Lecture 25: Security Issues in Structured Peer-to-Peer Networks

Lecture Notes on "Computer and Network Security"

by Avi Kak (kak@purdue.edu)

April 18, 2017
3:23pm

Goals:

- What are peer-to-peer (P2P) overlay networks

- Distributed hash tables (DHT)

- The Chord protocol

- The Pastry protocol

- The Kademlia Protocol

- The BitTorrent File Sharing Protocol

- Security Aspects of Structured DHT-Based P2P Protocols

- Anonymity in Structured P2P Overlay Networks

- An Answer to "Will I be Caught?"

CONTENTS

25.1: WHAT ARE PEER-TO-PEER OVERLAY NETWORKS?

- Services in traditional networks (such as the internet) are typically based on the client-server model. Examples include web services provided by your web servers (such as the HTTPD servers) and the browsers that act as clients vis-a-vis the servers. Another common example would be the email servers that are in charge of transporting (sending and receiving) email over the internet and the client email programs running on your personal machine that download email from designated servers.

- Therefore, the traditional services on the internet are based on the concept of central repositories of information; those who wish to see this information must make download requests to the central repositories. This is the same as the relationship between a library and you as a user/member of that library.

- Services in peer-to-peer (P2P) networks are based more on the notion of a book club. All the participants in a P2P network share equally all the information of mutual interest. [Sharing in the context of a P2P book-club could mean that, for the sake of overall efficiency in storage, a member

participating in a network may choose to store only that chapter that he/she is currently reading. When he/she decides to look at a chapter that is not currently in his/her own computer, the computer would know automatically how to fetch it from one of the other members participating in the P2P network.]

- Since P2P networks work in a decentralized fashion, there is no machine that acts as a coordinator in the network. All the machines in a P2P network possess the same capability as far as the network is concerned. The machines participating in a P2P network are frequently referred to as **nodes**. [Note that the earliest P2P systems that made this acronym virtually a household word did possess centralized components. **Napster** was the first P2P system that became very popular for sharing music files. Its functioning required a central database for mapping the song titles to the hosts where the songs were actually stored. Then came **BitTorrent** for P2P downloading of large multimedia objects such as movies. The earliest version of BitTorrent also required the notion of a central coordinator that was called the **tracker** which kept track of who had what segments of a large movie file. If you allow for centralized coordinators, constructing a P2P system for file sharing is a relatively easy thing to do. Let's say you are a content provider and you want your files to be downloaded through P2P file sharing. All you have to do is to provide at your web-site a tracker that keeps track of who has requested what file and a client program that folks can download. It would be the job of the client program to talk to the tracker program at your website. As a user, your client program will request a file from the tracker and the tracker would supply your client program with a list of all users currently in possession of the various segments of the file you want (and, at the same time, add you to the list of users who could be in possession of some segments of the file in question). Your client program would then request the various segments of the file from their keepers and assemble

them back into the file that you were looking for.]

- As we will see here, the nodes in a P2P network are also **self-organizing**. That is, each new incoming node knows where to place itself in an overall organization of all the participating nodes.

- In addition to being self-organizing, and partly because of it, P2P protocols can allow for such networks to scale up easily.

- Because all nodes participating in a *modern* P2P network operate in an identical fashion and without the help of any sort of a central manager, P2P networks can be characterized as **distributed systems**. [The distributed nature of P2P networks also makes them more **fault tolerant**. That is, the sudden failure of one or more nodes in a network does not bring down the network. When node failures do take place, the rest of the network adapts gracefully. For that reason, P2P systems can also be called **adaptive**.]

- P2P networks are usually overlaid on top of the internet. For that reason, they are also referred to as **overlay networks** or just **overlays**.

- We can therefore talk about routing in the underlying network (usually the internet) and routing in the overlay.