Overlay Networks

Toward Information Networking

Sasu Tarkoma

CRC Press
Taylor & Francis Group
AN AUERBACH BOOK

Overlay Networks

Toward Information Networking

Sasu Tarkoma

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
AN AUERBACH BOOK

Auerbach Publications
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number: 978-1-4398-1371-3 (Hardback)

Library of Congress Cataloging-in-Publication Data

Tarkoma, Sasu.
Overlay networks : toward information networking / Sasu Tarkoma.
p. cm.
Includes bibliographical references and index.
ISBN 978-1-4398-1371-3 (hardcover : alk. paper)
1. Information networks. 2. Peer-to-peer architecture (Computer networks) 3. Electronic data processing--Distributed processing. 4. Multimedia communications. I. Title.

TK5105.5.T366 2010
004.6'52--dc22 2009046412

Overlay Networks

Toward Information Networking

Preface

Data and media delivery have become hugely popular on the Internet, with well over 1 billion Internet users. Therefore scalable and flexible information dissemination solutions are needed. Much of the current development pertaining to services and service delivery happens above the basic network layer and the TCP/IP protocol suite because of the need to be able to rapidly develop and deploy them.

In recent years, various kinds of overlay networking technologies have emerged as an active area of research and development. Overlay systems, especially *peer-to-peer* systems, are technologies that can solve problems in massive information distribution and processing tasks. The key aim of many of these technologies is to be able to offer deployable solution for processing and distributing vast amounts of information, typically petabytes and more, while at the same time keeping the scaling costs low.

The aim of this book is to present the state of the art in overlay technologies, examine the key structures and algorithms used in overlay networks, and discuss their applications. Overlay networks have been a very active area of research and development during the last 10 years, and a substantial amount of scientific literature has formed around this topic.

This book has been inspired by the teaching notes and articles of the author in content-based routing. The book is designed not only as a reference for overlay technologies, but also as a textbook for a course in distributed overlay technologies and information networking at the graduate level.

About the Author

Sasu Tarkoma received his M.Sc. and Ph.D. degrees in Computer Science from the University of Helsinki, Department of Computer Science. He is currently professor at Helsinki University of Technology, Department of Computer Science and Engineering. He has been recently appointed as full professor at University of Helsinki, Department of Computer Science. He has managed and participated in national and international research projects at the University of Helsinki, Helsinki University of Technology, and Helsinki Institute for Information Technology (HIIT). He has worked in the IT industry as a consultant and chief system architect, and he is principal member of research staff at Nokia Research Center. He has over 100 publications, and has also contributed to several books on mobile middleware.

Ms. Nelli Tarkoma produced most of the diagrams used in this book.

Contents

1

Introduction

1.1 Overview

In recent years, various kinds of overlay networking technologies have emerged as an active area of research and development. Overlay systems, especially *peer-to-peer (P2P)* systems, are technologies that can solve problems in massive information distribution and processing tasks. The key aim of many of these technologies is to be able to offer deployable solution for processing and distributing vast amounts of information, typically petabytes and more, while at the same time keeping the scaling costs low.

Data and media delivery have become hugely popular on the Internet. Currently there are over 1.4 billion Internet users, well over 3 billion mobile phones, and 4 billion mobile subscriptions. By 2000 the Google index reached the 1 billion indexed web resources mark, and by 2008 it reached the trillion mark.

Multimedia content, especially videos, are paving the way for truly versatile network services that both compete with and extend existing broadcast-based medias. As a consequence, new kinds of social collaboration and advertisement mechanisms are being introduced both in the fixed Internet and also in the mobile world. This trend is heightened by the ubiquitous nature of digital cameras. Indeed, this has created a lot of interest in community-based services, in which users create their own content and make it available to others.

These developments have had a profound impact on network requirements and performance. Video delivery has become one of the recent services on the Web with the advent of YouTube [67] and other social media Web sites. Moreover, the network impact is heightened by various P2P services. Estimates of P2P share of network traffic range from 50% to 70%. Cisco's latest traffic forecast for 2009–2013 indicates that annual global IP traffic will reach 667 exabytes in 2013, two-thirds of a zettabyte [79]. An exabyte (EB) is an SI unit of information, and 1 EB equals 10^{18} bytes. Exabyte is followed by the zettabyte ($1 \text{ Z} = 10^{21}$) and yottabyte ($1 \text{ Y} = 10^{24}$). The traffic is expected to increase some 40% each year. Much of this increase comes from the delivery of video data in various forms. Video delivery on the Internet will see a huge increase, and the volume of video delivery in 2013 is expected to be 700 times the capacity of the US Internet backbone in 2000. The study anticipates that video traffic will account for 91% of all consumer traffic in 2013.

According to the study, P2P traffic will continue to grow but will become a smaller component of Internet traffic in terms of its current share. The current P2P systems in 2009 are transferring 3.3 EB data per month. The recent study indicates that the P2P share of consumer Internet traffic will drop to 20% by 2013, down from the current 50% (at the end of 2008). Even though the P2P share may drop, most video delivery solutions, accounting for much of the traffic increase, will utilize overlay technologies, which makes this area crucial for ensuring efficient and scalable services.

A P2P network consists of nodes that cooperate in order to provide services to each other. A pure P2P network consists of equal peers that are simultaneously clients and servers. The P2P model differs from the *client-server* model, where clients access services provided by logically centralized servers.

To date, P2P delivery has not been successfully combined with browser-based operation and media sites such as YouTube. Nevertheless, a number of businesses have realized the importance of scalable data delivery. For example, the game company Blizzard uses P2P technology to distribute patches for the *World of Warcraft* game. Given the heavy use of network, P2P protocols such as BitTorrent offer to reduce network load by peer-assisted data delivery. This means that peer users cooperate to transfer large files over the network.

1.2 Overlay Technology

Data structures and algorithms are central for today's data communications. We may consider circuit switching technology as an example of how information processing algorithms are vital for products and how innovation changes markets. Early telephone systems were based on manual circuit switching. Everything was done using human hands. Later systems used electromechanical devices to connect calls, but they required laborious preconfiguration of telephone numbers and had limited scalability. Modern digital circuit switching algorithms evolved from these older semiautomatic systems and optimize the number of connections in a switch. The nonblocking minimal spanning tree algorithm enabled the optimization of these automatic switches. Any algorithm used to connect millions of calls must be proven to be correct and efficient. The latest development changes the fundamentals of telephone switching, because information is forwarded as packets on a hop-by-hop basis and not via preestablished physical circuits. Today, this complex machinery enables end-to-end connectivity irrespective of time and location.

Data structures are at the heart of the Internet. Network-level routers use efficient algorithms for matching data packets to outgoing interfaces based on prefixes. Internet backbone routers have to manage 200,000 routes and more in order to route packets between systems. The matching algorithms include *suffix trees* and *ternary content addressable memories (TCAMs)* [268], which have to balance between matching efficiency and router memory. Therefore, just as with telephone switches, optimization plays a major role in the development of routers and routing systems.

The current generation of networks is being developed on top of TCP/IPs network-layer (layer 3 in the *open systems interconnection (OSI)* stack). These so-called overlay networks come in various shapes and forms. Overlays make many implementation issues easier, because network-level routers do not need to be changed. In many ways, overlay networks represent a fundamental paradigm shift compared to older technologies such as circuit switching and hierarchical routing.

Overlay networks are useful both in control and content plane scenarios. This division of traffic into control and content is typical of current telecommunications solutions such as the *session initiation protocol (SIP)*; however, this division does not exist on the current Internet as such. As control plane elements, overlays can be used to route control messages and connect different entities. As content plane elements, they can participate in data forwarding and dissemination.

An *overlay network* is a network that is built on top of an existing network. The overlay therefore relies on the so-called *underlay* network for basic networking functions, namely routing and forwarding. Today, most overlay networks are built in the application layer on top of the TCP/IP networking suite. Overlay technologies can be used to overcome some of the limitations of the underlay, at the same time offering new routing and forwarding features without changing the routers. The nodes in an overlay network are connected via logical links that can span many physical links. A link between two overlay nodes may take several hops in the underlying network.

An overlay network therefore consists of a set of distributed nodes, typically client devices or servers, that are deployed on the Internet. The nodes are expected to meet the following requirements:

1. Support the execution of one or more distributed applications by providing infrastructure for them.
2. Participate in and support high-level routing and forwarding tasks. The overlay is expected to provide data-forwarding capabilities that are different from those that are part of the basic Internet.
3. Deploy across the Internet in such a way that third parties can participate in the organization and operation of the overlay network.

Figure 1.1 presents a layered view to overlay networks. The view starts from the underlay, the network that offers the basic primitives of sending and receiving messages (packets). The two obvious choices today are UDP and TCP as the transport layer protocols. TCP is favored due to its connection-oriented nature, congestion control, and reliability.

After the underlay layer, we have the custom routing, forwarding, rendezvous, and discovery functions of the overlay architecture. Routing pertains to the process of building and maintaining routing tables. Forwarding is the process of sending messages toward their destination, and rendezvous is a function that is used to resolve issues regarding some identifier or node—for example, by offering indirection support in the case of mobility. Discovery is an integral part of this layer and is needed to populate the routing table by discovering both physically and logically nearby neighbors.

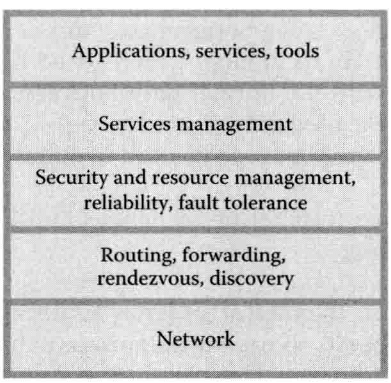

FIGURE 1.1
Layered view to overlay networks.

The next layer introduces additional functions, such as security and resource management, reliability support, and fault tolerance. These are typically built on top of the basic overlay functions mentioned above. Security pertains to the way node identities are assigned and controlled, and messages and packets are secured. Security encompasses multiple protocol layers and is responsible for ensuring that peers can maintain sufficient level of trust toward the system. Resource management is about taking content demand and supply into account and ensuring that certain performance and reliability requirements are met. For example, relevant issues are data placement and replication rate. Data replication is also a basic mechanism for ensuring fault-tolerance. If one node fails, another can take its place and, given that the data was replicated, there is no loss of information.

Above this layer, we have the services management for both monitoring and controlling service lifecycles. When a service is deployed on top of an overlay, there need to be functions for administering it and controlling various issues such as administrative boundaries, and data replication and access control policies.

Finally, in the topmost layer we have the actual applications and services that are executed on top of the layered overlay architecture. The applications rely on the overlay architecture for scalable and resilient data discovery and exchange.

An overlay network offers a number of advantages over both centralized solutions and solutions that introduce changes in routers. These include the following three key advantages:

Incremental deployment: Overlay networks do not require changes to the existing routers. This means that an overlay network can be grown node by node, and with more nodes it is possible to both monitor and control routing paths across the Internet from one overlay node to another. An overlay network can be built based on standard network protocols and existing APIs—for example, the Sockets API of the TCP/IP protocol stack.

Adaptable: The overlay algorithm can utilize a number of metrics when making routing and forwarding decisions. Thus the overlay can take application-specific concerns into account that are not currently offered by the Internet infrastructure. Key metrics include latency, bandwidth, and security.

Robust: An overlay network is robust to node and network failures due to its adaptable nature. With a sufficient number of nodes in the overlay, the network may be able to offer multiple independent (router-disjoint) paths to the same destination. At best, overlay networks are able to route around faults.

The designers of an early overlay system called *resilient overlay network (RON)* [361] used the idea of alternative paths to improve performance and to route around network faults. Figure 1.2 illustrates how overlay technology can be used to route around faults. In this example, there is a problem with the normal path between A and B across the Internet. Now, the overlay can use a so-called *detour path* through C to send traffic to B. This will result in some networking overhead but can be used to maintain communications between A and B.

Overlay networks face also a number of challenges and limitations. The three central challenges include the following:

- The real world: In practice, the typical underlay protocol, IP, does not provide universal end-to-end connectivity due to the ubiquitous nature of firewalls and *network address translation (NAT)* devices. This means that special solutions are needed to overcome reachability issues. In addition, many overlay networks are oblivious to the current organizational and management structures that exist in applications