

CSC547 Homework Assignment #2

Akruti Sinha (@asinha6) || Priya Andurkar (@pandurk)

Problem 1:

Problem 3.6:

Consider the internet, IP, its layer 3 protocol, and two layer 4 protocols, namely TCP and UDP. The system in the definition of virtualization we gave in Section 3.1.1 is the internet.

1. Explain how TCP virtualizes the internet. More specifically, what is the property TCP gives to the internet, that the internet does not (natively) have?
2. Describe, in any detail you like, what mechanisms TCP uses in order to achieve the property.
3. Who (not which hat) benefits from this virtualization?
4. Which hat benefits from this virtualization?
5. Which hat pays the price?
6. Explain why UDP does not virtualize the internet like TCP does.
7. Search “what protocols use UDP” and “what protocols use TCP”. Explain why some protocols don’t care to use an internet full of virtues...

Solution 1:

1. As per the definition in the textbook, we know that virtualization is the process of giving an additional property to a system that it does not have. The internet, inherently, does not have the property of reliability. This drawback is overcome by TCP (Transmission Control Protocol) which adds reliability and connection-oriented communication. Following is how TCP provides virtualization:
 - a. Acknowledgement and retransmission: An acknowledgement is sent back by the recipient to the sender. If not received, the sender retransmits data.
 - b. Packet Sequencing: TCP ensures that the data packets are numbered so that the recipient can reconstruct the data correctly.
2. The following are the details on the above provided properties:
 - a. The data from the sender is broken down into segments. The segment received at the receiver is checked for errors using checksum, to make sure of the integrity of data. If the segment is error free and in correct order, the receiver would send an acknowledgement to the sender with the header set to indicate the next segment that needs to be sent. On the sender’s end, if an acknowledgment is not received within a timeout period, it would retransmit the segment. This mechanism ensures the successful receipt of data packets.
 - b. A sequence number in the packet header is added to identify the position of the data within the overall stream of data that is being transmitted. Buffers are maintained at the sender’s and receiver’s end to deconstruct the data and reconstruct the data respectively. With the help of acknowledgements being sent for all the packets being delivered, TCP ensures that an error free and complete message is received by the recipient.
3. The users of the applications that adopt the TCP protocol would benefit from this. For example, a user waiting for a message on an application that does not use TCP, there is a possibility that they receive a message that is incorrect or incomplete.

4. This virtualization benefits the Cloud Architects, Cloud Application developers from the cloud consumers perspective.
5. The price is paid by the network engineers who need to implement these features in the protocols from the cloud provider's side.
6. UDP (User Datagram Protocol) fails to virtualize the internet like TCP for the following reasons:
 - a. It does not establish any connection between the sender and receiver before sending data which reduces the guarantee delivering the data. This makes the protocol unreliable in nature.
 - b. There is no sequencing mechanism that is maintained in UDP which means that each packet is sent independently. This can pose a problem where data needs to be received in a specific sequence.
 - c. Unlike TCP, UDP has no error recovery. Hence, there is no automatic mechanism to detect if the data being sent is corrupt. It further does not help with retransmission of corrupt packets either. This also results in unreliability.
7. UDP users:
 - a. DNS (Domain Name System): This protocol is used for translating domain names which are in human-readable formats to IP addresses. For the sake of simplicity of the query/response mechanism the DNS protocol uses UDP. It proves to be faster and more efficient in this use case.
 - b. Online Gaming: Many multiplayer games utilize the UDP protocol, to achieve faster communication between users. Here the use of quick data transmission is considered more important and traded off with occasional packet loss.
 - c. TFTP (Trivial File Transfer Protocol): It is a protocol with the underlying usage of UDP, due to its low overhead. UDP helps in achieving lightweight files like configuration files.

TCP users:

- a. SSH (Secure Shell): TCP is important for establishing secure connections to the remote login and command executions.
- b. IMAP (Internet Message Access Protocol): It is an email protocol which is important for accessing and managing emails on a remote server. It relies on TCP for reliable communication.
- c. HTTP (Hypertext Transfer Protocol): The TCP is an important underlying protocol for the HTTP which is used extensively across the world wide web. It leverages the reliability, security, and many more features to provide services to the end users.

Problem 2:

Problem 3.2:

The abstract concept of service. Consider the bus service at NCSU.

1. Describe the service in your own words.
2. Who is the provider?
3. Who is the consumer?
4. What are the benefits to the provider?
5. What are the benefits to the consumer?
6. How does the service get discovered by the consumer?

7. How does the provider “advertise” the service?

Solution 2:

1. The Wolfline bus service is a service provided by NCSU (the provider) to its students and workers (the receiver) that allows the receiver to go from one point to another on or near the NCSU campus. The bus stations are intentionally located to maximize the number of individuals who can use the service. The frequency of buses on this service varies depending on the amount of traffic predicted (greater frequency on more popular routes; presence/absence of major routes on weekends, etc.).
2. The provider is NC State.
3. The consumers are the staff members and students at NC State.
4. The benefits to the provider include:
 - a. a reduction in the number of vehicle accidents on or around campus.
 - b. Improved student safety, particularly at night.
 - c. Reduce greenhouse gas emissions on campus.
 - d. Increased lecture attendance, which is associated to higher grades, which is related to more lucrative job offers, which is related to the institution's reputation, which is related to more money for the university.
5. The consumer advantages include:
 - a. The elimination or reduction of the demand for private transportation.
 - b. Avoidance of maintenance and other hassles connected with owning an own vehicle. These expenses are passed on to the provider.
 - c. Removal of the requirement to obtain a driver's license. This need is passed on to the company's hired drivers.
6. The service is discovered via the TransLoc mobile app or Google Maps under the Buses tab.
7. The provider advertises the services through the following methods:
 - a. Logos on buses: For a completely new and ignorant NCSU student and employee, this is the first indication that such a service exists.
 - b. Wolfline bus stop signs: These signs inform riders that an NCSU bus stopping nearby may be able to take them close to their destination for free. These signs frequently include QR codes that direct users to the TransLoc smartphone application.
 - c. LED bus boards: These boards indicate the route number as well as the main destination of that route (usually popular destinations such as the Hunt Library, Wilson College, and so on).

Problem 3:

Problem 3.5

The concept of microservice. Consider the banking application of servicing ATM transactions.

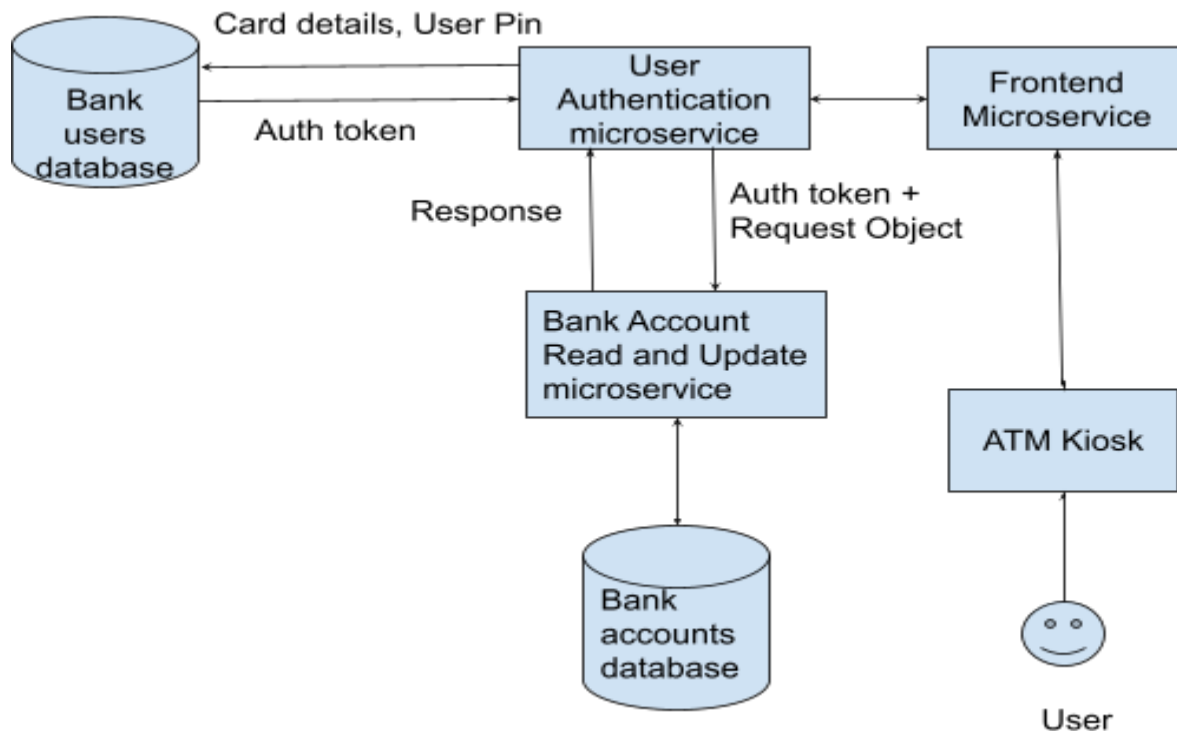
1. Describe the logic that this application has to implement.
2. Define a set of microservices that can be used to implement this logic.

3. Can you split one of these microservices into at least two “smaller” ones?
4. Define interfaces for each microservice.
5. Supply a diagram that depicts “who talks to whom”.

Answer 3.5

1. A banking application for servicing ATM transactions would have multiple functionalities and their logic can be as follows:
 - a. Authenticating user: Once the user swipes their card, a user should be authenticated with their card pin number. A valid pin entry would enable the user to access the subsequent services, else they would need to re-enter the pin.
 - b. Providing service menu: The user should be able to see all the services available to them in the form of a menu to select from. Services could include, cash withdrawal, cash deposit, check deposit, balance inquiry, etc.
 - c. Cash withdrawal: Once the user gives the input of the total amount of money that needs to be retrieved, the application must check the existing balance. If the balance is greater than or equal to the requested amount, the ATM should disburse the amount else throw an error message.
 - d. Cash deposit: Once the user is authenticated and deposits the money into the kiosk; the application must count the amount, specify the denominations, and after verification from the user add the amount to their total balance.
 - e. Check deposit: Once the user is authenticated and deposits the check into the kiosk; confirm the validity of the check, deduct the amount from the sender's account, and add the amount to the recipient's account.
 - f. Balance enquiry: Once the user is authenticated; the application must be able to return the total available balance.
 - g. Transaction enquiry: Once the user is authenticated; the application must be able to retrieve the last 10 transactions of the customer.
2. The above mentioned logic can be implemented with the help of the following microservices:
 - a. Front-end microservice: this microservice would be responsible for providing a user interface and rendering all the information on the screen of the application for a user to see and interact with.
 - b. Authentication microservice: this microservice would be responsible for reading the card details and authenticating the user against the card pin number entered by the user. It would be a gateway application between the user interface and the bank database. It would return a unique identification token which should be subsequently used by all the other API calls to the DB.
 - c. Bank database microservices: After the user has been successfully authenticated, these set of RESTful endpoints would be used for connecting to the bank database specifically for querying user data, user account information and overview, transaction history, and more. It can also be used to add or update data in the bank database such as, deduction of balance after withdrawals or addition of balance after deposits.
3. For the microservice which validates a check and deposits the money into the account of the user, we can further break it down as follows:
 - a. One microservice can be used to validate the correctness of a check on the basis of the unique check number, validating the signature of the issuer, checking if the

- check is expired, and more. This separate microservice can be made extremely robust to validate outstanding checks from any issuer bank.
- b. The other microservice can take the validation output from the above-mentioned microservice and subsequently proceed to deduct the amount from the payer's account and add the amount to the recipient's account.
 4. The interfaces for each microservices for the above mentioned application would be as follows:
 - a. HTTP/ REST apis
 - b. Database Interface
 - c. Authentication Interface
 - d. Graphical User Interface
 - 5.



Problem 4:
Problem 3.9

?aaS. The following are offerings of services. The question is what kind? Justify your answer. (Google the name, if you are not familiar with the company that offers the service.)

1. Overleaf?
2. Dropbox?
3. Google docs?
4. Gmail?
5. Netflix?

Solution 4:

1. **Overleaf** is a Software as a Service (SaaS).
 - a. "Overleaf is a collaborative cloud-based LaTeX editor used for writing, editing, and publishing scientific documents," according to Wikipedia. It is, at its core, a software for editing LaTeX texts.
 - b. This is also consistent with the NIST definition of SaaS: "the capability provided to users." The consumer will use the provider's cloud-based infrastructure."
2. **Dropbox** can be classified as SaaS or IaaS (Software, Infrastructure), depending on the usage.
 - a. Dropbox can be used to store, share, and collaborate on files. Dropbox is classified as SaaS because it is an end-user service for file storage and access management.
 - b. However, Dropbox can also be classified as IaaS because their APIs can be used as storage and/or backup services as part of other application stacks (<https://www.dropbox.com/developers>).
3. **Google Docs** is a SaaS application.
 - a. Google Docs allows users to create and modify documents online while collaborating in real time with other people. Spell checking, autocomplete, grammar correction, revision history tracking, and more capabilities are also available.
 - b. Google Docs, like Overleaf, is a document editing program at its heart, but for a broader range of applications. This is also consistent with the NIST definition of SaaS.
4. **Gmail** is a SaaS.
 - a. Gmail provides electronic mail exchange software as a service through the internet, and they are responsible for the infrastructure, maintenance, and enhancement.
 - b. Because customers exclusively use Gmail's services, it adheres to the NIST definition of SaaS once more.
5. **Netflix** can be classified as a SaaS
 - a. It is because it offers subscriptions to their software to watch licensed content on demand.
 - b. It could be argued that users are actually paying for the content, but they are receiving personal recommendations and on demand content (some of which are proprietary to Netflix) curated and accessible by the software provided by Netflix.

Problem 5:
Problem 3.19

Burstable compute instances. Search the AWS documentation about burstable compute instances.

1. Describe the key concepts of this service.
2. State succinctly what is the benefit for the consumer.
3. State succinctly what is the benefit for the provider.

Solution 5:

1. Burstable compute instances are a type of compute instance supplied by AWS for efficient workload management and to improve CPU utilization.
 - a. The key issue addressed by AWS's Burstable compute instances is underutilization of CPUs, resulting in CPU cycle waste.
 - b. According to AWS, "many general purpose workloads are not busy on average, and do not require a high level of sustained CPU performance" (<https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/burstable-performance-instances.html>).
 - c. Burstable compute instances are "T instances" having a baseline performance identical to other "T instances" but the ability to "burst" over this baseline at any time and for as long as necessary.
2. The baseline CPU is already defined to meet the needs of the majority of general purpose workloads
 - a. The T instances offer a balance of compute, memory, and network resources AWS also claims that *"They can save you up to 15% in costs when compared to M instances, and can lead to even more cost savings with smaller, more economical instance sizes, offering as low as 2 vCPUs and 0.5 GiB of memory"*.
 - b. In brief, consumers can get the majority of their needs addressed at a reduced cost, as well as the flexibility to handle "bursts" in CPU consumption.
3. The benefits for the provider are:
 - a. Burstable compute instances enable AWS to oversell server hardware while preserving performance.
 - b. They can reduce resource waste; and
 - c. They can acquire and supply for more tenants at lower costs.

Bonus Problems:

Problem 7:
Problem 3.8

Feedback-based control. Give an example from your personal, college life in which feedback was used to achieve an objective. Describe the problem using the terminology of the MAPE framework we presented in Figure 3.1, page 57. More specifically:

1. State the objective clearly.
2. What is the managed element?
3. What is the sensor and actuator?
4. What analysis is done? Where? Who did the analysis?
5. How the results of the analysis were communicated to the actuator?
6. Discuss the benefits, if any.
7. Discuss the challenges, if any

Solution 7:

1. The goal of my research project was to improve the energy usage of a smart home automation system while preserving user comfort and convenience.
2. In this case, the managed element was the full smart home automation system, including sensors, actuators, and the central control unit.
3. Sensors: The sensors in the smart home included devices such as motion detectors, temperature sensors, and light sensors. These sensors gathered information on occupancy, temperature, and lighting conditions. Smart thermostats, lighting controls, and HVAC systems were among the actuators, which could be programmed to change temperature, lighting, and other factors.
4. Monitoring data from sensors was used to determine occupancy patterns, temperature variations, and illumination preferences. This data was analyzed using algorithms developed by our research team to make energy optimization decisions.
5. The analytical results were conveyed to the actuators via the central control unit. Based on the collected data and optimization targets, this machine used algorithms to alter the settings of heating, cooling, and lighting systems in real-time.
6. The benefits are:
 - a. Energy Efficiency: We achieved significant energy savings without sacrificing user comfort by continuously monitoring sensor data and changing smart home systems.
 - b. Cost Savings: Lowering energy use resulted in cheaper utility costs, which is helpful to homeowners.
 - c. Reduced energy use adds to a lower carbon footprint, supporting environmental sustainability.
7. The challenges are:
 - a. Complexity: Developing and fine-tuning real-time energy optimization algorithms was a difficult undertaking.
 - b. Handling and processing massive amounts of sensor data necessitated effective data management and processing techniques.

Problem 8:

Problem 3.20:

Service pricing. Find out the cost per hour for using any type of compute instance in any cloud provider.

1. Find out the cost of buying a (desktop, laptop) computer with similar features.
2. Calculate when a consumer will break even. State your assumptions clearly.

Solution 8:

For this solution, we are assuming that a student needs to deploy and run a data-intensive RESTful API application on a compute server. We are assuming the following for the application to run successfully:

- A. Needs 10GB RAM
- B. Application sees a daily spike for 2 hours of the day
- C. Minimum number of instances for workload: 1
- D. Maximum number of instances for workload: 2
- E. Needs a machine with at least 8 core CPU

Based on these needs, we decided to use the AWS EC2 compute instance for deploying and running this application. The following are the configuration for the EC2 instance:

- A. Instance Family: General Purpose t3.nano
- B. RAM: 10.5 GiB
- C. vCPUs: 8
- D. Minimum instances: 1
- E. Maximum instances: 2
- F. Tenancy: Dedicated Instances

With these specifications, the AWS cost estimator gives us the estimate that the hourly usage would cost 0.0052 dollars/hour.

1. Let us consider the Macbook Air M1 chip laptop. It contains a 8-Core CPU and 16GB RAM. The one-time cost of buying this laptop would be 1199.0 USD.
2. To break even with the cost provided by AWS, the student would have to run his application on AWS for approximately $(1199/0.0052)$ hours. This would be close to 230576.92 hours. This would be close to 9608 days.