AWS 이론 4
- Load Balancing

Sung-Dong Kim,

School of Computer Engineering,

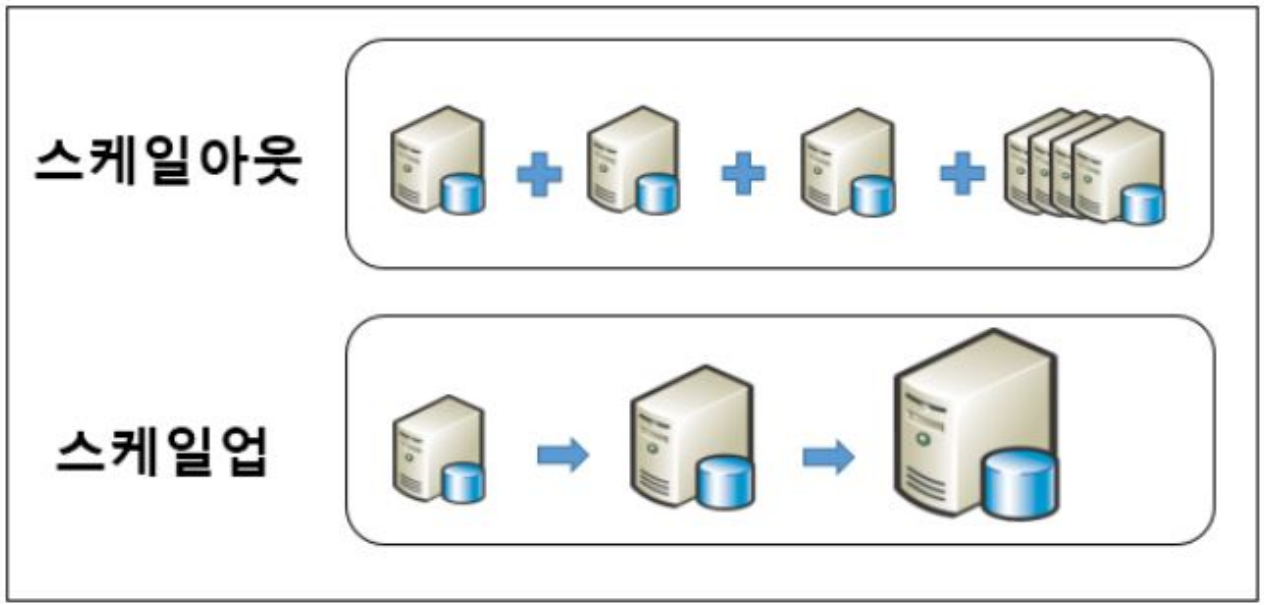
Hansung University

What

- 네트워크 기술의 일종
- 네트워크 트래픽을 하나 이상의 서버나 장비로 분산하는 기술
- 로드 밸런서(load balancer): 로드 밸런싱을 수행하는 SW, HW

Scale up/out (1)

- 웹 트래픽 증가에 대한 처리 방식



https://m.blog.naver.com/islove8587/220548900044

Scale up/out (2)

- Scale up
  - 더 높은 성능을 가진 서버로 upgrade
  - 비용, 가용성 문제
- Scale out
  - 저렴한 여러 노드로 cluster 구성
  - 가용성이 높음
  - 로드 밸런싱

로드 밸런싱 방식

- round robin: server로의 session 연결을 순차적으로 수행
- hash: client와 server 간에 연결된 session을 계속 유지
- least connection: 가장 작은 session을 보유한 server로 session을 연결
- response time
  - resource와 connection의 차이가 있는 환경
  - 빠른 응답시간을 제공하는 서버로 session 연결

AWS ELB

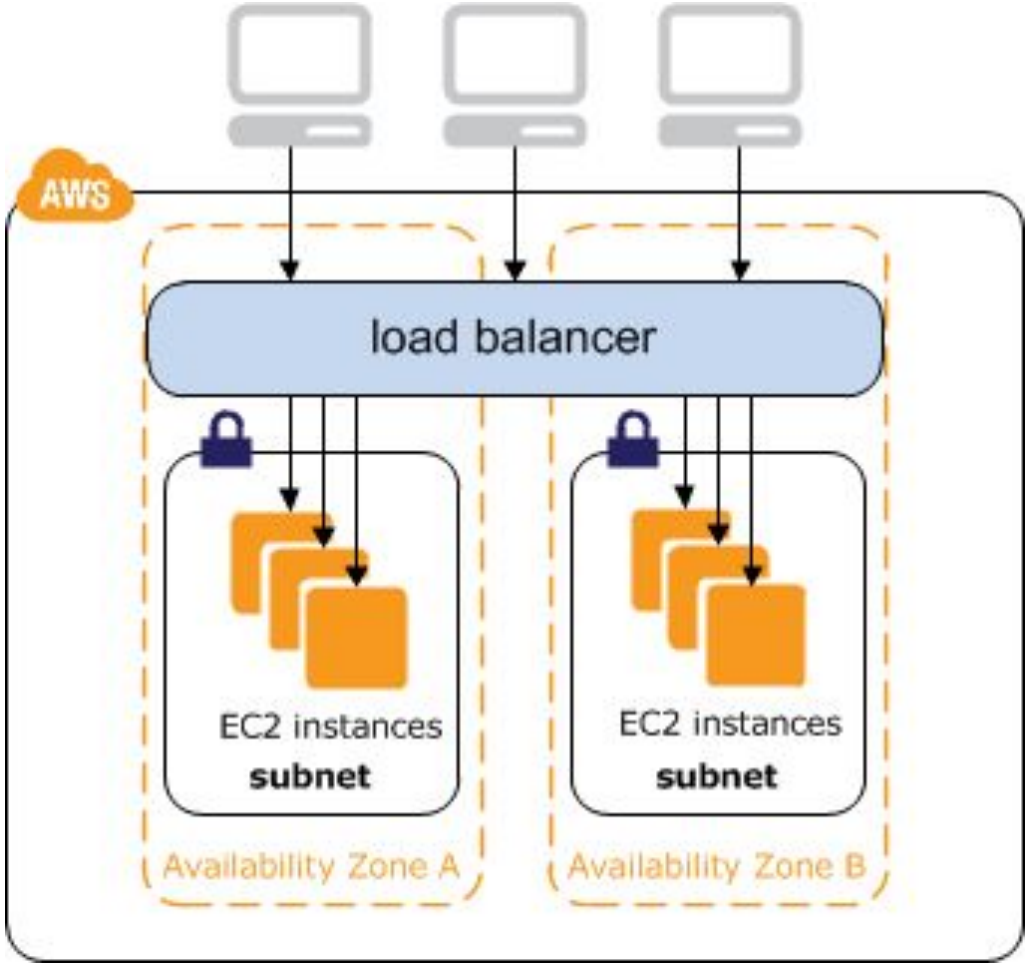
- EC2 instance, container, IP address 같은 여러 대상에 대해 수신 application 또는 network traffic을 여러 AZ에 배포
- workload를 다수의 computing resource로 분산
- application의 가용성, 내결함성 향상

특징 (1)

- health check (상태 확인)
  - ELB와 연결된 instance의 연결 상태를 수시로 체크
  - HTTP, HTTPS: 웹 페이지 접속 시도에 대한 응답 코드(200) 정상 반환 여부 확인
- sticky session
  - client가 처음 접속한 서버로 계속 연결시켜줌
  - 처음 연결된 client에 별도의 HTTP 기반의 쿠키 값 생성, 이용

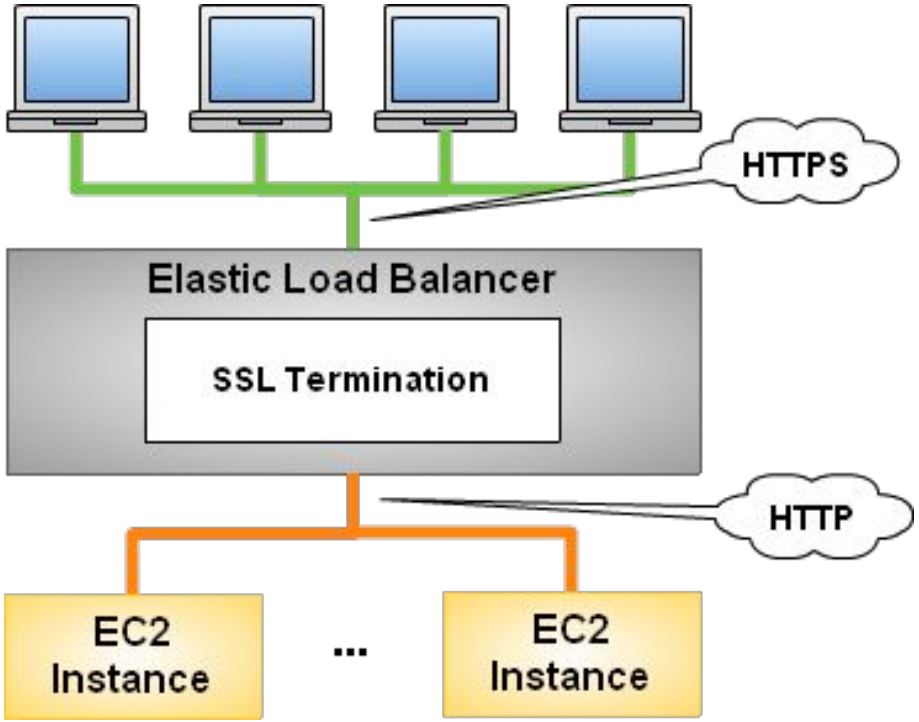
특징 (2)

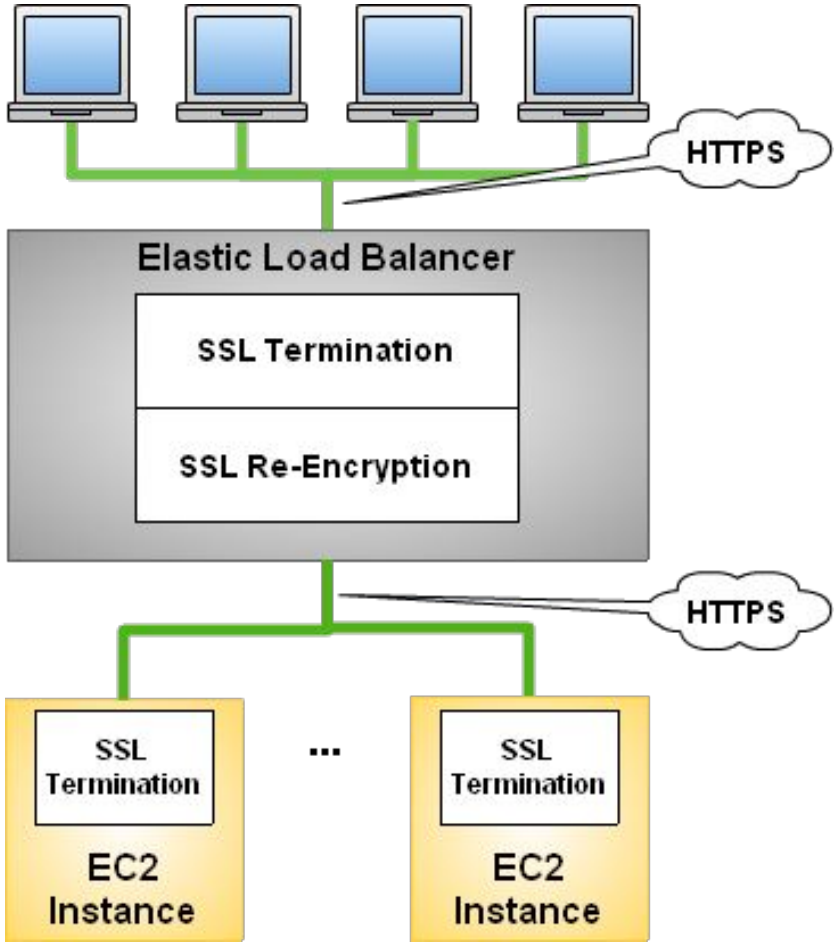
- 고가용성 구성
  - Route53 등의 AWS의 다른 서비스와의 연계를 통해 가용성 서비스 제공


AWS
load balancer
EC2 instances
subnet
EC2 instances
subnet
Availability Zone A
Availability Zone B

특징 (3)

- SSL termination 및 보안 기능
  - HTTPS와 같은 암호화 통신: 웹 서버에 별도의 공인인증서 필요
  - 개별 EC2 instace에서 SSL 암호화 및 복호화를 직접 처리
  - ELB의 SSL termination: 개별 instance에 SSL 인증서를 직접 설치할 필요 없음


HTTPS
Elastic Load Balancer
SSL Termination
HTTP
EC2 Instance
...
EC2 Instance


HTTPS
Elastic Load Balancer
SSL Termination
SSL Re-Encryption
HTTPS
SSL Termination
EC2 Instance
...
SSL Termination
EC2 Instance

종류

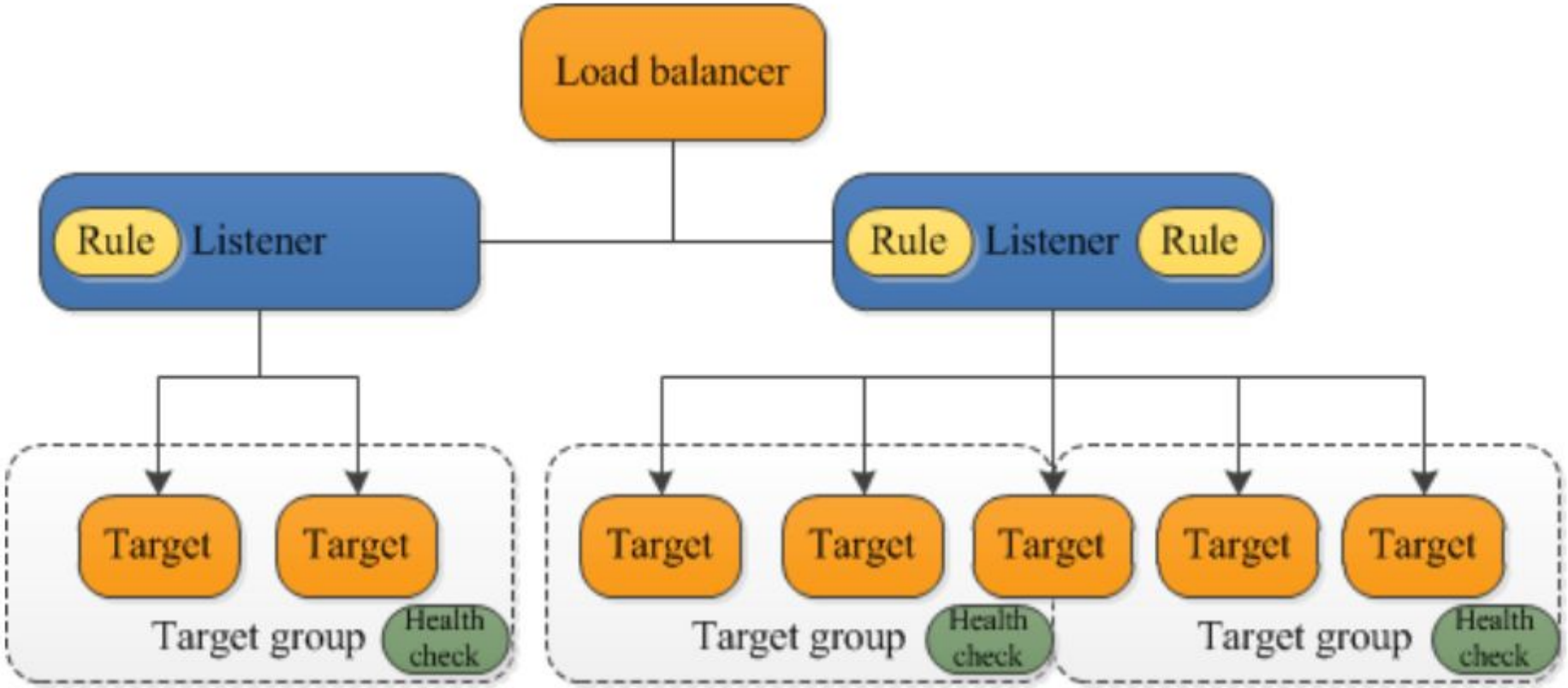
- Application Load Balancer
- Network Load Balancer
- Classic Load Balancer

Classic/Network Load Balancer

- classic load balancer
  - EC2-classic 네트워크 내에 구축된 application 대상
  - OSI 4-layer (transport layer), 3-layer (network layer)에서 작동
- network load balancer
  - OSI 4-layer (transport layer)에서 작동, TCP 트래픽 로드 밸런싱
  - 짧은 지연 시간

Application Load Balancer (1)

- load balancer: single point of contact for clients
- listener
  - clinet의 접속 요청 검사 → target group으로 보냄
  - rule을 정의: target group, condition, prority
- target group: 등록된 target으로 요청을 보냄 (protocol, port 이용)


Load balancer
Rule
Listener
Rule
Listener
Rule
Target
Target
Target group
Health check
Target
Target
Target
Target
Target
Target group
Health check
Target group
Health check

Application Load Balancer (2)

- OSI layer의 7번 째 application layer에 서 작 동: HTTP/HTTPS
- request → listner rule 결 정 → target group에 서 target 결 정

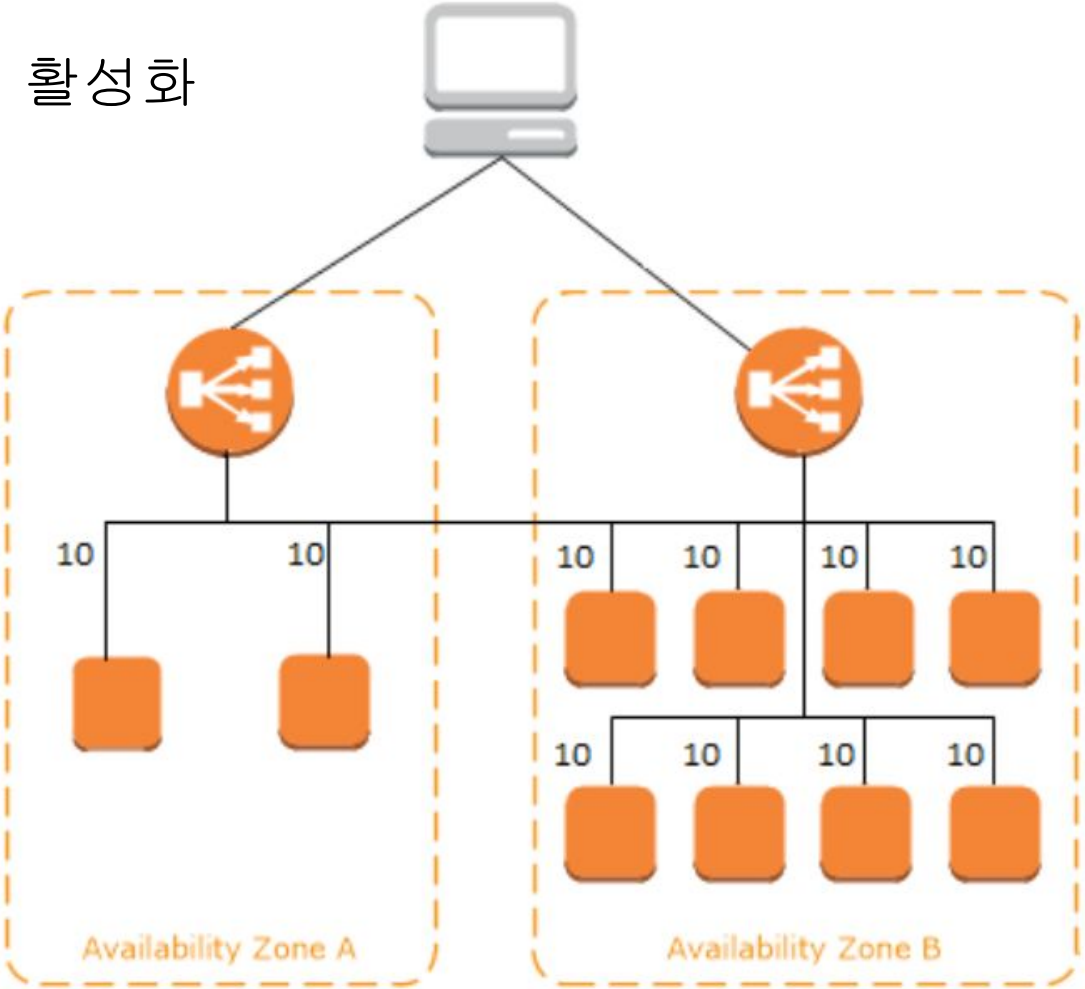
작동 방식

- client에서의 traffic을 받아, target group의 instance에 요청을 라우팅
- target group instance의 상태 모니터링
- listner
  - incomming traffic을 허용
  - protocol + port number

가용 영역 및 로드 밸런서 노드

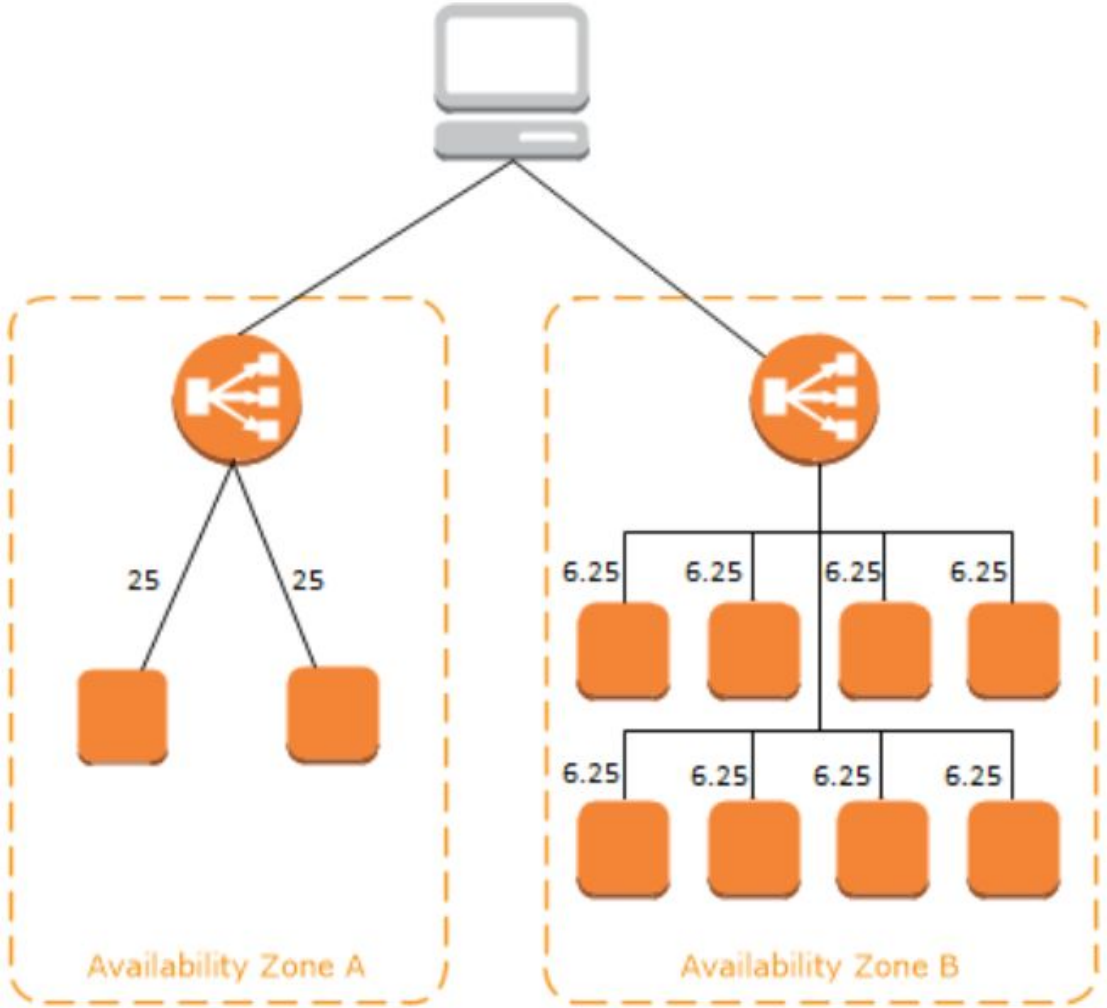
- 여러 개의 가용 영역 활성화: 각 AZ에 하나 이상의 대상 등록
- 교차 영역 로드 밸런싱
  - 모든 AZ에 등록된 대상으로 트래픽 분산
  - Application load balancer에서는 항상 활성화 상태
  - Network load balancer에서는 기본적으로 비활성화 상태

교차 영역 로드 밸런싱 활성화



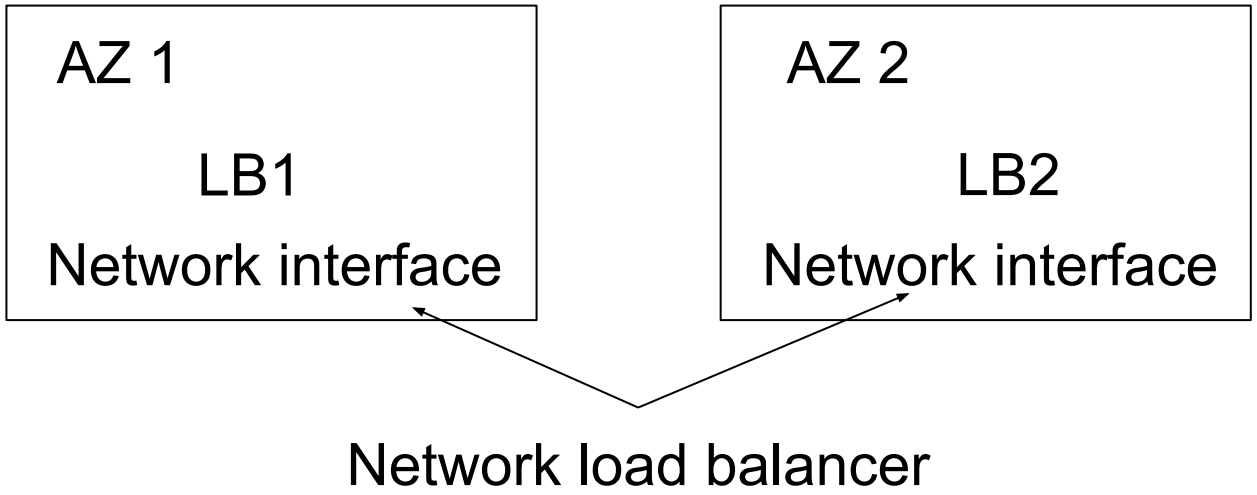
교차 영역 로드 밸런싱

비활성화



라우팅 요청 (1)

- client → DNS server → domain 이름 해석 → LB에 요청을 보냄: LB ← static IP from NI



라우팅 요청 (2)

- client의 요청을 받은 LB는 상태가 양호한 대상을 선택 → private IP를 사용하여 요청 전송
- 라우팅 알고리즘
  - priority에 따라 listner의 규칙 평가 후 규칙 결정
  - round-robin algorithm으로 대상 그룹에서 대상 선택

로드 밸런서 체계 (1)

- 내부 로드 밸런서 vs 인터넷 경계 로드 밸런서
  - private IP를 사용하여 요청을 라우팅
  - public IP가 없는 target instance도 요청을 수신할 수 있음

로드 밸런서 체계 (2)

- 인터넷 경계 로드 밸런서
  - public IP, private IP
  - VPC 내부 및 internet을 통해 client의 요청을 라우팅
- 내부 로드 밸런서
  - private IP
  - VPC 내부만 접속