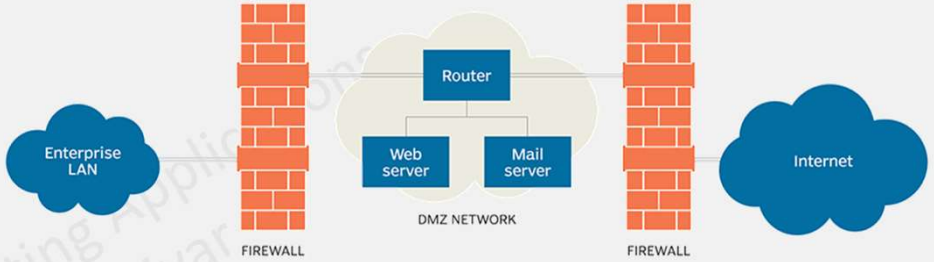
CLOUD COMPUTING APPLICATIONS

VPC: Subnets

Prof. Reza Farivar

Subnets

- Create subnets to isolate resources per the project requirement
  - DMZ/Proxy
  - Load balancer
  - web applications
  - Mail servers
  - Databases



- E.g. have a public subnet to host internet-facing resources and a private subnet for databases that accept web requests
- Create multiple subnets (public or private) in multiple AZs to host a high availability multi-AZ infrastructure and avoid a single point of failure
  - each subnet can communicate with every other subnet in the same VPC

Private Subnets

- Any incoming traffic from the internet cannot directly access the resources within a private subnet
- Outgoing traffic from a private subnet cannot directly access the internet
  - Restricted; or
  - Routed through a NAT
- Each resource (instance) gets a private IP
  - From the CIDR range associated with the subnet
- Technically, a subnet is private if there is no route in the routing table to an internet gateway

Public Subnet

- A subnet that has access to an internet gateway defined in the routing table
- Each resource in a public subnet gets a private IP within the CIDR range, AND a public IP accessible from the internet
  - the public IP can be dynamic (only remains valid while the instance is alive, and then AWS reclaims it), or
  - It can be an elastic IP, where you pay for it, and it will remain yours even if the instance shuts down
- Outgoing traffic can directly access internet
  - Unlike Private, which needs a NAT to access internet

Route Tables

- Each VPC has an associated “Main Route Table”
  - The Default VPC has a route in its “Main Route Table” to an internet gateway
  - Custom VPCs usually only have the local route in the MRT
- Subnets can have their own custom route table
  - If no custom route table is explicitly associated with a subnet, then it is associated with the VPC's main route table
- Possible route table targets
  - Local, IGW, a NAT device, A VGW, a peering connection, or a VPC endpoint (e.g. S3)
- Selection of the optimum route for network traffic is done based on the longest prefix match
  - the most specific routes that match the network traffic