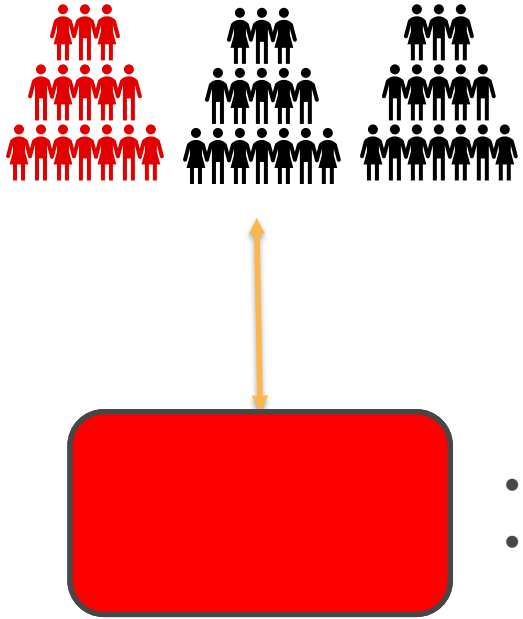
Elastic Load Balancing

Classic, Application, Network

Chandra Lingam

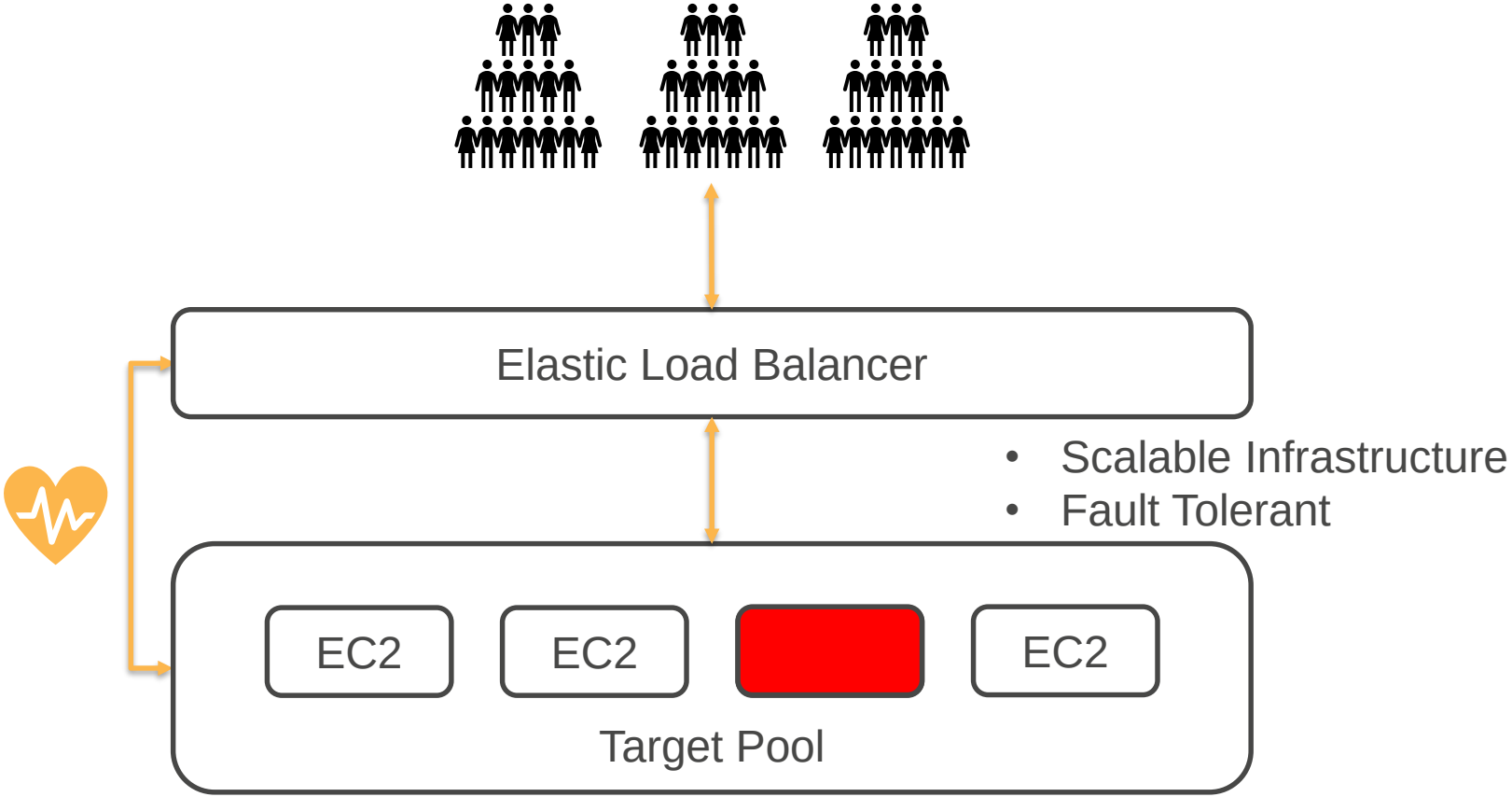
Cloud Wave LLC

Elastic Load Balancing Motivation

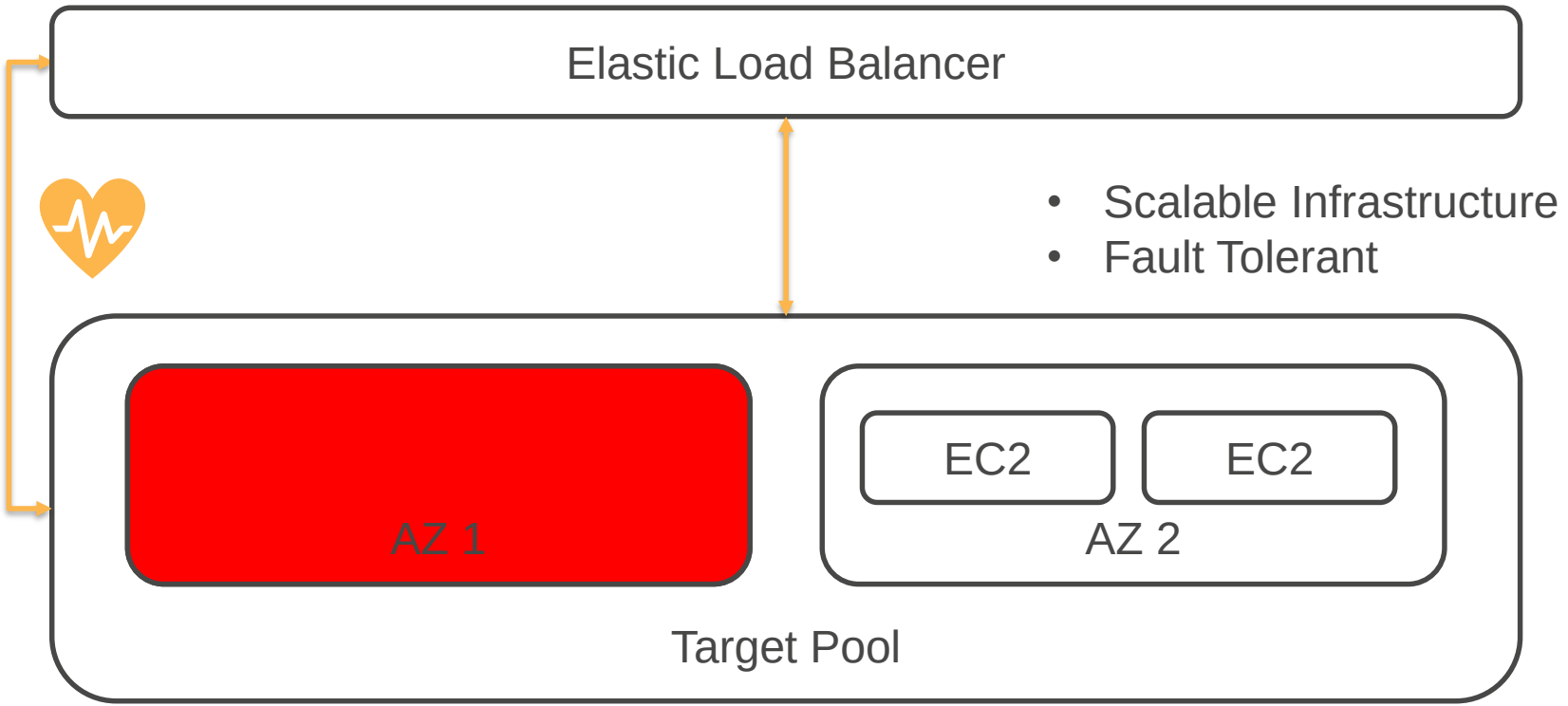


- Scalability Challenges
- Single point of failure

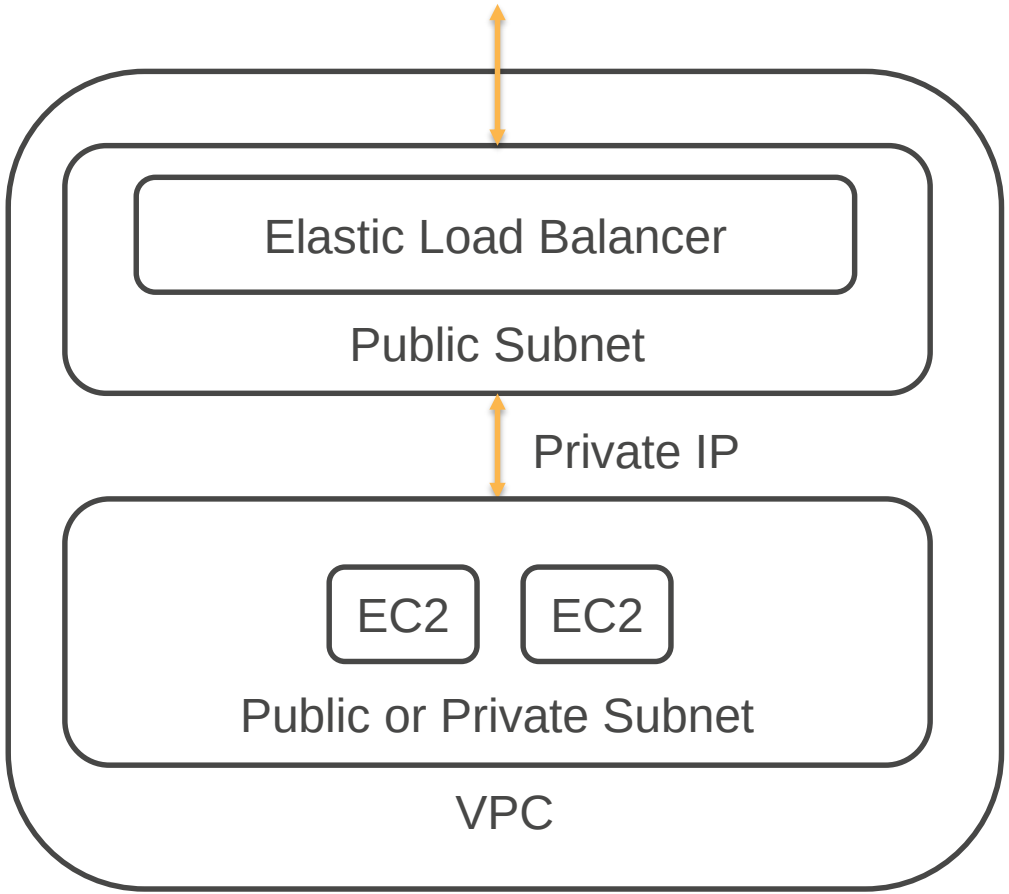
Elastic Load Balancing



Elastic Load Balancing

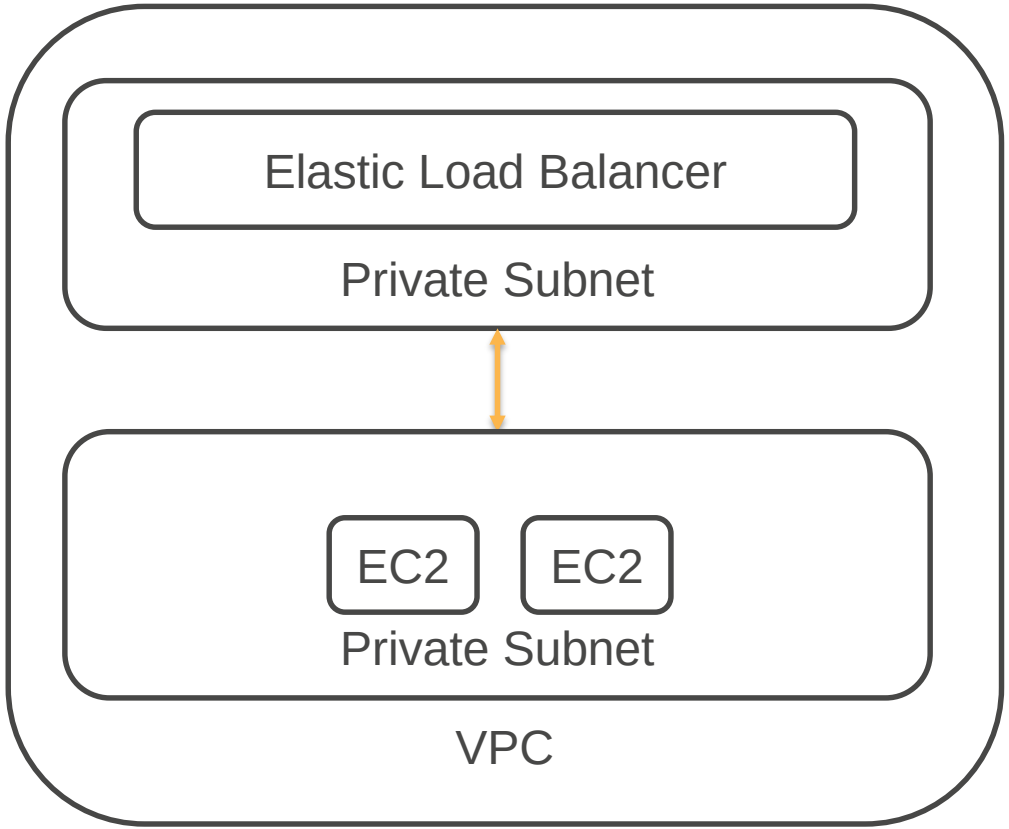


Elastic Load Balancing – Internet Facing



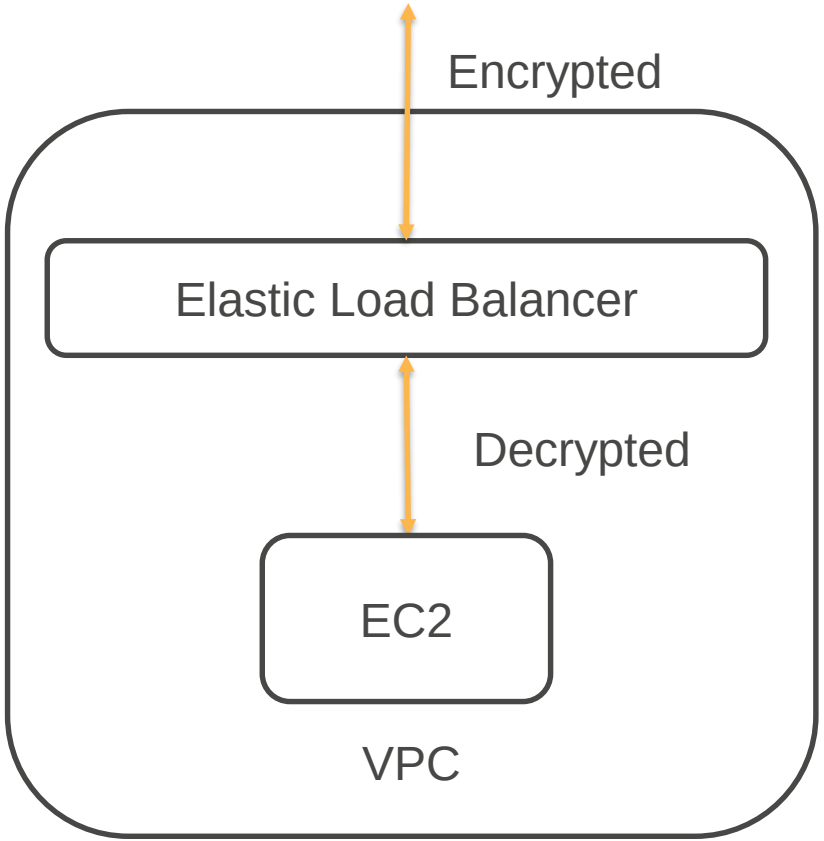
- Load Balancer is accessible from the internet
- Load Balancer talks to EC2 instance using Private IP
- EC2 instances can be in public or private subnet
- Reduces attack surface – EC2 instance configured only for private traffic
- DDOS Protection

Elastic Load Balancing – Internal Facing



- Load Balancer is accessible only inside VPC

Elastic Load Balancing – Security



- Offload SSL/TLS
- Integrated Certificate Management
- User Authentication – Cognito (Application Load Balancer)
  - Internet Identity Providers
  - SAML
  - OpenID Connect

Features

CloudWatch Monitoring

- Real time monitoring of key metrics

Connection Draining

- When deregistering instance, allow in-flight requests to complete
- Default wait time is 5 minutes (300 seconds)
- After wait time elapses, instance is deregistered

Additional Concepts

Sticky Sessions

- Route requests from a client to same target
- Used for stateful application - servers cache user data
- Disabled by default

HTTP/2

- Multiple requests sent on the same connection
- Efficient use of network resources

Additional Concepts

WebSockets

- Long running TCP Connection
- Bi-directional
- Server to Client Push notification support

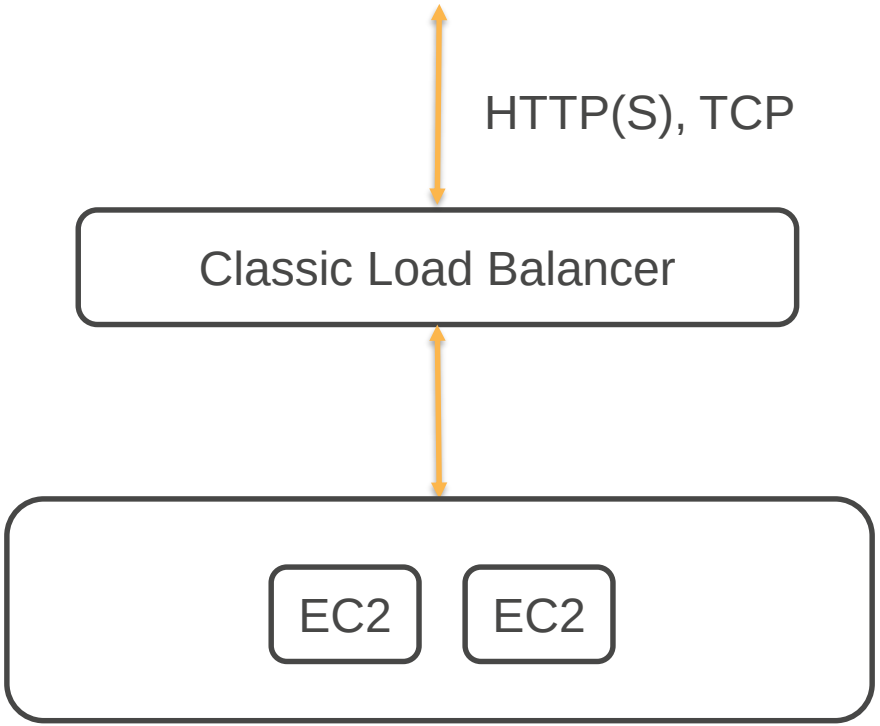
Cross Zone Load Balancing

- Enabled – distribute traffic evenly across all EC2 instances
- Disabled – distribute traffic evenly across availability zones

Load Balancer Types

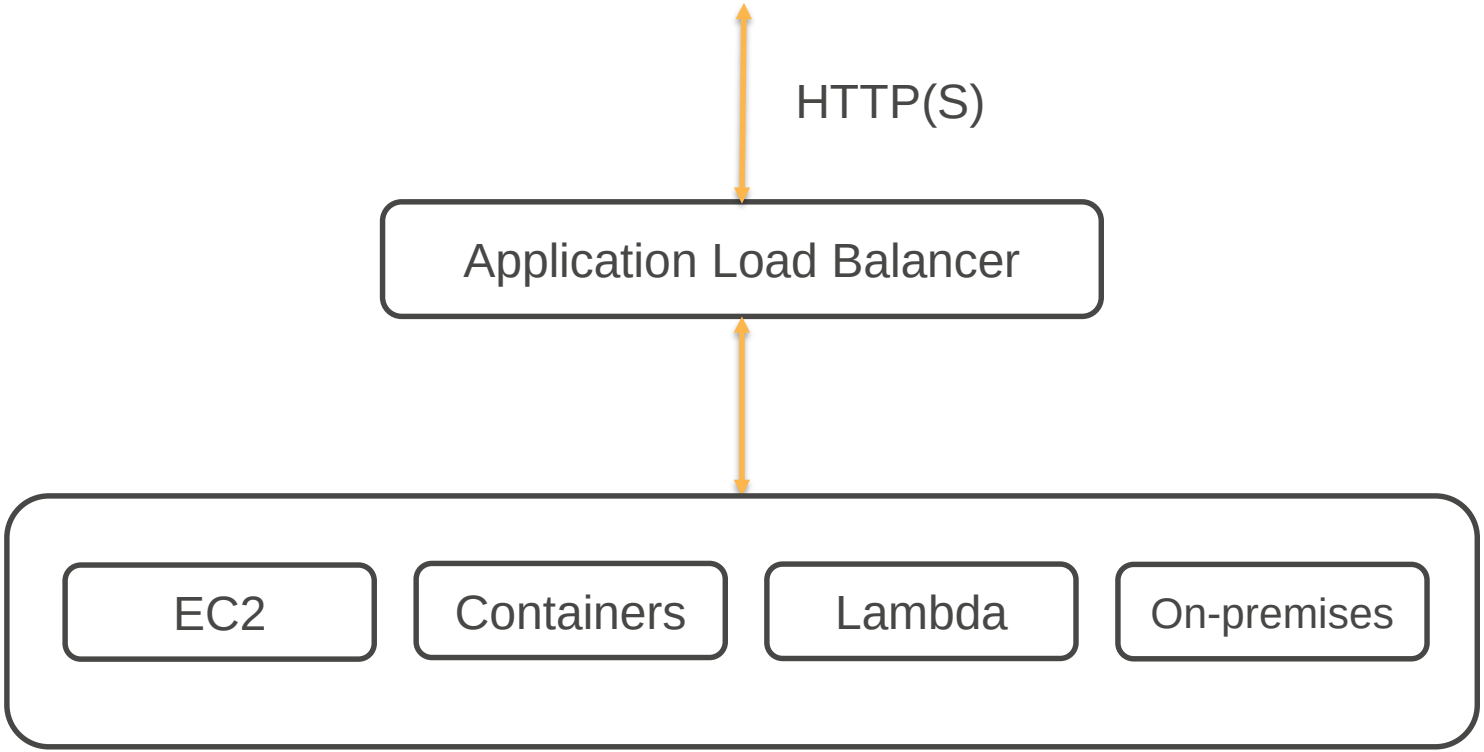
Classic, Application, Network

Classic Load Balancer

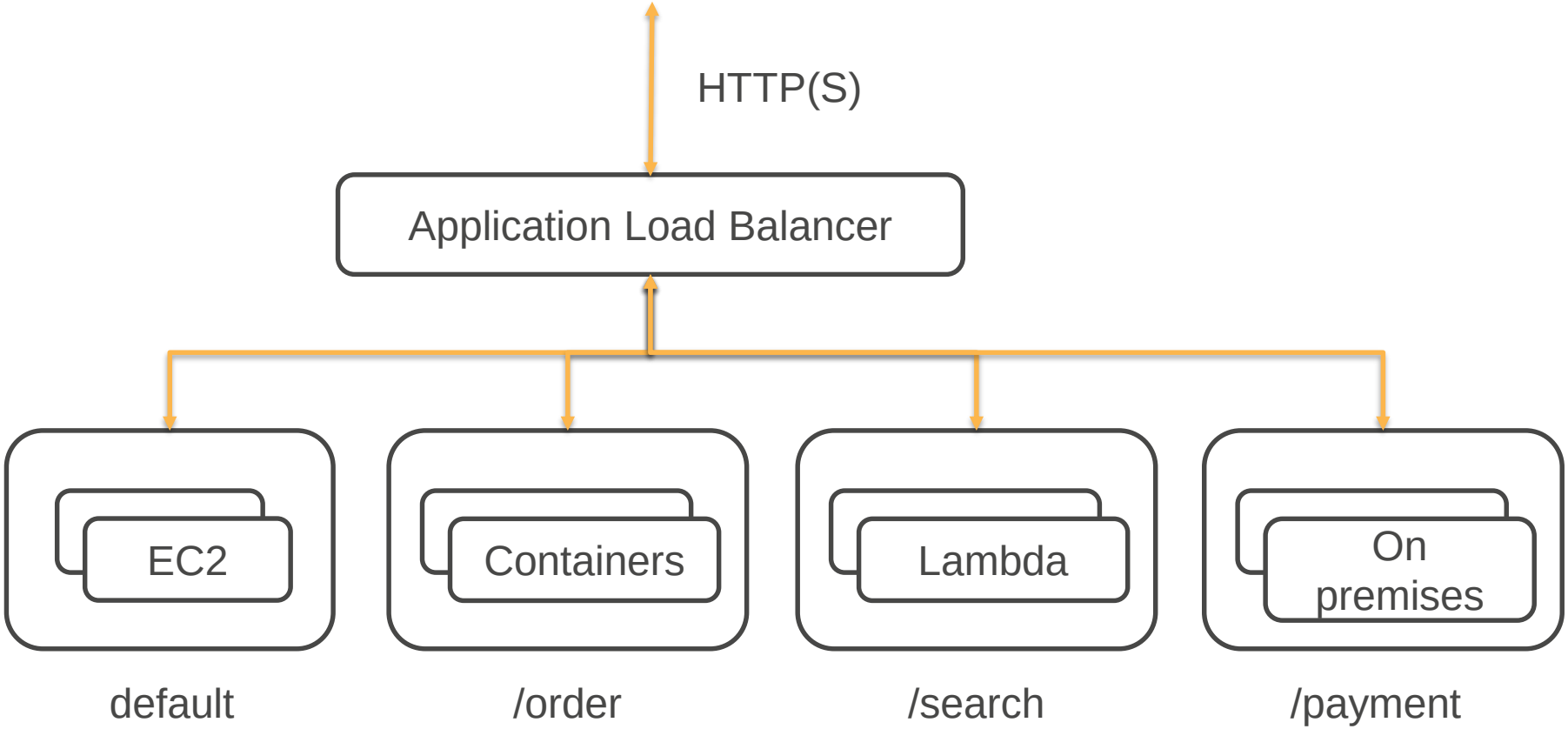


- Basic Load Balancing across multiple EC2 instances
- Supports HTTP(S) (Layer 7) and TCP (Layer 4) traffic
- Works both on EC2-Classic and VPC
- Previous generation product – recommended only for EC2-Classic

Application Load Balancer



Application Load Balancer - Routing



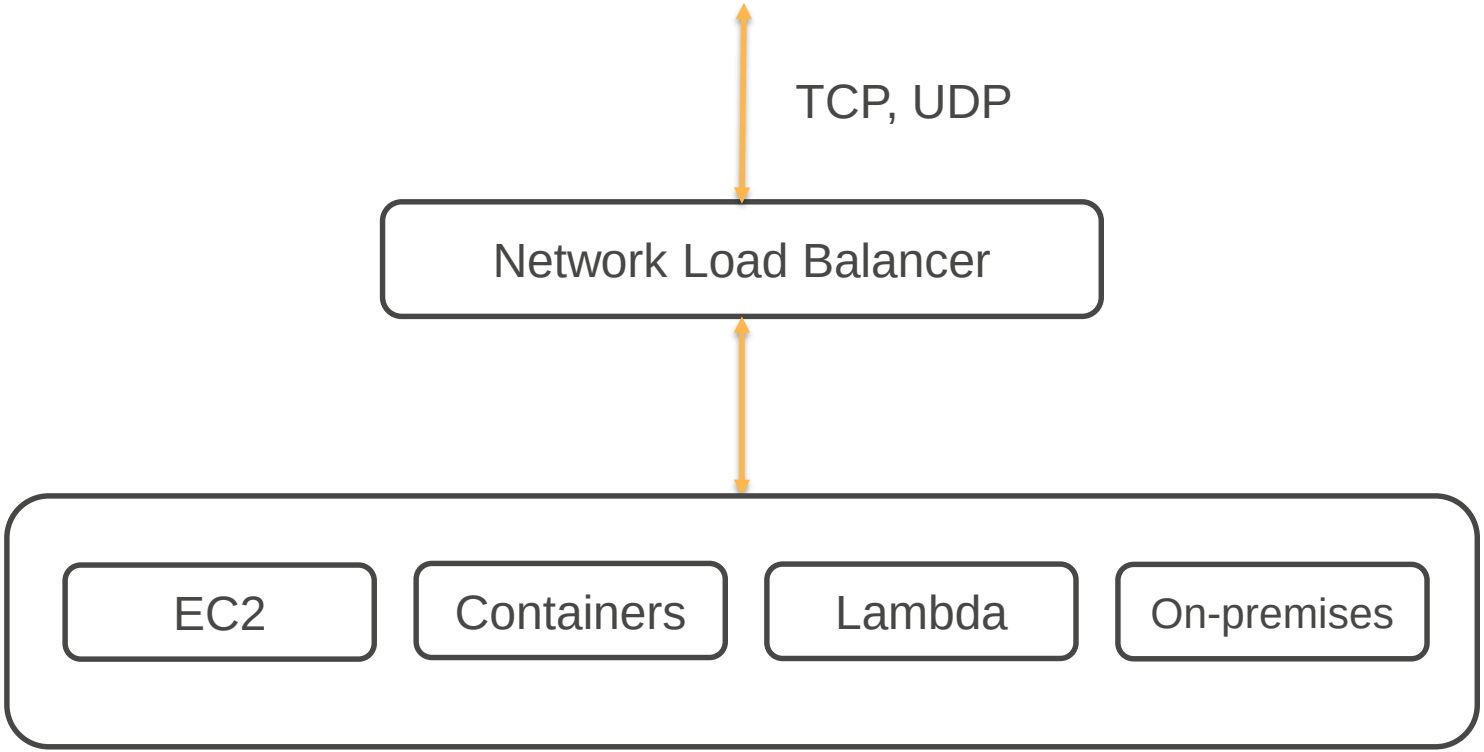
Application Load Balancer

- Ideal for load balancing HTTP(S) traffic (Layer 7)
- Advanced routing – for microservices, containerized applications, hybrid infrastructure
- HTTP/2 and WebSocket Support
- Request Tracing – track individual request by unique ID across various services
- Support for hosting multiple websites (Server Name Indication)
- User Authentication - Cognito

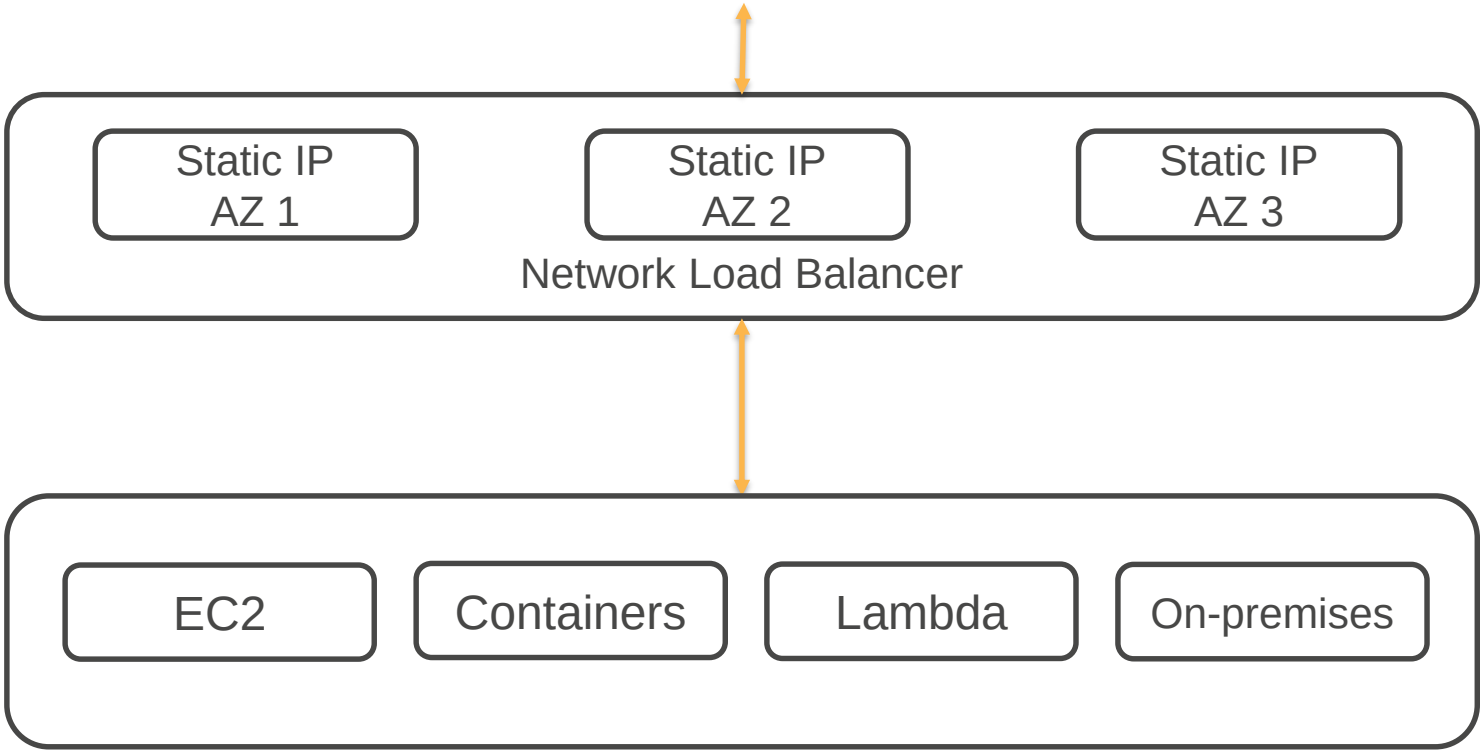
Application Load Balancer - Routing

- Path based
- Host HTTP header (support for multiple domains)
- Any standard or custom HTTP header
- Query string parameter based
- Source IP based (from where request is originating)

Network Load Balancer



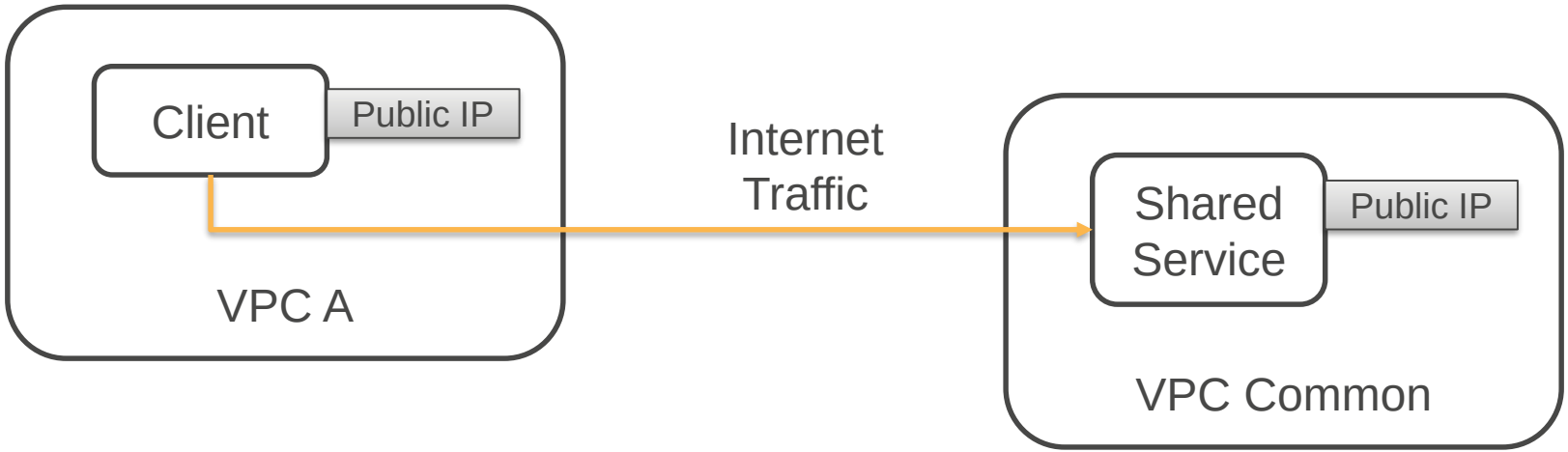
Network Load Balancer – Static IP



Network Load Balancer

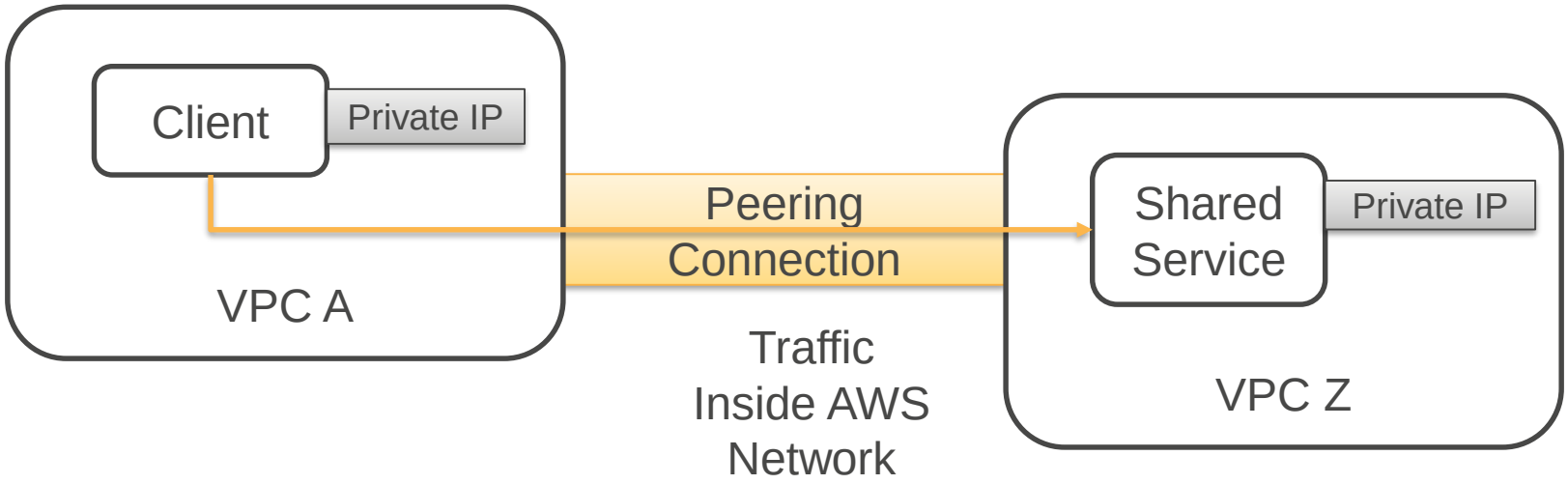
- Ideal for load balancing TCP and UDP traffic
- Scales to millions of requests/sec
- Handles Volatile traffic patterns
- One Static IP (or Elastic IP) per Availability Zone
    - Easy to whitelist Load Balancer IP in your Client
- Preserves Client IP (Source IP) – your application can use this for further processing
- WebSocket Support
- Private Link Support – Private communication between VPCs

Private Link - Motivation



- Shared Service accessed over the internet
- Your applications in AWS are communicating over internet
- Potential threat vectors (Denial of Service attacks on service)

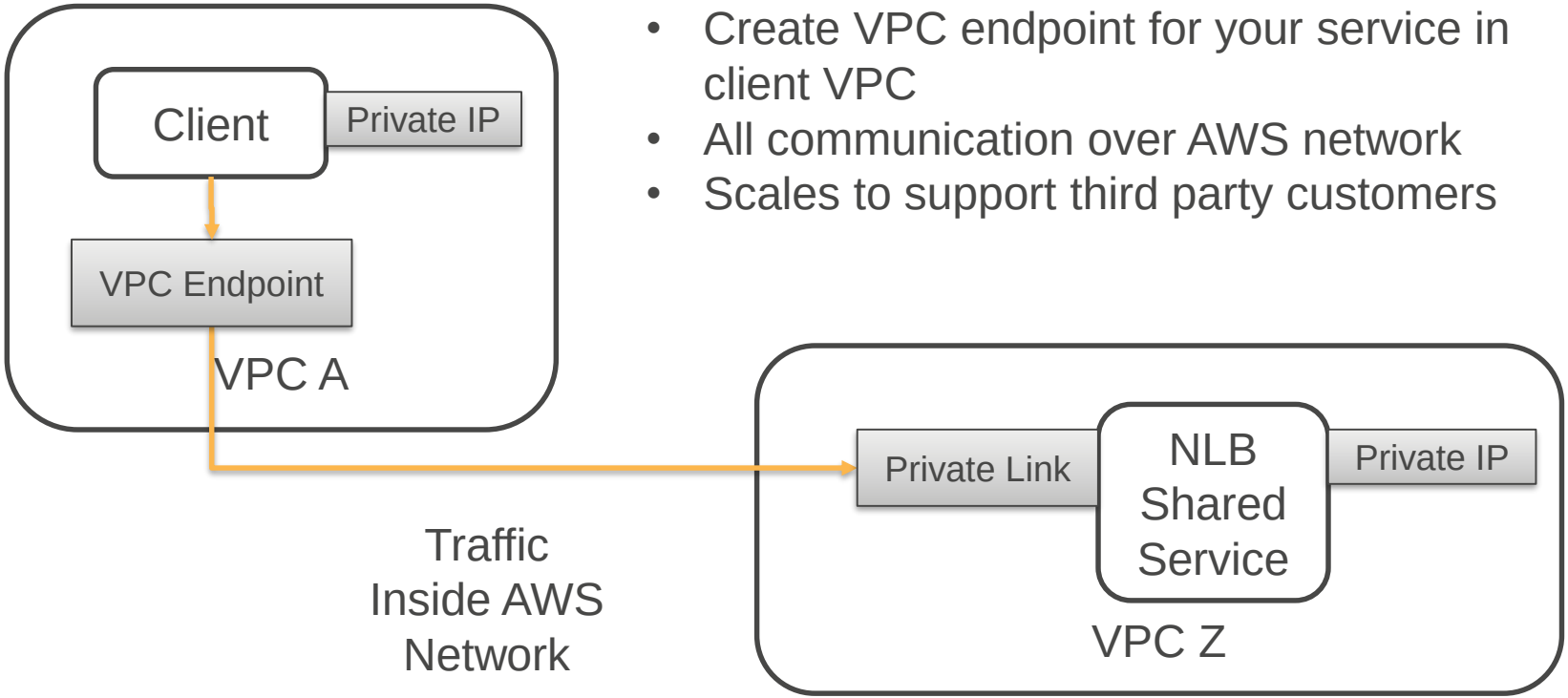
Private Link - Motivation



- Shared Service accessed over a peering connection with Private IP
- Traffic stays inside AWS network
- Network peering exposes resources on both sides – defeats the purpose of creating separate VPCs
- Not an option if shared services are used by third party customers in AWS

Private Link

- Network LB based Shared Service
- Configure as PrivateLink Powered Service
- Create VPC endpoint for your service in client VPC
- All communication over AWS network
- Scales to support third party customers



Access Logs

“Elastic Load Balancing provides access logs that capture detailed information about requests sent to your load balancer.

Each log contains information such as the **time the request was received, the client's IP address, latencies, request paths, and server responses**.

You can use these access logs to analyze traffic patterns and troubleshoot issues.”

“**Access logging is an optional feature of Elastic Load Balancing that is disabled by default**.

Load Balancer - Types

Load Balancer	Use
Classic	• Basic load balancing across multiple EC2 instances • HTTP(S) and TCP Support • Recommended for legacy applications on EC2-Classic network
Application	• Load Balance across EC2 instances, Containers, Lambda, and Hybrid infrastructure • HTTP(S) traffic support (Layer 7) • Route traffic to target based on the content of the request
Network	• Load Balance across EC2 instances, Containers, Lambda, and Hybrid infrastructure • TCP, UDP traffic support (Layer 4) • Extreme performance

Lab – Classic Load Balancer

- Load Balance traffic across EC2 instances
- Health Checks
- Simulate Error

Lab – Application Load Balancer

- Load Balance traffic across EC2 instances
- Health Checks
- Path based routing

Lab – Network Load Balancer

- Load Balance traffic across EC2 instances
- Health Checks
- Static IP

Chandra Lingam

50,000+ Students

Up-to-date Content