# Methodology

In order to accomplish our goal we have designed a messaging protocol that will provide anonymity through the use of a distributed network of peers and multiple layers of encryption. In order to test our design we have implemented a test platform in C for Linux. This test application will allow us to perform testing on our design to ensure that it meets our goals and does not have a huge performance overhead. The details of our design and testing methods will be discussed in detail below.

## Messaging Protocol

The Anonychat messaging protocol consists of a set of distributed clients that act as peers in transmitting messages, as well as a centralized name server to inform clients of available peers. These clients will send messages amongst each other using controlled flooding, encryption between the source and destination peers, and encryption between each intermediate peer. In order for this to work, we needed a way to ensure that each peer has a connection to all of the other peers. This is the role of our name server; it distributes peer lists to each peer, and make sure that they can all communicate.

### Anonychat Name Server

The Anonychat Name Server (ANS) is responsible for creating a network of peers that provides each peer with a route to each other. Our protocol does not require a specific implementation for creating this network, but there are certain conditions that must remain true. The peer network must be changed periodically, allowing peers to send messages on different paths. The peer network must assign peers in a random fashion; it is recommended that peers with similar IP address prefixes not be assigned to each other. The peer network must not isolate any peers or group of peers; they must all form a small-world graph. The ANS must also maintain connection state on all peers, and be able to react accordingly when a peer drops from or joins into the network.

In our test implementation we maintained state by requiring each peer to uphold a TCP connection with the server. This allowed us both to determine when a peer has connected/disconnected and to send out peer updates to the peer when required. Our node network was implemented using a connection graph, which is detailed below:

1. An internal list of clients is re-arranged to occur in a random order. This should ensure that connections appear random as well for the following steps.
2. If there are more than five nodes connected, each of the nodes is arranged in a ring, and is connected to the two nodes ahead and behind it in the ring as well as the node directly across from it. Otherwise all of the nodes are simply connected to all of the other nodes. These connections are bidirectional, meaning that if node A is able to send to node B, node B is also able to send to node A. This situation is incidental, however and should not be assumed by the clients to always be true.
3. When the central server’s graph is completed, a message is sent to every node informing it of its assigned peers. No node is informed of who is able to send messages to it.
4. This cycle will repeat either (1) every time a client joins or leaves if the number of clients in the network is less than a number set by the ANS, or (2) every 30 seconds if there is a larger number of clients. The second case implies that, in larger networks, new clients may have to wait up to 30 seconds before it is able to broadcast messages, but this should also keep the name server from flooding the network. Messages in transit when connections are re-assigned behave as if nothing happened.

### Anonychat Client

Our protocol implements two levels of encryption, message level and peer level. Both levels of encryption will use asymmetric encryption, implemented by the RSA algorithm. The first level of encryption, between the destination and source, guarantees that the destination knows the message is for itself. The source does not know any information about the destination other than its public key, and the fact that no one else can read the message. At this level of encryption keys are maintained by the user. The second level of encryption, between each intermediate peer, ensures that the source and destination of the messages remain anonymous. This level also ensures that the signature of the message changes as it is sent across the network. However, the content of the message remains the same. This approach is similar to onion routing; because the message signature is constantly changing, it disallows the message from being traced through the network. The keys used at this level are hidden from users.

Peers communicate with each other via controlled flooding. Specifically, they communicate by broadcasting the message to all of their peers, and their peers will then broadcast the message to their peers and so on. To determine if a received message is new, a hash of the source/destination encrypted message can be compared against a hash of previously received messages. When a peer receives a message that it has never seen before, it will always broadcast the message to its peers, regardless of whether it was the intended destination or not.

The message protocol that used by our messaging platform is relatively simple. It consists of a ten character or less command with the body following, delimited by a space. All messages have a fixed size to prevent an adversary from extrapolating data about users from message size. In our implementation each message is padded with a string of random characters, before the peer level encryption. This ensures that the padding is not obvious and randomness will not create a unique signature on messages.

## Testing

To verify that our protocol is viable, has sufficient performance, and works as intended, we have run several tests on various attributes of the protocol. These tests include measuring message latency, latency introduced by encryption, time differences in successful and unsuccessful decryption, and the network utilization of our controlled flooding implementation. These experiments were run via a series of simulations on both a local machine and four computers spread out on different networks. The latter is a necessity to get accurate latency results.

### Latency

Given the indirect method our protocol uses to transmit messages, we expected to incur additional round trip time (RTT). However, as a metric for success we didn’t want this additional latency to be much bigger than the latency found in direct communications. To ensure our latency is acceptable, we set up test implementation on our four test machines. We then sent a series of ten messages between each machine using our messaging protocol, and recorded the RTT of each message. Next, we will sent another ten messages between the each of the four hosts themselves using ping, and again record the RTT of each message. After collecting the RTT of both messages, we compared them and verified that the RTT using our protocol is not greater than five times the direct RTT.

### Encryption Overhead

Encryption plays a big part in our messaging protocol, but it can also be a very expensive computation. For the sake of rapid communication, we didn’t want encryption to add a huge overhead to our protocol. To ensure this is not the case, we tested the time it takes for messages of various different sizes to be encrypted and then decrypted. We sent messages with lengths in a range starting from 20 to 1000, increasing by increments of 20 each iteration. Using these tests, we calculated the time each of these messages took to encrypt and decrypt. This will add additional latency to message sending, but we hoped it would be less than half of the network latency.

In addition to testing the overhead added by encryption, we also needed to test that the time difference between a successful and unsuccessful decryption is not noticeable. If this difference is noticeable, it could allow an adversary to determine if a particular node was the intended recipient of a message. To do this we performed a similar test to our total encryption/decryption overhead test mentioned above. However, this time we decrypted the message twice: once with the correct private key, and the second time with an invalid private key. Comparing these values allow us to determine if a successful decryption is noticeable.

### Network Utilization

The network utilization of our protocol also needed to be tested. We don’t want our protocol to use a vast majority of the available bandwidth by flooding messages across the network. In order to test our utilization we ran a simulation with around 25 clients running locally. The simulation lasted for five minutes with each client periodically sending messages to other clients. To do this, each client picked a random time from five to 35 seconds. The client then sent a message after this time expired and repeated the process. As this occurred we analyzed the network using Wireshark and took notes of how many packets were in the network for 5 second intervals. We also ran a five minute simulation of normal computer use, such as web browsing, and heavy network use such as a file transfer. This allowed us to compare our protocol’s network usage against these two, giving us a sense of how our network utilization compares to normal and heavy network use.

### Scalability

As a chat program could possibly be used by different amounts of people at a time, it seemed paramount we test how well our messaging protocol will scale. To do this we performed a test similar to the network utilization test; however, instead of analyzing the network, we analyzed the resource usage of each node and the ANS. During this simulation we recorded the number of messages each node processed, and the processing time of each of those messages. We then recorded the RTT of each message to determine the effect more clients have on it. We also analyzed the ANS, measuring how many times it reshuffles the peer list and how long each of those peer reshuffles took. Unlike the network utilization test, we ran this test on a range of clients from six to 18, increasing by six clients each time. This gave us a clear measure of how well our protocol scales with varying numbers of clients.

### Anonymity

The final test that we performed is the anonymity test of the source of our messages. To do this, we ran a simulation with around ten clients sending messages back and forth to each other at varying intervals. We used Wireshark to examine the sent packets, and took note of whether the packets were encrypted as well as confirmed there were no identical packets sent between peers. This helped us tell if it is possible to determine the source of a particular message by analyzing the packets sent between nodes.