CS 6650 - Scalable Distributed Systems

Mahvash Maghrabi

Homework # 2

Late Days Left : 5

II. **Answer the following questions using explanations and diagrams as needed. No implementation needed**

1. **Consider a basic client-server program you implemented using socket programming. How does RPC improve the same? And how does Java RMI improve it? Why do we need RPC and RMI and what value are they adding to this client-server interprocess communication?**

   Remote Procedure Call (RPC) is a protocol through which a computer program can request a service from another computer located on a network, without the need to understand the network's details of the other computer. RPC uses the client server model for requesting and providing services. It improves the basic client-server program in such a way that the client can call a function on the server like it is a local function which makes it easier for the client to communicate with the server and the client is not required to know the details of the network communication. Due to this the client can simply call the function on the server and the Remote Procedure Call will handle the communication details.

   Remote Method Invocation (RMI) is a Java API through which an object can invoke methods on another object that is present on another java virtual machine) on the network. It can invoke methods like they were local methods to the object. It is a java specific implementation of RPC that provides additional functionality such as security and serialization. It allows transparent invocation which does not require the client to know that it is communicating with a remote object which makes it easier to write client code, and allows a better interaction between the client and the server objects. It also provides automatic support for distributed garbage collection which allows for the client and server to release resources when they are no longer needed.

   RPC and RMI are needed as they provide different processes to communicate with each other even if they are executed on different machines they also add value to client-server interprocess communication as the process does not need to worry about the network and the programmer can focus on the logic of the application. They also provide a higher-level programming model making it easier to create and maintain distributed systems and they also provide additional functionality such as security and serialization.

2. **What are the main arguments for adopting a super node approach in Skype?**

   For adopting a super node approach in Skype the main arguments are :

   1. **Scalability:** Using the super node approach will allow for a distributed network that can handle larger numbers of users as well as more connections as it will distribute the workload across multiple nodes.

2. **Reliability:** A super node approach can also improve the reliability of the network by distributing the load across multiple nodes and there will be less chances of failure. If one super node becomes unavailable the workload can be handed over to another node. The system will function without any issues and will provide consistent results.

3. **Efficiency:** A super node can also act as a relay which means that they can forward messages and connections on behalf of other nodes in the network which improves the efficiency of communication and messaging as it reduces the number of hops the message has to take to reach the destination thereby improving the speed of delivery of messages. It also caches the results which again improves the performance.

4. **Flexibility:** Super nodes can also be used to perform efficient search of the global index of users. The global index of users refers to a database or directory of all the registered users in the system like Skype here. When a user starts searching for another user on Skype the search request is sent to a super node which then searches across other super nodes until the requested user is found. This allows for efficient and fast searching in the entire database.

5. **Bypassing Firewalls:** Sometimes connection requests are blocked in these cases the super nodes are used to establish a connection by using intermediary nodes which allows the clients to bypass the firewall and establish a connection, even if their connection requests are blocked.

6. **Authentication:** Authentication makes sure that only users that are authorized have access to the system and unauthorized users do not have access for security and privacy reasons. In Skype the users are required to authenticate themselves using a well-known login, username and password only then they can make contact with a selected super node.

3. Outline an implementation for the Election service that ensures that its records remain consistent when it is accessed concurrently by multiple clients.

An implementation for the Election service would be to create a java vector object to store the votes. The vector will hold the votes in the form {String vote, int number}. A method that will check if a given voter number has already voted can be implemented. The method can iterate through the vector and check if the vote number already exists. The synchronized keyword in Java which locks a specific method or code so that only one thread can execute at a time can be used. The Vector class in Java has a built-in synchronization mechanism so that multiple threads cannot access and update the vector simultaneously and that will prevent conflicts between concurrent clients. A method that can append a vote to the Vector has to be implemented this method can first check if the voter number has already voted using the previous method and if not the method will append the vote to the Vector. A method to count the votes and announce the results has to implemented.This method can iterate through the Vector and count the number of votes for each voter. A monitoring and alert system to detect and resolve any potential issues or failures within the service should also be implemented.

4. A client makes remote method invocations to a server. The client takes 5 milliseconds to compute the arguments for each request, and the server takes 10 milliseconds to process each request. The local operating system processing time for each send or receive operation is 0.5 milliseconds,

and the network time to transmit each request or reply message is 3 milliseconds. Marshalling or unmarshalling takes 0.5 milliseconds per message.

Calculate the time taken by the client to generate and return from two requests:

(i) if it is single-threaded;

(ii) if it has two threads that can make requests concurrently on a single processor.

You can ignore context-switching times. Is there a need for asynchronous invocation if the client and server processes are threaded?

i) if it is single-threaded

Time per call = calc. arguments + marshal args + OS send time + message transmission +OS receive time + unmarshall args + execute server procedure + marshall results + OS send time + message transmission +OS receive time + unmarshal args
= 5+ 4*0.5 + 4*0.5 + +2*3 + 10 ms = 5+2+2+6+10 =25ms.
Time for two calls = 50 ms.

ii) two threads

calc. args + marshal args + OS send time (call 1) = 5+0.5+0.5 = 6
then calc args + marshal args + OS send time (call 2) = 6
6 + 6= 12 ms
message transmission + OS receive time + unmarshal args = 6+ 3+.5+.5 = 10 ms
Execute : 10 +10+1 =21 ms
21 + 10+1 = 32 ms
client : 3+1 = 4 then 32 +4 = 36 ms
Time for two calls = 36 ms.

III. Answer the following questions using explanations and diagrams as needed. No implementation needed

5.Suggest a design for a notification mailbox service that is intended to store notifications on behalf of multiple subscribers, allowing subscribers to specify when they require notifications to be delivered. Explain how subscribers that are not always active can make use of the service you describe. How will the service deals with subscribers that crash while they have delivery turned on?

A mailbox service is designed in such a manner that multiple users can access it and receive notifications whenever necessary. The service is as such the the users receive the notification even if they are not currently Online or using the service. The first step to using a service is it to register using unique credentials of an individual. When a user registers for a service the register() method is called and the service returns a registration object which is a reference for the subscriber to interact with the services here the mailbox.
The user can access the RemoteEventListener by using the registration object provided by the mailbox Service. The RemoteEventListener is like a bridge between the event generator and the user of the Service. It facilitates the flow of notifications from event generators to the user's device.
The service also gives the users the permission to turn the notification on or off depending on their way of using. Some users like to receive notifications while others do not hence they have the ability to do so and they do not Receive tons of notifications that might distract them. If the subscribers are crashed they can restore the registration object whenever they restart and then they continue to receive notifications.

6. Consider the version of the *FireAlarm* program written in JMS (Section 6.4.3). How would you extend the consumer to receive alarms only from a given location?

Java Messaging Service (JMS) is an API for messaging between two systems. It can be used to send and receive messages between applications. The key components of JMS are :

1. JMS client - It is a java program that produces or consumes messages
2. JMS producer - creates and produces messages
3. JMS consumer - receives and consumes messages
4. JMS provider - any system that implements the JMS specification
5. JMS message - used to communicate information between clients
6. JMS destination - is an object that supports indirect communication in JMS.

To extend the JMS consumer to receive alarms only from a given location one can use message selectors in JMS. A message selector allows us to filter messages based on both the header properties and the message properties. In the FireAlarm program one can set the header property or the message property for the location of the alarm. For instance one can set the header property called as ‘Location’ to specify the location of the alarm and in the Consumer one can use the message selector and only messages from the specified location will be received.

7. Explain in which respects DSM is suitable or unsuitable for client-server systems.

Distributed Shared Memory (DSM) allows the processes that are being run on different computers to use And also share the data like they have a common physical memory. There will be no need to write passing explicit messages which will in turn make it easier for the programmers to program any distributed system. It also makes sure that any changes or any changes made by one computer are also made by others although they share a different physical memory. It appears as a single shared memory space to all connected clients in a network.

The DSM can be suitable for client-server system in the following ways :

Easy Programming : The programmers can easily program distributes systems as they do not have to think about The local memory of individual systems. The applications can easily access and share the data like they are Present in a single memory space.

Scalability : With the use of DSM there can be larger shared memory compared to a single system which will allow for more scalability.

Performance : There will be lesser burden on a single system as the data will be distributed across multiple Systems.

The DSM can be unsuitable for client-server system in the following ways :

Complexity : Having a DSM comes with large amount of coordination and management which will increase the Complexity.

Performance Overhead : As the complexity increases there will be additional communication and coordination which will lead to performance overhead.

Limited consistency : There is also difficulty to make sure that all the nodes have a consistent view of the shared memory at any given time.

Hence DSM is suitable for parallel applications as the programmer does not need to worry about message passing and the processes are like a single shared memory. DSM is less appropriate in client-server systems if the resources are viewed as abstract data and a request is made to access them as DSM requires sharing directly. The efficiency of DSM depends on the number of computers as the quality decrease as the number increases

8. **Compare the request-reply protocol as described in Section 5.2 with the implementation of client-server communication in SOAP. State two reasons why the use of asynchronous messages by SOAP is more appropriate over the Internet. To what extent does the use of HTTP by SOAP reduce the difference between the two approaches?** 9.1

A request - reply protocol is a communication protocol in this protocol one process sends a message Which is a request to another process and it then waits for the reply. The process to message or request again is blocked until a reply is received and then it can continue. This protocol is often used in client-server communication.
The Simple Object Access Protocol is a protocol for interchanging information in web services.
In this protocol a client sends a message to a server asking for a service or asking some information and the server replies with the desired information or service that was asked for. The SOAP message is encoded in a XML format and can be transferred over different networks like HTTP SMTP etc. The exchange of information or the communication between the server and the client takes place in an XML format. This allows exchange of data in a standardized way. The SOAP method is more complicated as compared to the simple request-reply protocol. Reasons the use of asynchronous messages by SOAP are more appropriate are :

1. The communication is asynchronous which allows the client and server to handle multiple requests and responses which improves the scalability
2. There is reliability in the SOAP messaging protocol.
3. Better error handling as both the server and client send as well as receive messages independently
4. There are no timeout risks involved as the communication can take place even if any one the client or the server is not available temporarily.

HTTP by SOAP reduce the difference between the two approaches to a certain extent as the messaging pattern followed is still similar to the request-reply protocol where the client sends a request message to the server and the server sends a reply to the client. HTTP provides the basic infrastructure for communication. HTTP gives a standardized way for transmitting data over the internet and SOAP uses HTTP as the protocol to interchange messages between the client and server. Hence SOAP has many features as well as the benefits of HTTP which includes the reliability, security, and support for both the synchronous as well as the asynchronous communication.

9. Explain the role of a servlet container in the deploying of a web service and the execution of a client request.

A servlet container is a component in a web server and it provides a runtime environment where the Java servlets can be executed. The servlet container is responsible for initializing the servlets and also executes them when the client requests also and manages the lifecycle of the servlet. It also simplifies the development of web applications and allows developers to focus on the programming logic of their application.

A servlet container is like a runtime environment for executing the servlets and deploying web services. It provides an environment that enables the web service to respond to client requests. When the client makes the request the servlet container receives it and then makes an instance if there isn’t one already and gives the request to the servlet for processing. The request is answered and the response is given back via the servlet container. When the servlet processes the request and gives back a reply the skeleton translates back into a SOAP reply.

The servlet container also gives a management interface. The management interface shows the names of servlets that have been deployed and also provides operations to manage and access the information about every service provided. When the servlet is stopped the variables get reset whenever it is restarted. The servlet container also provides access to the service description that is expressed in XML notation. This way the servlet container plays an important role in deploying and also executing the web service and it enables the client to handle requests and provides a response.