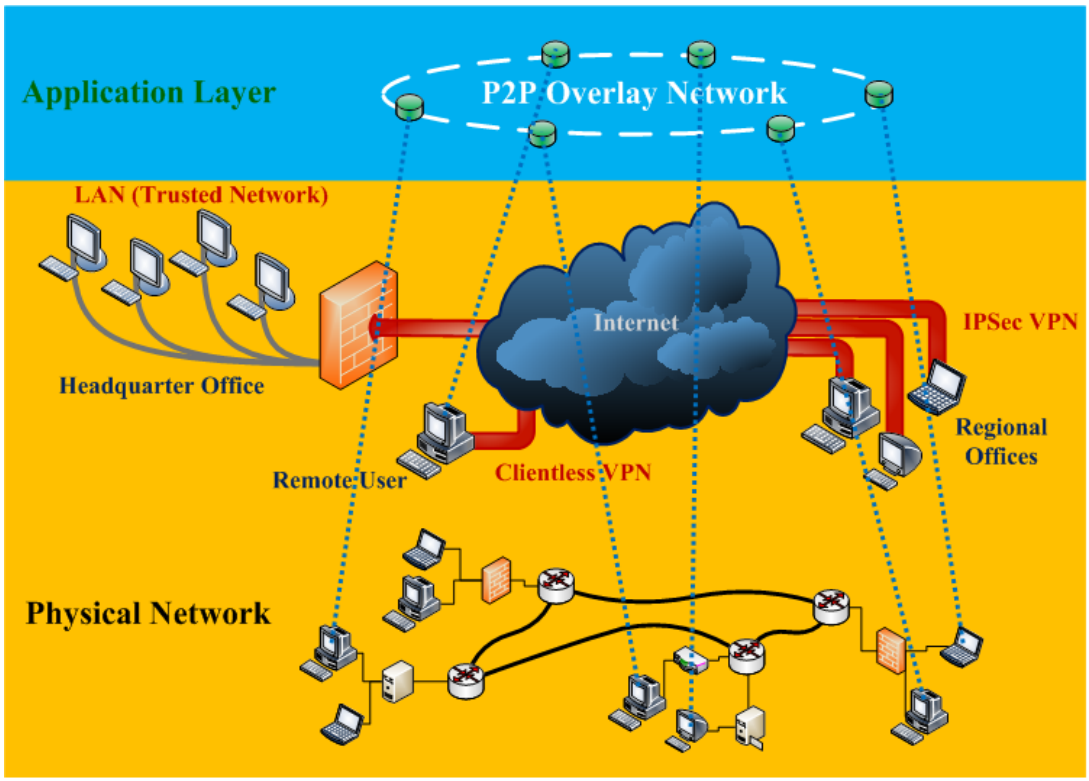

Figure 2.2: P2P overlay network vs. VPN.

2.1.1 Advantages of Overlay Networks

An overlay network offers several advantages over centralized approaches and router-based solutions. The key advantage of overlay technology is its easy deployment, as overlay networks do not require changes to the existing routers and

other network infrastructure. This easy deployment of overlays makes many other implementation issues easier as well. Moreover, overlay networks are adaptable and can utilize a number of metrics to make routing and forwarding decisions. Hence, it is possible to take application-specific concerns and performance metrics such as delay and bandwidth into account in routing and forwarding process. Such capabilities are not offered by the regular Internet infrastructure (Tarkoma, 2010).

Robustness is another key advantage of overlay networks. In fact, an overlay network is robust to node and network failure due to its adaptable nature. An overlay network with sufficient number of nodes can offer multiple independent paths to the same destination. Thus, overlay networks are able to route around faults, congestions, transient outages, or suboptimal paths through dynamic routing (Tarkoma, 2010).

2.1.2 Limitations of Overlay Networks

Overlay networks also have some limitations. For instance, overlay networks cannot be as efficient as dedicated routers in processing and routing packets due to the heterogeneous body of overlay networks. In fact, a typical overlay network consists of devices across the Internet, which introduce additional latency (stretch) into the communications. Moreover, the overlay network may not have adequate information about the Internet topology to properly optimize the routing process.

Another challenge is to overcome the reach-ability and connectivity issues in the real world. In practice, IP networks do not typically provide universal end-to-end connectivity due to the ubiquitous nature of NATs (Network Address Translators)

and firewalls. In addition, the practical deployment of overlay networks requires a proper management and administration interface that may become nontrivial when several parties are involved. Overlay networks are also vulnerable to 'malicious' nodes and other security problems (Galan-Jimenez & Gazo-Cervero, 2011; Tarkoma, 2010).

2.1.3 Types of Overlay Networks

Based on the provided services, overlay networks can be categorized into different classes as follows:

- **P2P File Sharing:** This type of overlay network is used for sharing media and data. ***BitTorrent*** ("BitTorrent," 2001), ***Gnutella*** ("Gnutella," 2000), ***KaZaA*** ("KaZaA," 2001), and ***Napster*** ("Napster," 1999) are the most popular examples of P2P file sharing overlays.

- **Content Distribution Networks (CDN):** CDNs are used for content caching and distribution to reduce delay and cost of content distribution. ***Akamai*** ("Akamai," 1999) and ***Limelight*** ("Limelight," 2001) are two well-known CDNs.

- **Routing and Forwarding Overlays:** Such overlay networks are used to reduce routing delays and costs. ***Resilient Overlay Network (RON)*** (Andersen, Balakrishnan, Kaashoek, & Morris, 2001) and ***Internet indirection infrastructure (i3)*** (Stoica, Adkins, Zhuang, Shenker, & Surana, 2004) are two examples of this type of overlay.

- **Experimental:** Experimental overlay networks offer testing grounds for new overlay technologies. For example, ***PlanetLab*** (Chun et al., 2003) is an experimental overlay network.

The focus of this thesis is on P2P file sharing overlays and Content Distribution Networks (CDNs). With this regard, content delivery in overlay networks will be explained in more detail in the following sub-section.

2.1.4 Content Delivery in Overlay Networks

Today, most content-delivery solutions utilize overlay technologies to solve various problems related to massive content-distribution and processing tasks. Compared to the traditional communication mechanisms such as IP multi-cast approaches, overlay networks offer an enhanced alternative for content delivery in terms of flexibility, scalability, and ease of deployment (Shen et al., 2009; Tarkoma, 2010). Both IP multicast and overlay multicast require active participation of several users. However, overlay multicast and content distribution is based on unicast communications and, therefore, they can work well with the current Internet (Mundinger, Weber, & Weiss, 2006). Recent P2P streaming systems have replaced the traditional client-server based video streaming solutions that incur high bandwidth provision cost on the server (Liu, Guo, & Liang, 2008).

Content distribution is considered as a key component in several overlay-network technologies, such as P2P file sharing, Content Distribution Networks (CDNs), and video streaming. Distribution in overlay networks leverages the uploading capacity of the receiving nodes to facilitate the content distribution

process. Traditional client-server file distribution systems depend on the store-and-forward mechanism, in which the content needs to be completely uploaded from the sender to the server, before it can be downloaded by the receivers. In contrast, in overlay content distribution once a node has received any portion of the content, it can redistribute that portion to any of the other receiving nodes. A general content distribution model in overlay networks is illustrated in Figure 2.3 (Feily et al., 2012; Noori Saleh, 2010).

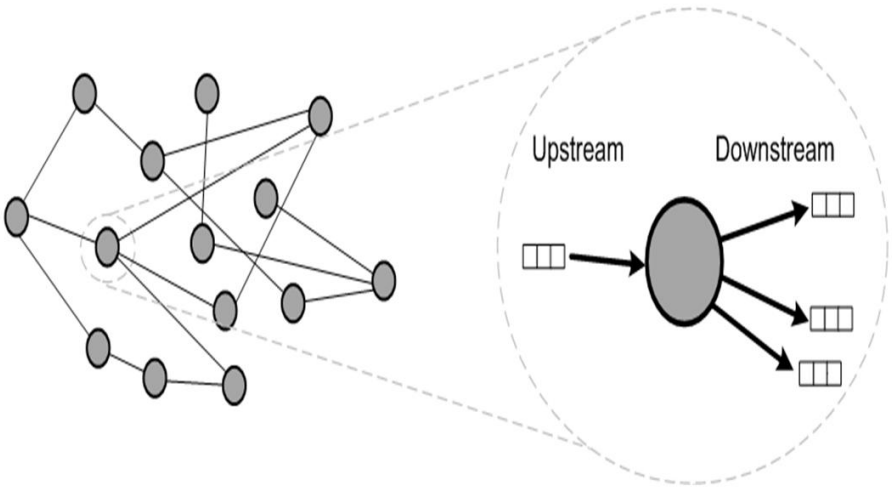

Figure 2.3: General content distribution model in overlay networks (Noori Saleh, 2010).

A P2P overlay network may consist of either homogeneous or heterogeneous peers. In contrast to homogeneous peers, which have identical network access and uploading bandwidth, heterogeneous peers have different types of network access and therefore, different uploading bandwidth (Noori Saleh, 2010; Noori Saleh, Feily, Ramadass, & Shahrestani, 2012). The high-level architecture of a typical Content Distribution network (CDN) is demonstrated in Figure 2.4.

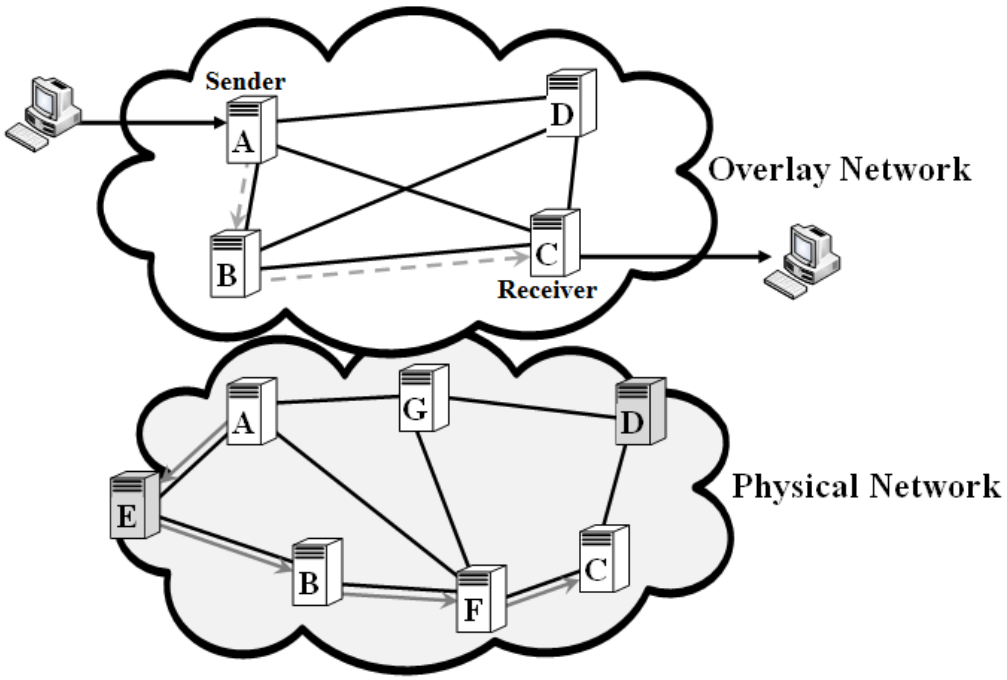

Figure 2.4: The high-level architecture of a typical Content Distribution Network (CDN).

The primary concept of P2P technologies is to encourage users or peers to act as both clients and server to achieve better scalability and robustness than traditional client-server based systems. In a P2P network, not only a peer downloads data from the network, but also it uploads the downloaded data to other peers in the network. Therefore, the uploading bandwidth of peers is efficiently utilized, and bandwidth burdens are reduced considerably (Liu et al., 2008; Wang, Wang, Yang, & An, 2011).

In this context, a viable content distribution model leads to efficient utilization of resources such as network bandwidth, and addresses the heterogeneity issue in overlay networks. Moreover, a proper content distribution model could considerably minimize the total download time of the content, whereas a poor model could result in longer distribution time (Feily et al., 2012; Noori Saleh et al., 2012; Noori Saleh, Feily, Ramadass, & Hannan, 2011).

2.1.4.1 Content Distribution Models

Content distribution in overlay networks is generally based on two models (Noori Saleh, 2010; Noori Saleh et al., 2011; Noori Saleh et al., 2012):

- **Fluid Content Distribution Model**
- **Chunk Content Distribution Model**

Fluid content distribution models provide continuous transfer of content from a source to multiple receivers. However, deploying a Fluid model in heterogeneous overlay networks requires special consideration due to the incorporation of tightly coupled connections between adjacent nodes in this model. Let us explain tightly coupled connections, first. In a Fluid model, a receiving node should distribute each single bit of the content once it has received that bit. This feature is denoted as tightly coupled connections and it cannot accommodate the network dynamics and asymmetric bandwidth in heterogeneous overlay networks properly. Tightly coupled connections will significantly degrade the performance of all overlay nodes in a heterogeneous overlay network, where participating nodes have different download time and bandwidth resources.

In contrast to Fluid content distribution models, in Chunk models all connections among peers are loosely coupled. The loosely coupled connections accommodate asymmetric bandwidth, and therefore the model suites the heterogeneity of the Internet, and especially heterogeneous overlay networks. In a Chunk model, in order to maximize the participation of each node in the overlay, large contents are typically divided into many small pieces called *"Chunks;"* these

chunks have significantly greater size than the IP packets. Chunks are the smallest units that are directly exchanged between the overlay nodes. A peer will not distribute a specific chunk until it receives that chunk entirely. Accordingly, overlay nodes have to wait to receive the entire chunk before they can forward it to other nodes. However, this may become untenable in that content transfer may take a long time, during which the upload capacity of downloading peers is wasted. In other words, peers' uploading bandwidth is not fully utilized in this model. Both content distribution models are shown in Figure 2.5.

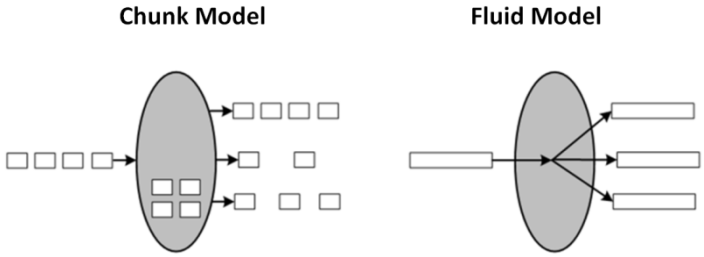

Figure 2.5: Chunk and Fluid content distribution models (Noori Saleh, 2010).

Each of these content distribution models has its own advantages and disadvantages. There are several proposals to improve the performance of Fluid and Chunk content distribution models by employing different strategies as discussed by (Noori Saleh, 2010; Noori Saleh et al., 2011; Noori Saleh et al., 2012). Overall, Chunk content distribution models are commonly used in P2P file sharing applications and Content Distribution Networks. Network coding approaches and scheduling mechanisms have been proposed to overcome the limitation of Chunk model. While, the feasibility of network coding in P2P file sharing and content distribution is doubted (Wang & Li, 2006), scheduling can increase the availability of chunks, and provide proper data dissemination. A viable scheduler can reduce the completion time, and enables efficient utilization of peers' resources such as local

storage and network bandwidth (Chan, Li, & Lui, 2005; Guo, Liang, & Liu, 2008; Ma & King-Shan, 2008; Mundinger, Weber, & Weiss, 2008; Ren, Li, & Chan, 2008).

Researchers (Han & Xia, 2009; Rodriguez & Biersack, 2002; Xu, Xianliang, Mengshu, & Chuan, 2005) have shown that using a parallel downloading scheme in P2P file sharing systems could result in higher aggregate download rates and thus shorter download times. In a parallel downloading scheme, an end user opens multiple connections to multiple file sources to download different portions of the file from different sources, and then it reassembles the file locally. The multiple connections in parallel download scheme need an effective scheduling algorithm, to get significant performance improvements in collaborative file sharing (Han & Xia, 2009).

2.1.4.2Transmission Models

In order to understand scheduling algorithms, one needs to distinguish two models used for controlling transmissions, which are:

- **Push-based Transmission Model**
- **Pull-based Transmission Model**

In both models, peers have to periodically exchange information about parts of the file or video stream they possess. However, in a push-based model, the sender peer determines which piece of data should be transmitted and which peer should receive that piece of the file or video. This scheduling approach was first adopted by

(Ma & King-Shan, 2008). On the other hand, in a pull-based model, the receiver node determines which pieces of the file it needs from other peers. Subsequently it sends request messages to the peers it chooses. The file source that receives request messages could either accept or reject these requests based on some policies, such as the available bandwidth and the contribution of the requesting peer (Noori Saleh et al., 2012).

Most of existing P2P applications, such as ***BitTorrent*** ("BitTorrent," 2001), use pull-based models. Nevertheless, this model has some disadvantages. First of all, a considerable amount of network bandwidth and processing time might be wasted if many request messages traverse the network at nearly the same time and cause congestion. In addition, it may happen that multiple peers decide to request the same piece of a file from the same source. In this situation, the queuing time at the source node will be increased dramatically, and request messages might even be rejected by the source node (Zhang, Zhang, Sun, & Yang, 2007).

2.1.5 Overlay Path Probing

In dynamic environments such as the Internet, overlay nodes need to perform path probing periodically to monitor the quality of paths to other nodes. Pair-wise probing is the straightforward solution, which is employed by the ***Narda*** protocol (Chu, Rao, Seshan, & Zhang, 2002) and ***RON*** (Andersen et al., 2001). In this solution, each overlay node periodically probes all paths to all other nodes to find all possible alternative paths. Although such full-scale probing is complete and accurate, the probing overhead (number of probing packets) is *O(n2)*, where *n* is the number of overlay nodes. Such probing overhead is too expensive for large overlay networks.

Specifically, pair-wise probing may incur high link-stress in sparse networks with low number of links, (i.e., compared to dense networks in which the number of links in each node is close to the total number of nodes in the network.) like the Internet (Faloutsos, Faloutsos, & Faloutsos, 1999), and overlay networks, where overlay paths usually share physical links. High link-stress affects cross-traffic on links, and also the probing results (Tang & McKinley, 2003).

Several approaches have been proposed to reduce probing overhead in large-scale overlay networks. A thread of research has focused on providing a balance between the probing cost and probing completeness. Different techniques such as approximation, aggregation, and hierarchy have been employed to improve the scalability of overlay path probing (Banerjee, Bhattacharjee, & Kommareddy, 2002; Braynard, Kostic, Rodriguez, Chase, & Vahdat, 2002; Rowstron & Druschel, 2001; B. Zhang, Jamin, & Zhang, 2002; Zhao et al., 2004). For instance, ***Pastry*** (Rowstron & Druschel, 2001) and ***Tapestry*** (Zhao et al., 2004) and have reduced the number of probing packets to $O(n \log n)$ in structured peer-to-peer networks by probing only a small subset of the possible paths in each round. In addition, scalable application level multicast systems such as ***NICE*** (Banerjee et al., 2002) and ***HMTP*** (Zhang et al., 2002) have reduced the total probing overhead to $O(n)$ by organizing overlay nodes in a hierarchy based on the distance between overlay nodes. Therefore, in order to find its optimal location in the hierarchy, each node periodically selects and probes a particular set of nodes with constant size. These approaches require information about network proximities such as node distance, node degree, *etc*.

On the other hand, researchers at Michigan State University (Tang & McKinley, 2003) have focused on the trade-off between probing cost and estimation accuracy in sparse networks such as the Internet. They proposed a probing method, which uses network-level path composition information to infer path quality without full-scale probing. This method periodically probes a subset of overlay paths. After probing each overlay path, it makes coarse quality estimation for all physical links on that path. Then, based on these estimations it infers the quality of other paths that contain those physical links. Similar to topology-aware overlay networks (Han, Watson, & Jahanian, 2005; Kwon & Fahmy, 2002), this solution requires prior network topology information. It assumes the availability of network topology information at end nodes. However, such information is not typically available at the end hosts. Besides, overlay network technologies are moving toward unstructured topologies and P2P networks. Thus, in order to be practical in a real Internet environment, an overlay path probing method should not depend on information about underlying network topology.

Overall, it seems more logical to exploit end-to-end path probing in overlay networks rather than performing topology-aware path probing. Therefore, this thesis will focus on overlay probing techniques that do not require prior knowledge about the physical network topology and proximities. In this regard, bandwidth estimation, related metrics, and measurement methods will be studied in the following sections (2.2 and 2.3).