582615 - Overlay and P2P Networks, Spring 2016

Exercise 3 model answers

Considering the following points in the answer granted points in the assignment.

Assignment 1 - Power law (5 points)

If the degree (number of connections) of nodes in a network follows a power law, what does this mean?

A power law means a relationship between two quantities where one quantity varies as a power of the other. When talking about degree distribution in a network, it means that such a relationship exists between the degrees on nodes present on the network, and the number of nodes with each degree. In networks, this typically means that there is a large amount of nodes with a small amount of connections, and a small amount of nodes with a very large amount of connections (i.e., the exponent is negative).

What are Small worlds and Scale-free networks? What features do they exhibit?

Small worlds are networks where the average shortest path between nodes *(network diameter)* is short. The average shortest path length grows proportionally to the logarithm of the number of nodes. These kinds of networks tend to contain cliques and hubs, i.e. have local, highly clustered parts that are less tightly connected to other similar parts.

Scale-free networks are networks where the degree distribution of the nodes follows a power law. This essentially results in a small world network, with the special condition that to satisfy the power law, scale-free networks tend to be even more clustered and have an even smaller diameter.

Diameters
Small world:

$$L \propto \log N$$

Scale-free:

$$L \propto \log \log N$$

Assume that nodes on a scale-free network are removed one at a time. How does it affect robustness if the removed nodes are chosen at random, or in a coordinated way? Which nodes' removal causes the largest impact?

On a scale-free network, the number of nodes with a small number of connections heavily exceeds the number of nodes with a large amount of connections. When choosing the node at random, with all nodes having an equal probability to get chosen, it's more likely that the chosen node only has a small number of connections and is less important to the robustness of the network.

A coordinated attack where the most heavily connected nodes *(hubs)* are taken down first, would have the largest impact on the network, as a large number of paths is likely to depend on these nodes.

Assignment 2 - Clustering coefficient (5 points)

The average clustering coefficient of a network is the average of local clustering coefficients of its vertices. The local clustering coefficient is computed from how many of connections exist between a vertice's neigbours, divided by how many connections there would be if they all were neighbors of each other. Another way to think of it is through the presence and number of triangles in the network. In this exercise, it was instructed to use 0 as the clustering coefficient for leaf vertices.

In A, the graph did not have any triangles, meaning that all vertices were either leaf vertices, or no vertice had neighbors that were neighbors to each other.

	Local clustering coefficients							
Network	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**avg**
A) Acyclic graph	0	0	0	0	0	0	0	**0**
B) Some clustering	0	0	0	1/3	1	1/3	0	**5/21** (~0.24)
C) Almost full mesh	1	5/6	1	5/6	5/6			**9/10** (0.90)

Assignment 3 - Freenet (5 points)

What is Freenet? What services does it provide to its users? (not what content can be found)

Freenet is a privacy overlay aimed to secure freedom of speech online. It provides a secure way of distributing and receiving content that may be blocked or monitored on other kinds of networks. Freenet hides its user's identities behind pseudonyms and provides services such as a publication platform for basically any kinds of documents and a pseudonymous email service.

When a user requests for a file, how does Freenet route the request?

Freenet uses a depth-first search with backtracking, trying the peers whose location is lexicographically closest to the key of the requested file. Freenet starts by checking the local storage for the content, if it did not already possess it, it will request the peer with the closest key for the content. They will then repeat this process until the content is found or a TTL value is reached. The result is then carried back the same route that the request took.

What steps does Freenet take to protect the anonymity of the requester and the data holder?

Freenet does a lot of things to protect the privacy of its users.

Chaum's mix-net scheme for anonymous communication is used for providing anonymity for the first step in the Freenet network. Messages are routed and forwarded through a path of peers instead of directly connecting to the target peer for the request. Each step is individually encrypted and each peer only knows about the previous step and the next step on the path. The peers in the mix do not know if the previous peer was the original peer or just another peer in the mix net, nor do they know if the next peer is the actual target peer. The purpose of this is to hide the real origin of the request.

Requests and responses also follow a path among the peers. All connections between peers are encrypted. An outside observer can't learn what content is being transferred between the peers.

All content is encrypted and keys in Freenet are semantic free, i.e. the similarity of two keys says nothing about how similar the files are. As peers carry files with keys similar to the peers location, it does not mean that the files they carry have anything to do with each other. This provides a kind of deniability to the owners of those peers; they can't really know what kind of files they are hosting.

Content can be signed with a private key; the pseudonymous origin of the content can be verified. Even though the pseudonym can't be tracked to a real person, "identity theft" of the pseudonym can be prevented.

Intermediate peers in the paths of messages may cache the messages, and place themselves as the origin of the forwarded message. This serves the purpose of providing anonymity to both the origin of the request, and the peer that originally held the data. It also serves to provide additional replicas for popular data.

What is Freenet's darknet operating mode, and how is it different from the default setup, the opennet?

The darknet operating mode is an alternative operating mode that enhances security at the cost of manual configuration and potentially limited availability. The purpose of a darknet is to isolate a peer from direct connections to untrusted peers.

In the regular opennet operating mode, Freenet bootstraps itself by connecting to a set of known seed nodes (list comes with the client), and then further connecting to peers on a list received from the seed nodes. In darknet, the client does not connect to seed nodes. Instead it connects only to peers from whom they have gotten a darknet invitation.

Another difference is that darknet supports location swapping, i.e. the nodes can compare each others location labels and swap them if it seems to result in a more optimal network. On opennet, this is not supported because it exposes topology and allows for certain other attacks. In the darknet, the peers are assumed to be trusted, which means that they are also trusted not to perform these attacks.

Assignment 4 - Future of overlays and P2P (5 points)

How is a P2P cloud different from a centrally hosted cloud? What are the benefits and the drawbacks?

The idea of a P2P cloud is to use ordinary office computers as building blocks for a cloud. A centrally hosted cloud is usually owned by somebody else, while a P2P cloud would likely be constructed on in-house hardware. The actual hardware itself would probably be somewhat different as well.

Benefits	Drawbacks
No need to trust service provider	Scalability limited by available workstations
Data remains on premises	Less controlled environment
Increased utilization of workstations	Demands higher-end workstations
Less need to invest in servers / datacenters	Increased local energy consumption
No running costs from cloud provider	Network probably lesser than at data center

How can SDN and overlays be integrated and what benefits can be obtained from their integration?

SDN is a way to control network data flows programmatically. Integrating SDN and overlays would involve integrating the overlay with an SDN controller, to give it instructions on how to serve the overlay better. Overlays can benefit from SDN integration e.g. by getting a better ability to learn about the structure and condition of the underlying network and being able to optimize their routing accordingly, or change the routing in the underlaying network to better suit the needs of the overlay.

How could a P2P cloud benefit from SDN-compatible networking?

The P2P cloud can be considered an example of an overlay that could benefit from SDN integration. Using one physical topology, SDN would allow casting different virtual networks for the P2P cloud virtual machines, and to optimize routing based on the topologies. SDN would also allow adapting the network i.e. if a virtual machine is migrated to a more powerful workstation, the data flows in the network can be reconfigured accordingly.

Turbo challenge - Tor (5 points)

How does Tor anonymise the browsing?

Tor disguises the location of the host on the Internet by using a Chaum's mix-net type of setup, and uses Onion routing to encrypt the traffic until it's no longer close to the original host. The outers in the mix-net are changed when connecting to different targets. Other than that, it doesn't. Several projects, such as the Tor browser (a special version of Firefox) exist to provide further anonymity on Tor.

How does *Onion routing* work? i.e. What happens while a message travels across the Tor network.

In Onion routing, a set of Onion routers is chosen from a list provided by a directory server. The servers are randomly organized into a path (or *circuit*). The original host contacts the first Onion router, and forms an encrypted link with that. The original host then makes the first router to contact the second router with encryption data from the original host. Then it makes the second router contact the third router with encryption data from the original host etc. Each step of the circuit is encrypted between the original host and the next Onion router. This results in an encryption setup that resembles layers of an onion, with different layers of encryption inside each other. The last router is an exit node. It will not send encrypted Tor traffic further, but will be the point where the original request reaches the Internet.

Once the circuit is formed, packets are encrypted with this multiple layers of encryption, and decrypted layer by layer as it traverses the mix-net hop by hop. The exit node decrypts the last layer and sends the original packets to the Internet. Responses travel back through the same circuit.

In addition to regular WWW, Tor can be used to access so called *hidden services*. How are they different from browsing regular WWW sites?

Tor *hidden services* are services that reside on the Tor network. They are different in the sense that they can't be accessed without going through the Tor network. Tor provides responder anonymity through a rendezvous mechanism. A Tor client can advertise that it hosts a service through a number of Onion routers that it chooses as *introductory points*. Another client can access them by finding an Onion router to work as their *rendezvous point*, and informing one of the introductory points about this rendezvous point. The former client can then contact the latter through their rendezvous point to provide the service. Location of both parties remains hidden from the other.

Tor provides anonymity. Does it also mean security? How can anonymity get compromised on Tor?

Tor uses a scheme similar to Chaum's mix-net for disguising the origin of a request, and uses onion encryption to hide the content of the messages from all nodes other than the exit node. This means that Tor does not provide end-to-end encryption, but only encrypts up until the exit node. If the traffic is not secured additionally by other means, the exit node is able to read (and alter) whatever was in that traffic. Tor provides security from observers close to the Tor client by encrypting the first hops, but there is no guarantee that the exit node or hosts in their immediate vicinity are any more trustworthy.

Tor does not filter the traffic passing through it. Browser cookies, host/user certificates, user typing in their credentials, applications doing things both inside and outside Tor etc. can still leak the identity of the user to the providers of the services that they access. The only thing that Tor hides is where the device is on the Internet.