NT-02

Build a better and more secure user experience at the edge

Toni Syvänen
Senior Game Tech Solutions Architect
AWS

Agenda

- Availability matters
- Engineered for high availability
- Architecture patterns for high availability

- This is a Level 300 talk

“Everything fails, all the time”

Werner Vogels
Amazon CTO

Availability matters

The average cost of downtime
is $5,600 per minute;
well over $300k per hour

Gartner

AWS Edge Networking Services

01 Amazon CloudFront

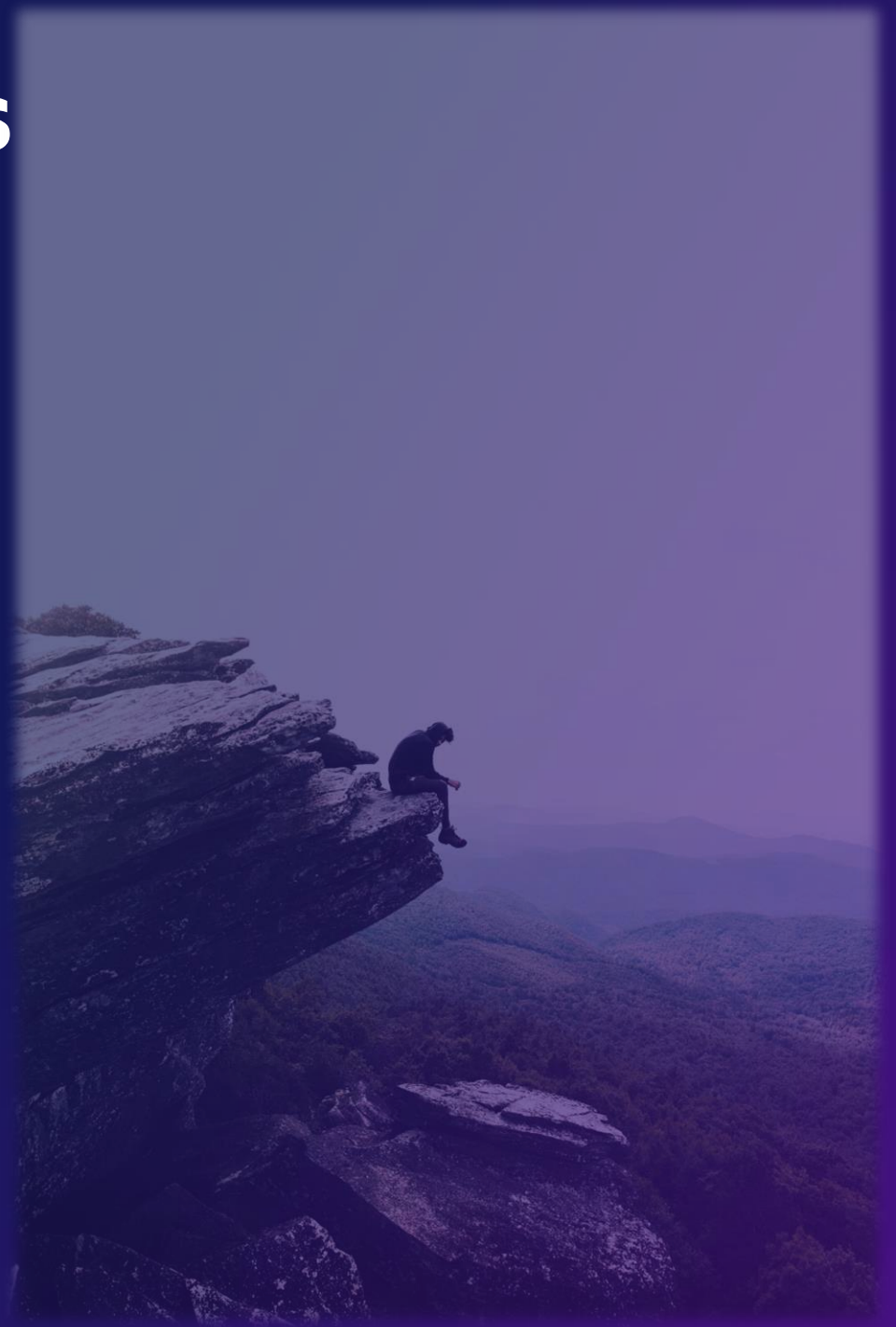
02 AWS Global Accelerator

03 AWS WAF

04 AWS Shield

05 Amazon Route 53

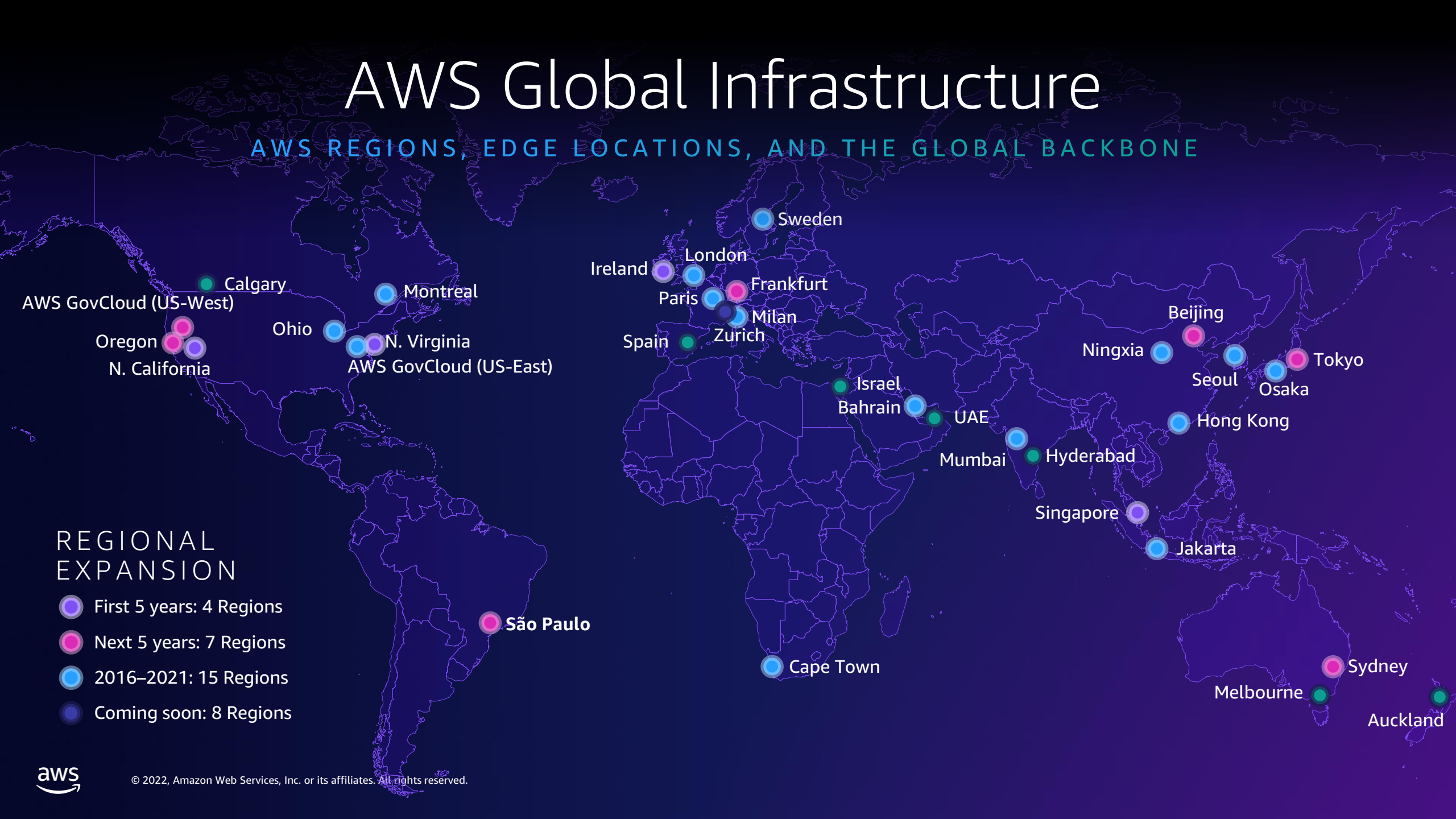
06 Lambda @Edge, CloudFront Functions



Engineered for High Availability

AWS Global Infrastructure

AWS REGIONS, EDGE LOCATIONS, AND THE GLOBAL BACKBONE

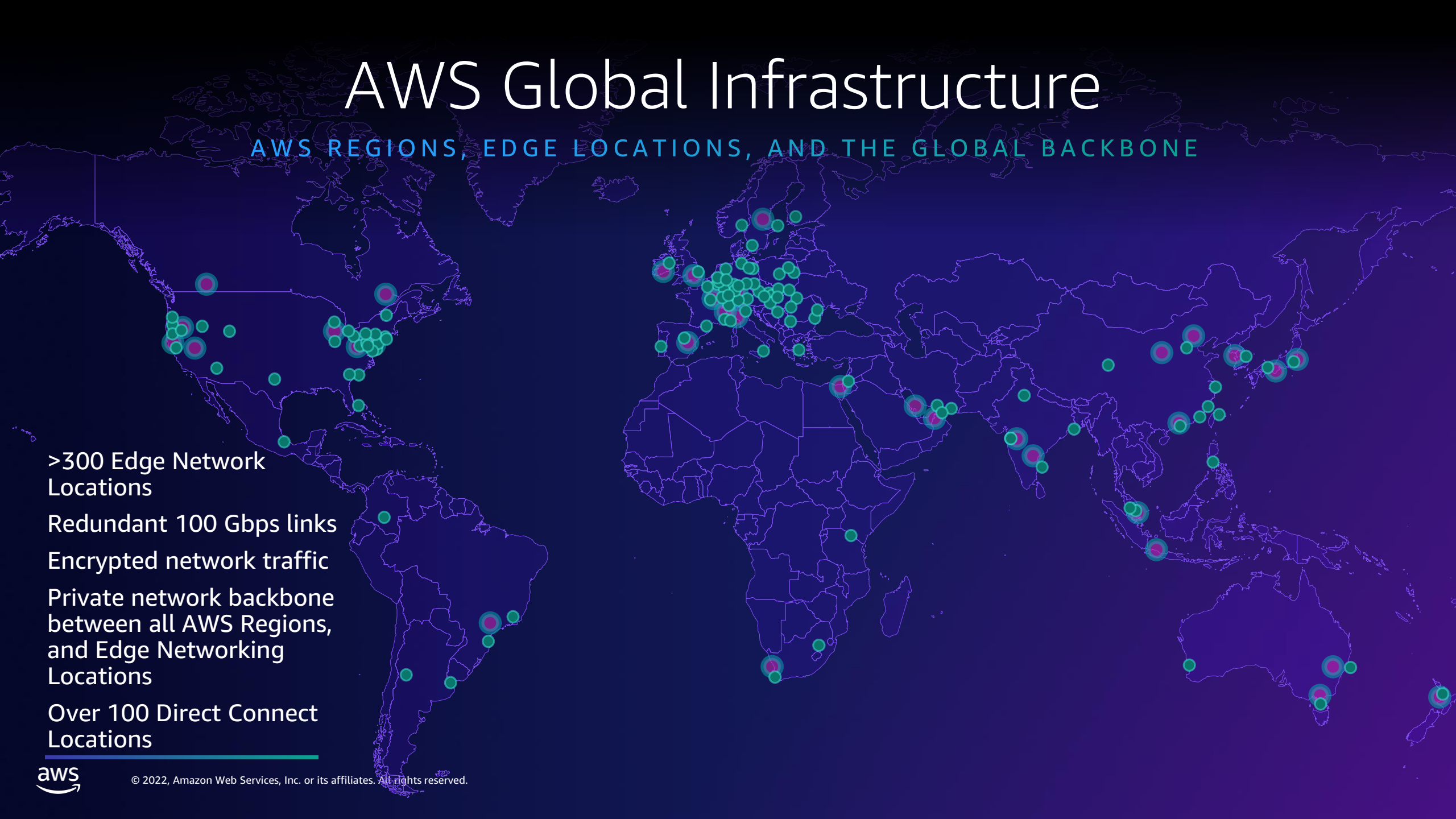


REGIONAL EXPANSION

- First 5 years: 4 Regions
- Next 5 years: 7 Regions
- 2016–2021: 15 Regions
- Coming soon: 8 Regions

AWS Global Infrastructure

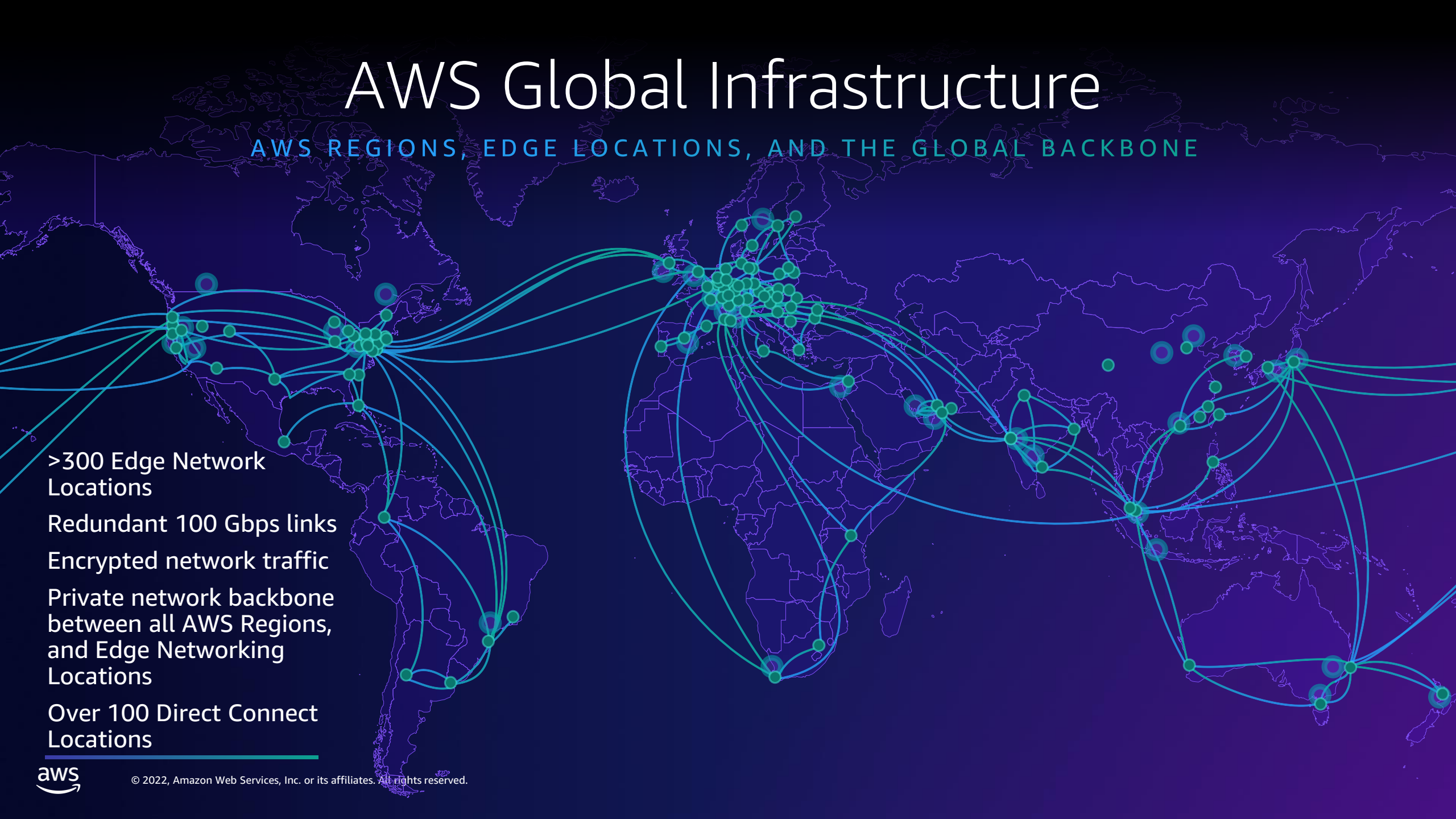
AWS REGIONS, EDGE LOCATIONS, AND THE GLOBAL BACKBONE



- >300 Edge Network Locations
- Redundant 100 Gbps links
- Encrypted network traffic
- Private network backbone between all AWS Regions, and Edge Networking Locations
- Over 100 Direct Connect Locations

AWS Global Infrastructure

AWS REGIONS, EDGE LOCATIONS, AND THE GLOBAL BACKBONE

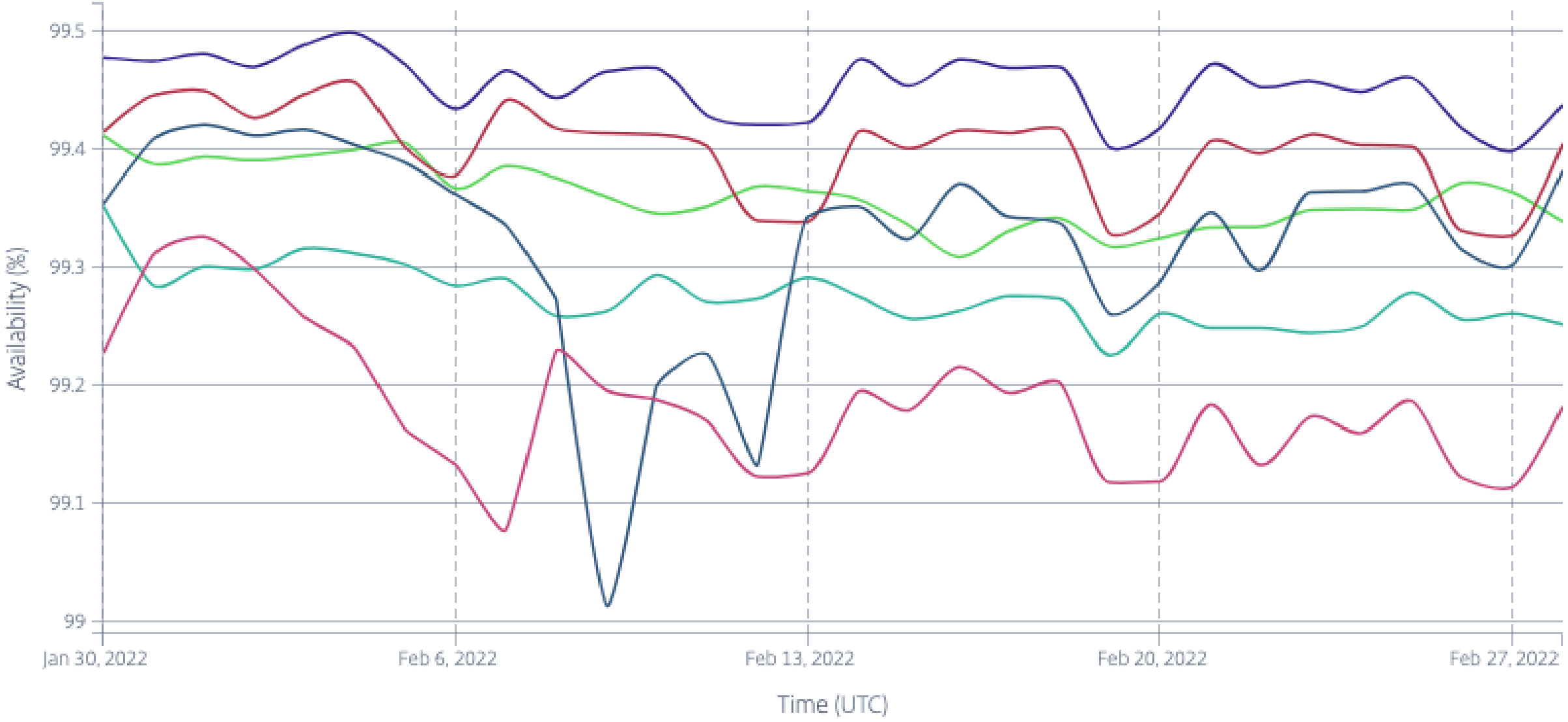


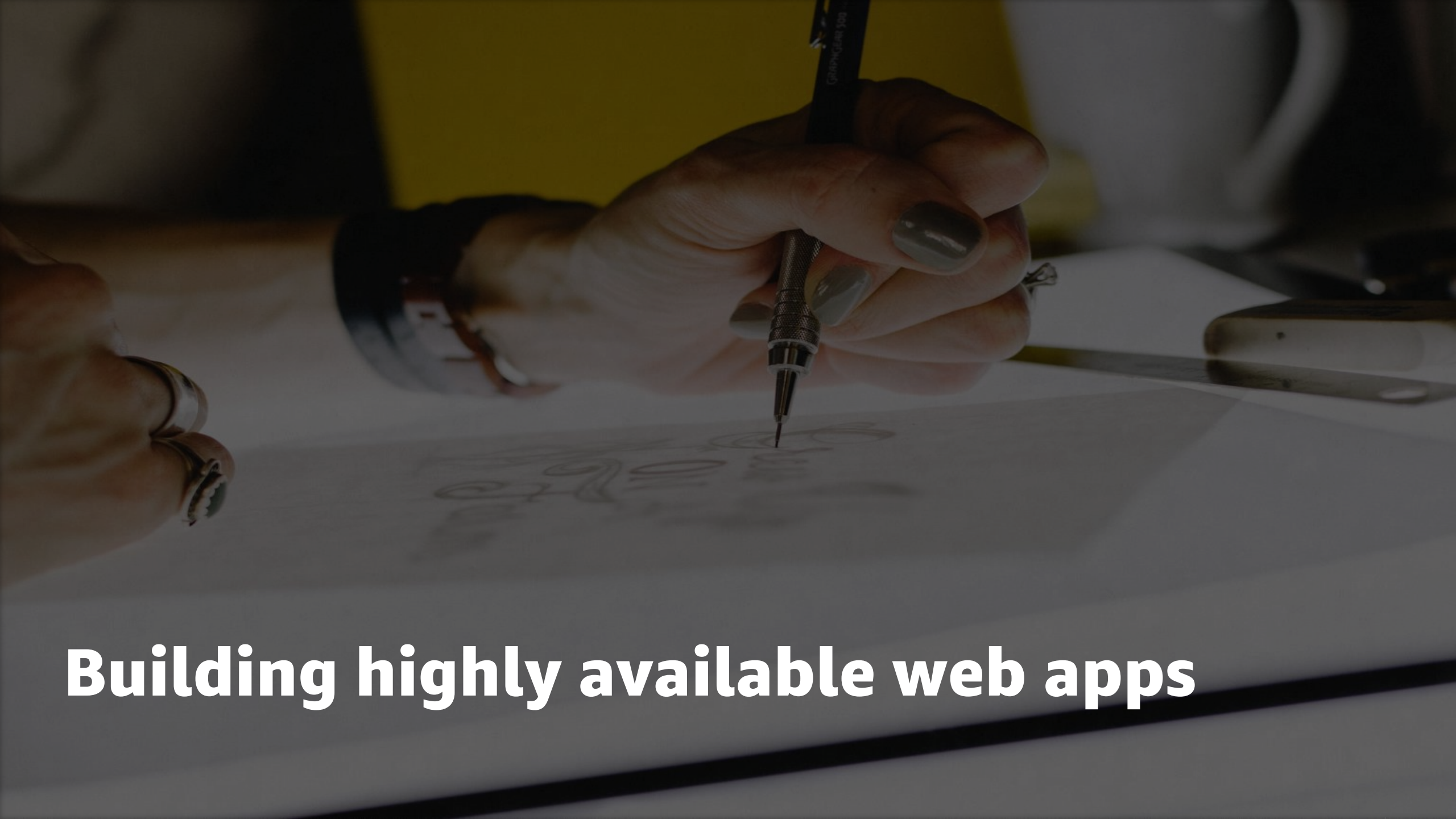
- >300 Edge Network Locations
- Redundant 100 Gbps links
- Encrypted network traffic
- Private network backbone between all AWS Regions, and Edge Networking Locations
- Over 100 Direct Connect Locations

AWS Under the hood: Design patterns for high availability

Route 53	Shuffle Sharding	Noisy neighboors
CloudFront	Food Tasting	Corrupted data
Global Accelerator	Striped CI/CD	Software bugs
CloudFront	Diversity	Mono culture

Filters: Last 30 Days, Entire Radar Community, Client IP, Availability, Platforms 8, Availability



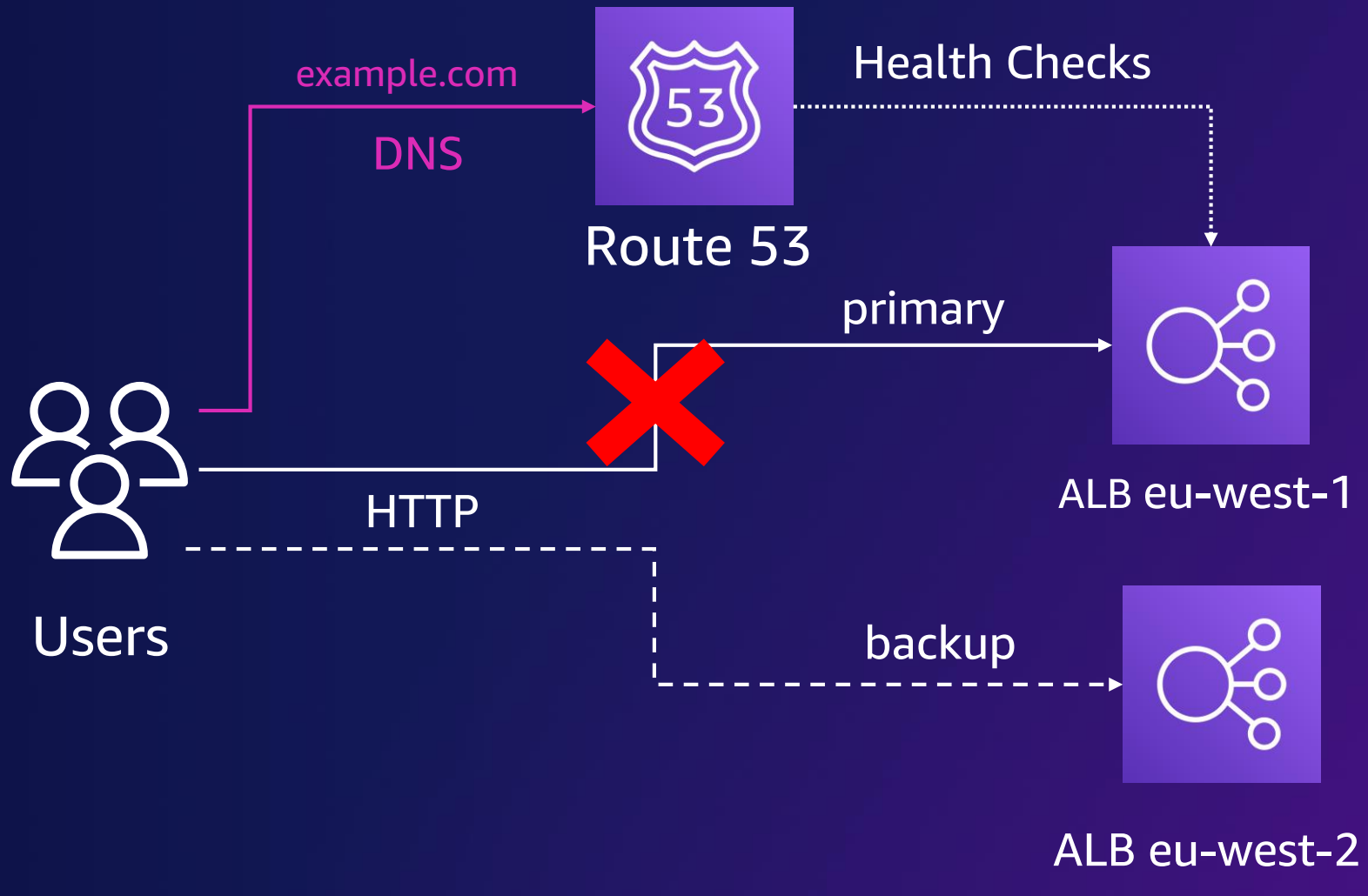


Building highly available web apps

Origin failover

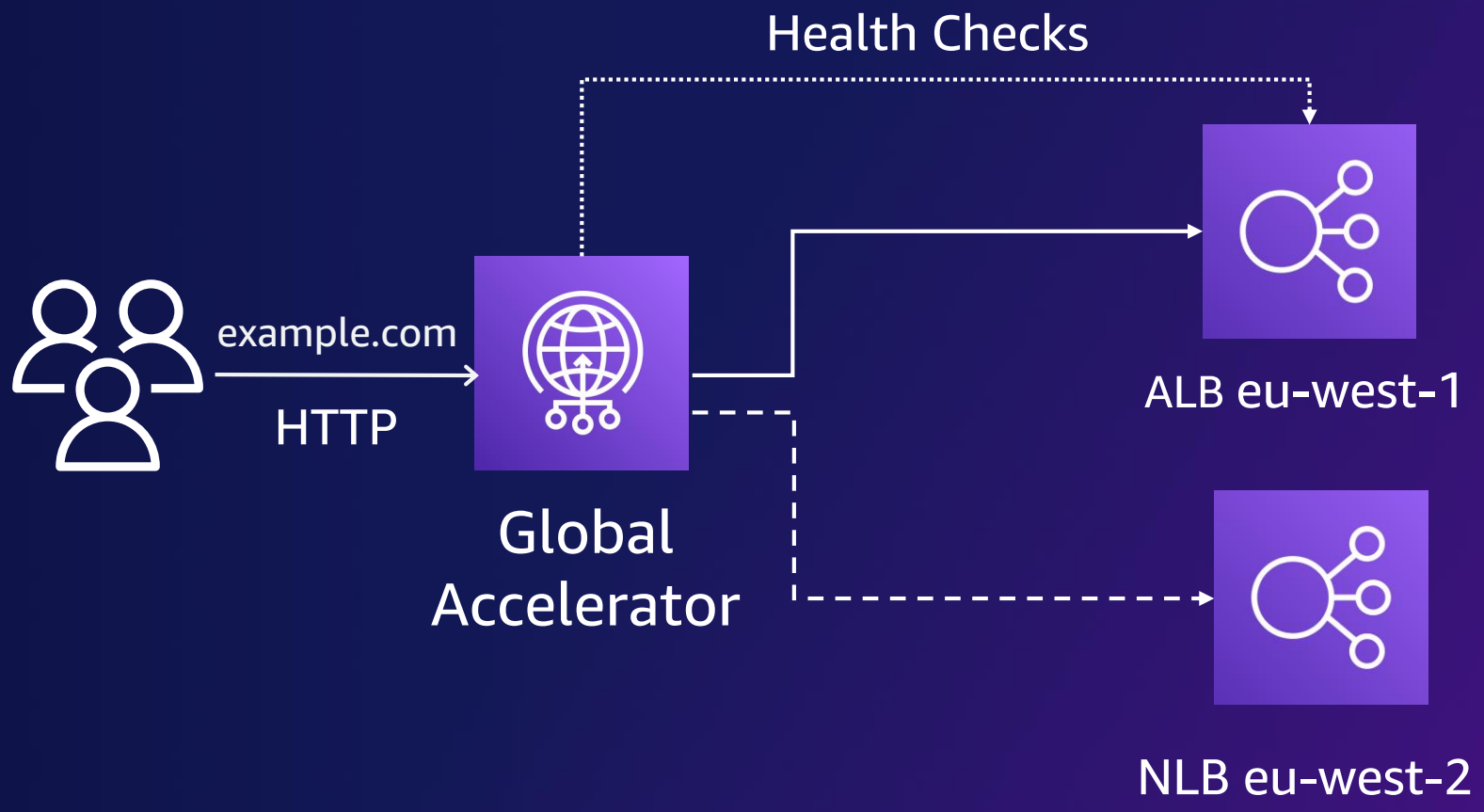
Origin Failover using Route 53

- Stateful
- Cost effective
- Custom origins
- Challenge with resolvers respecting DNS TTL



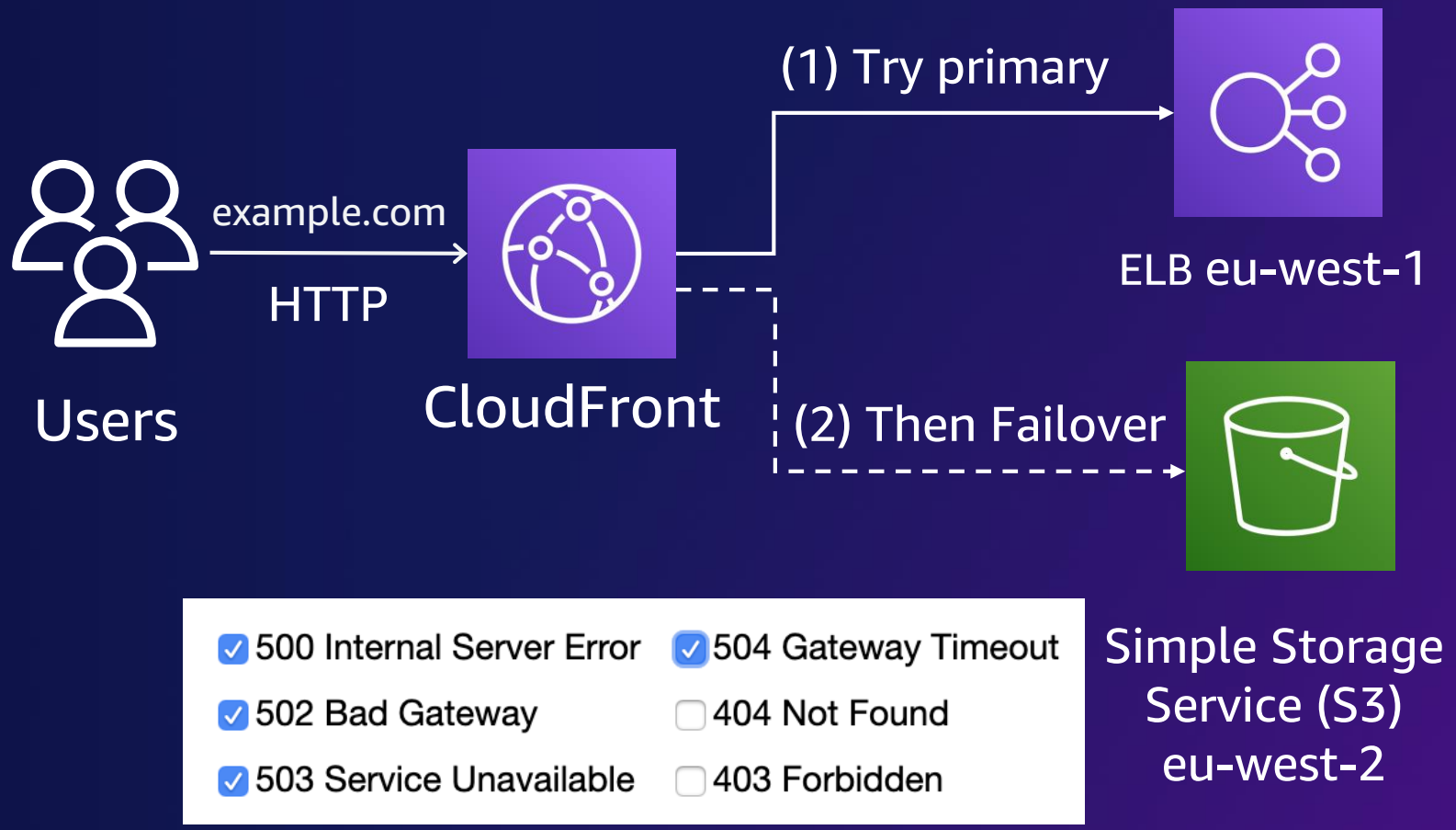
Origin Failover using Global Accelerator

- Stateful
- Works with non HTTP apps
- Failover in less than 30 seconds
- Premium DTO
- Works with EC2/ALB/NLB/EC2/ EIP

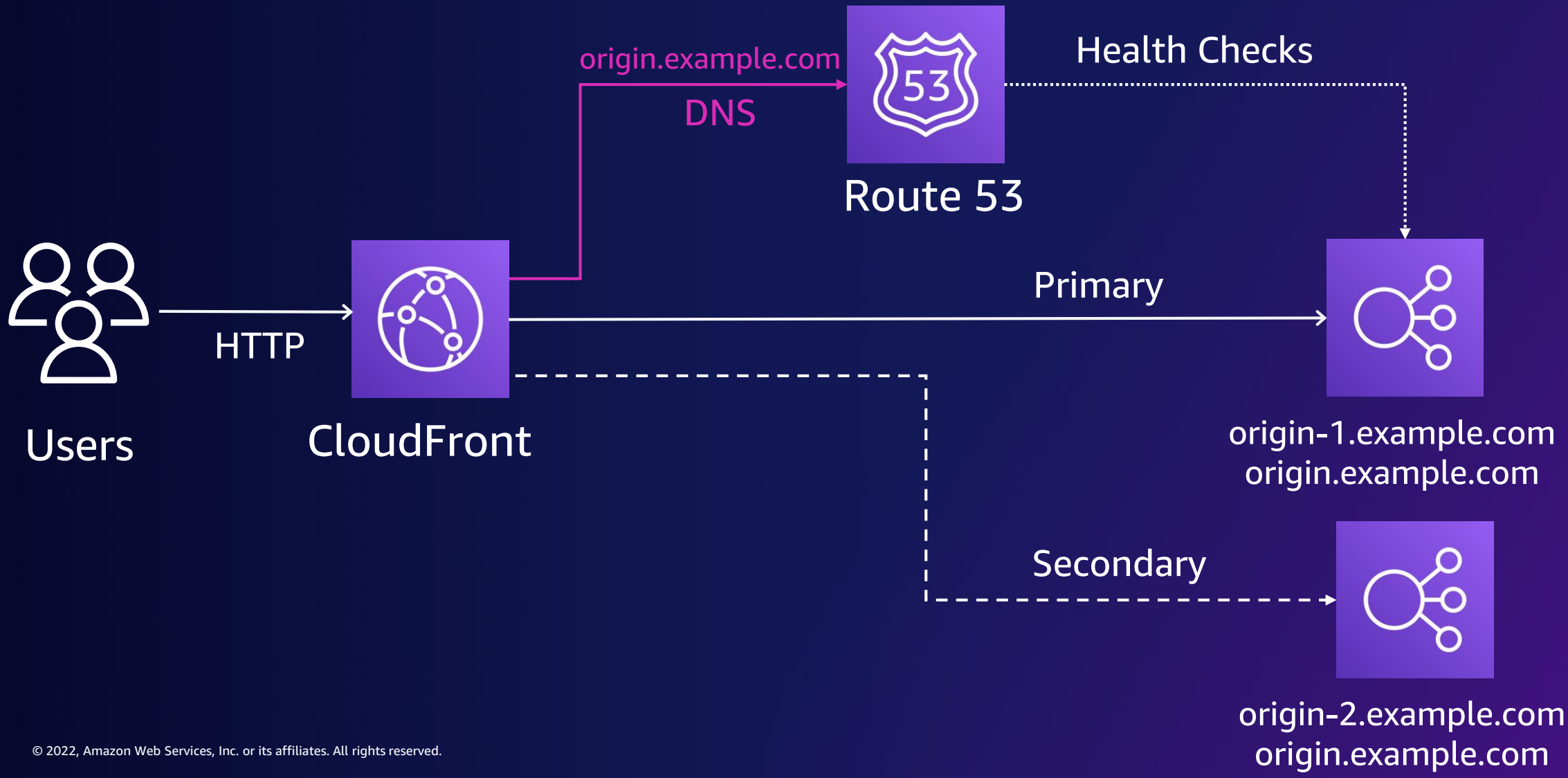


Origin Failover using CloudFront

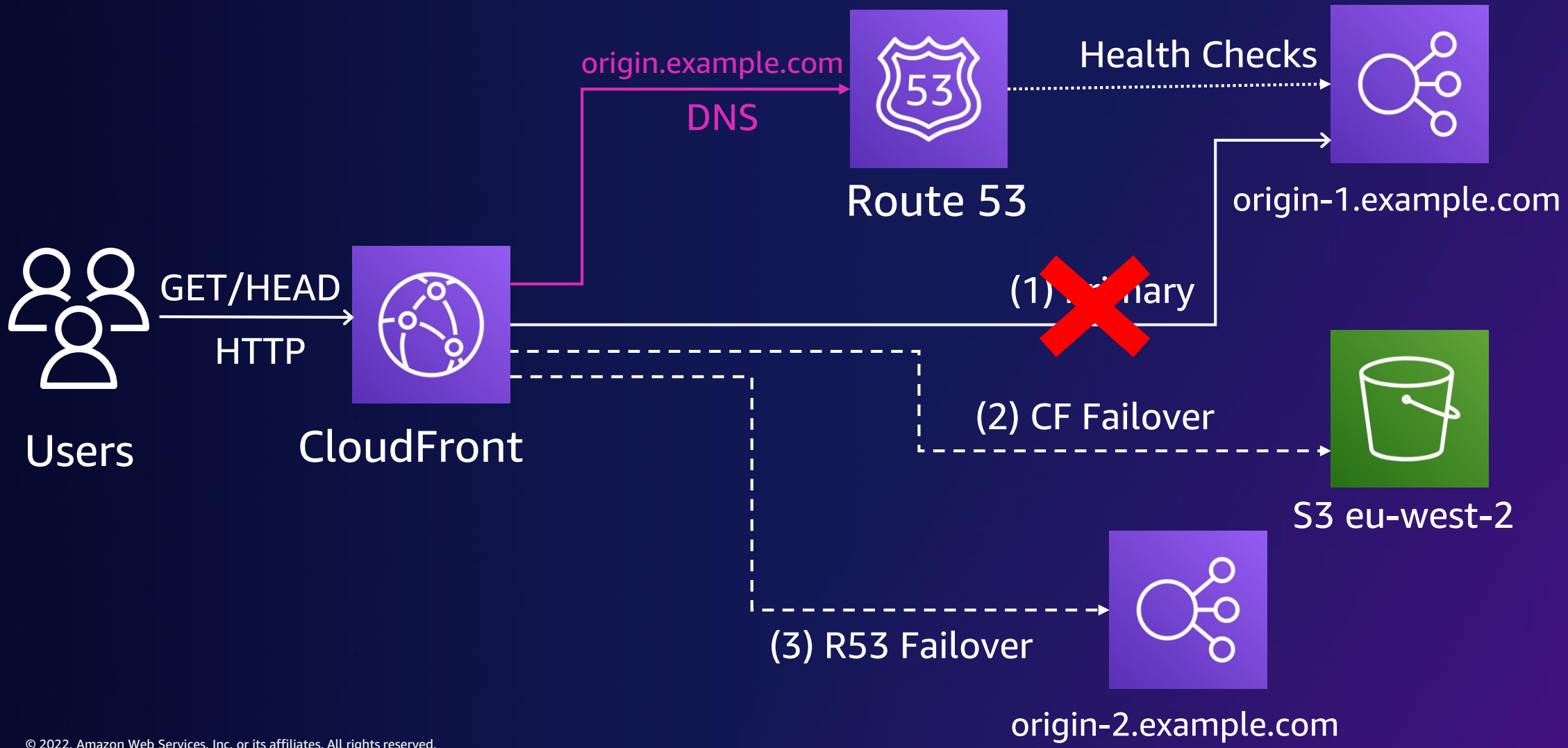
- Stateless
- Immediate (1s)
- CloudFront as proxy
- Any origin
- Only GET requests
- Stable User connection



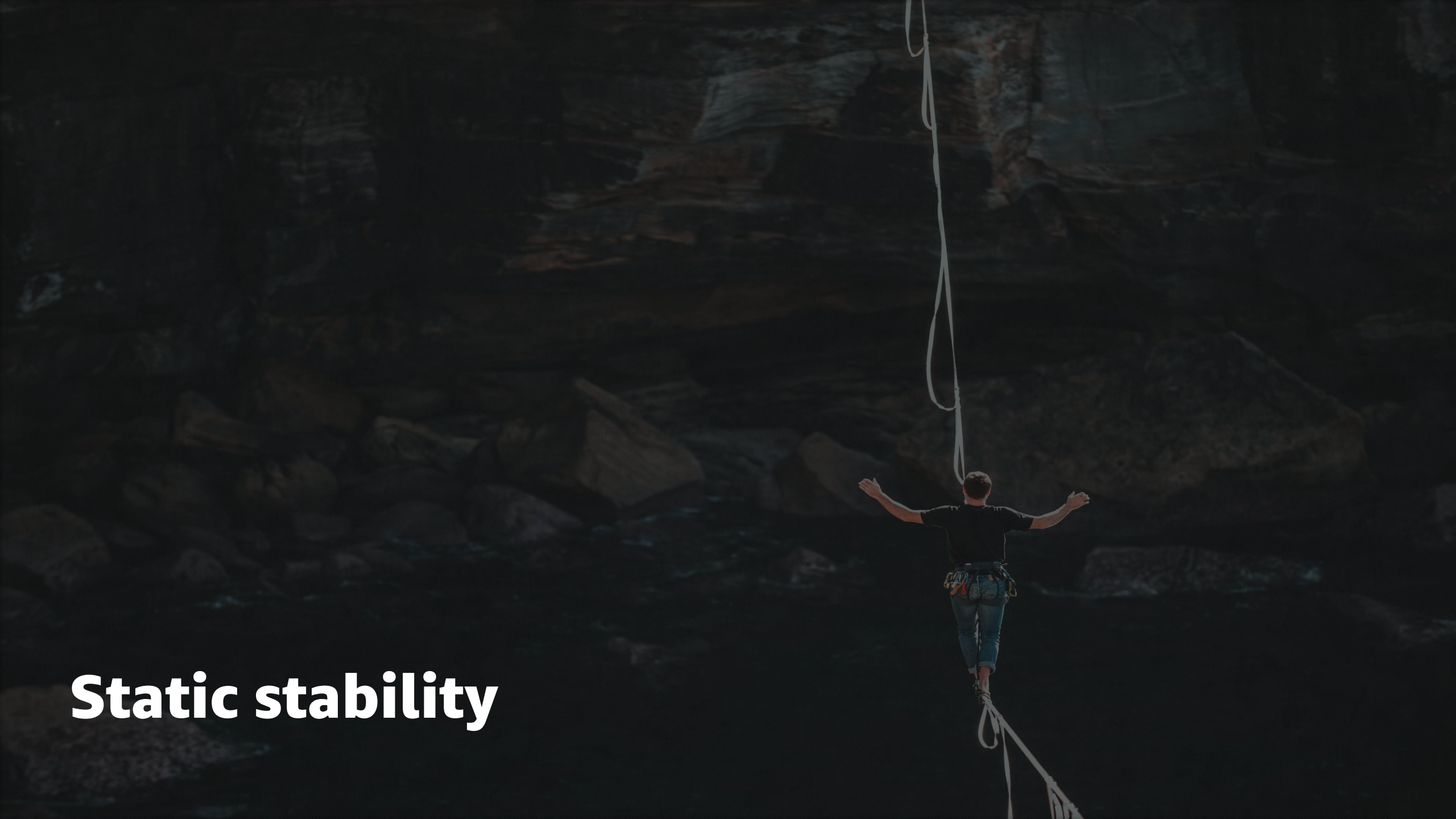
Combining CloudFront with Route 53



Combine CloudFront with Route 53 and S3



DEMO



Static stability

Static stability of CloudFront

Control Plane

- Create, update, remove distributions, invalidate content, reporting
- Prioritize strong consistency and durability
- Regional component in us-east-1

Data Plane

- Request processing
- Prioritize availability
- Global and distributed infrastructure

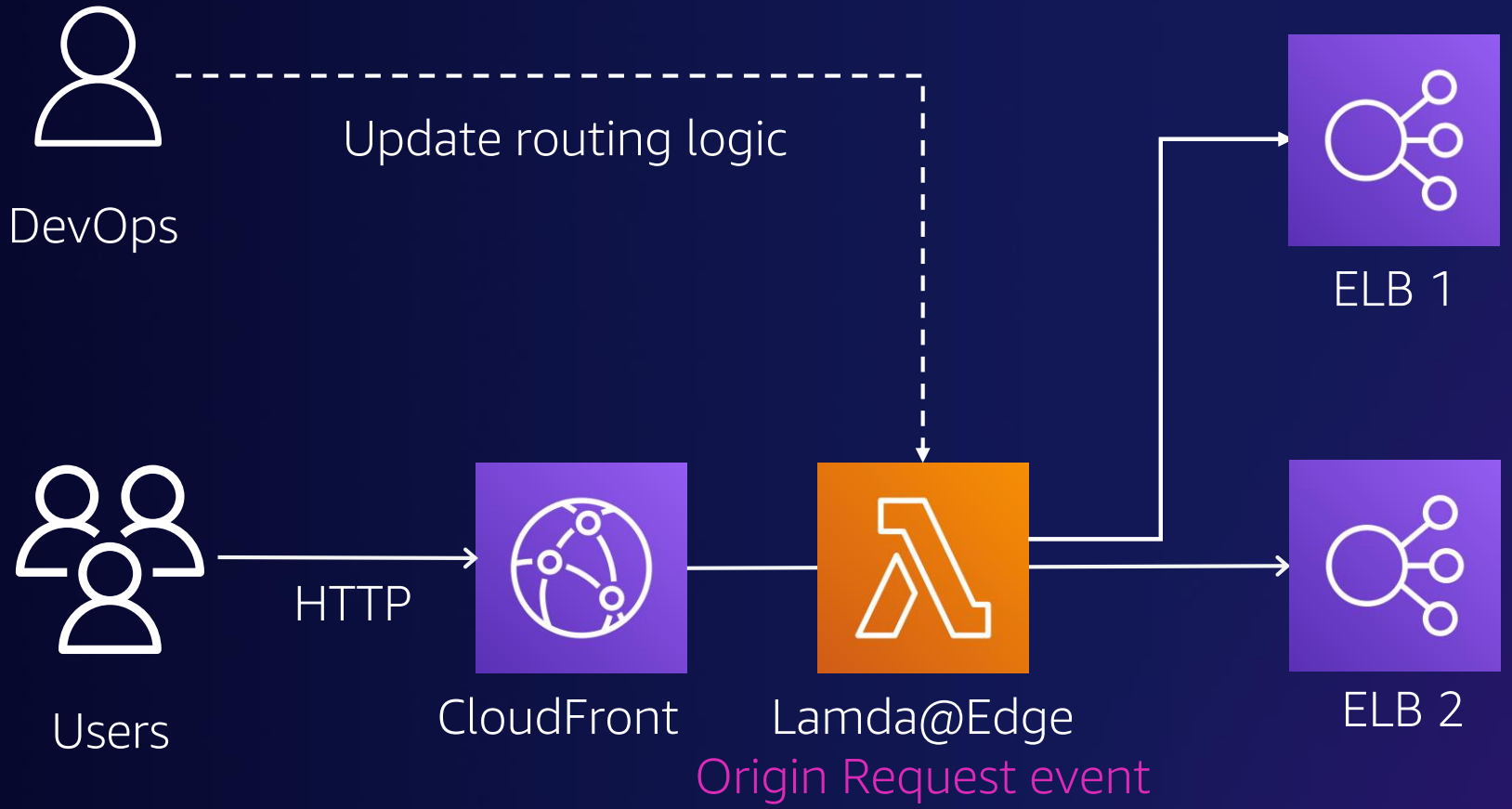
Summary of the Amazon Kinesis Event in the Northern Virginia (US-EAST-1) Region

November, 25th 2020

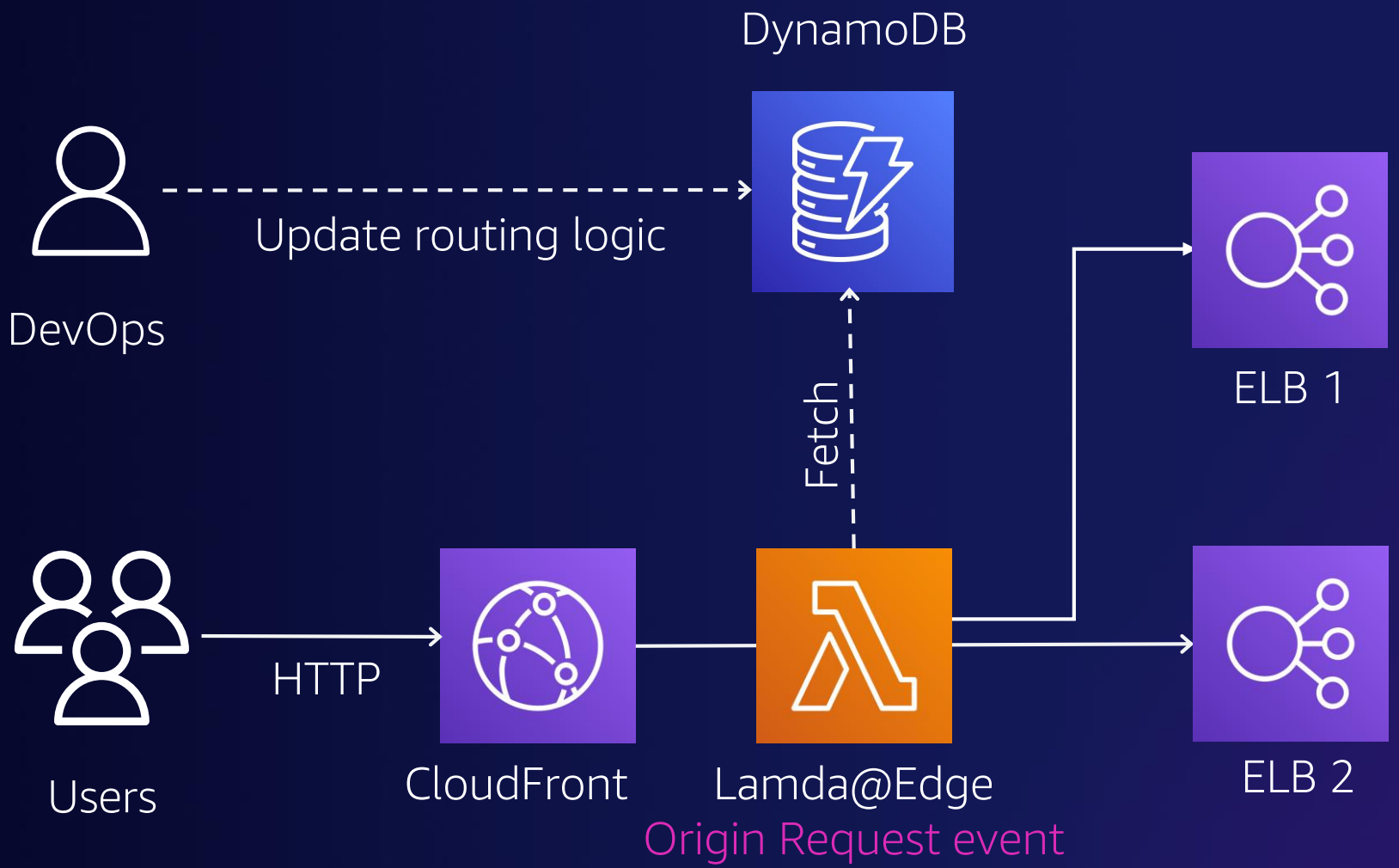
We wanted to provide you with some additional information about the service disruption that occurred in the Northern Virginia (US-EAST-1) Region on November 25th, 2020.

```
        <item>
         <title><![CDATA[Informational message: Change
Propagation and Invalidations Reporting Delay]]>
</title>
          <link>http://status.aws.amazon.com/</link>
          <pubDate>Wed, 25 Nov 2020 21:44:31
PST</pubDate>
          <guid
isPermaLink="false">http://status.aws.amazon.com/#cloud
front_1606369471</guid>
          <description><![CDATA[CloudFront Access Logs,
Metrics, and Reporting continues to be affected by the
Kinesis event but we are observing improving recovery.
CloudFront edge locations are serving traffic as
expected. Change propagation and cache invalidation
times are operating within normal time windows.]]>
</description>
        </item>
```

Case study: Dynamic origin routing



Case study: Dynamic origin routing

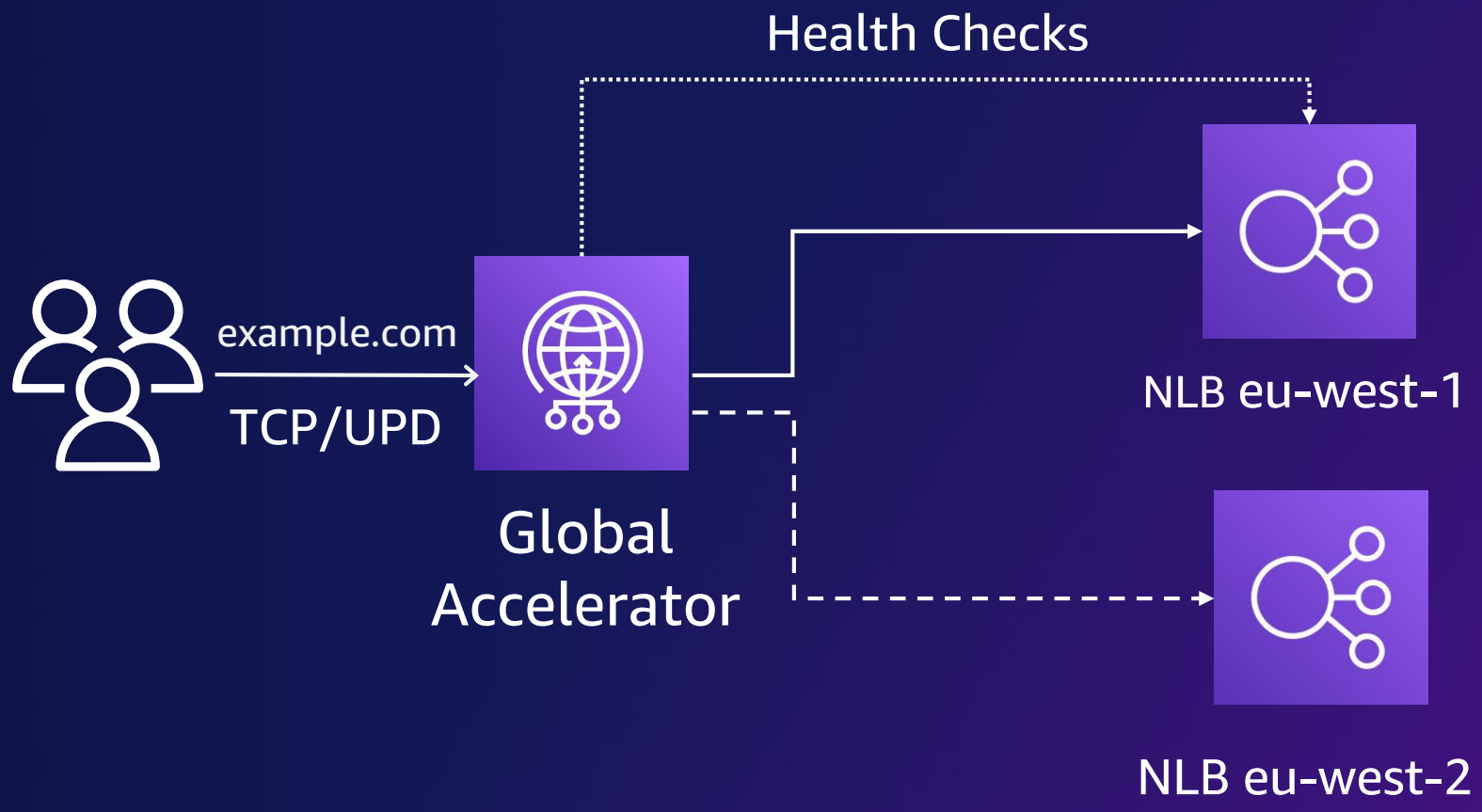


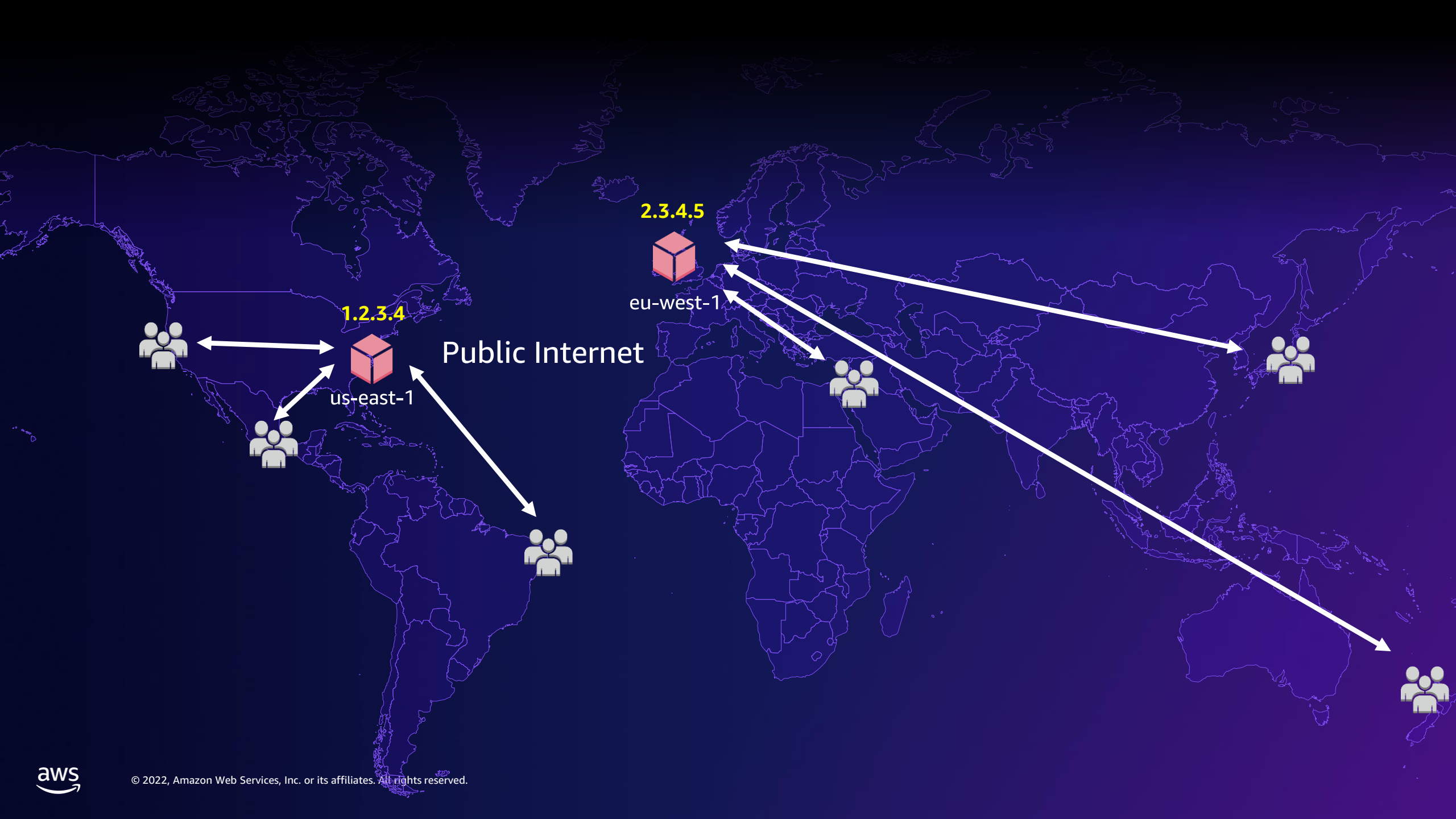
outsystems

Optimizing global performance

Origin Failover using Global Accelerator

- Stateful
- Works with non HTTP apps
- Failover in less than 30 seconds
- Premium DTO
- Works with EC2/ALB/NLB/EC2/ EIP





Performance

First byte latency (FBL)

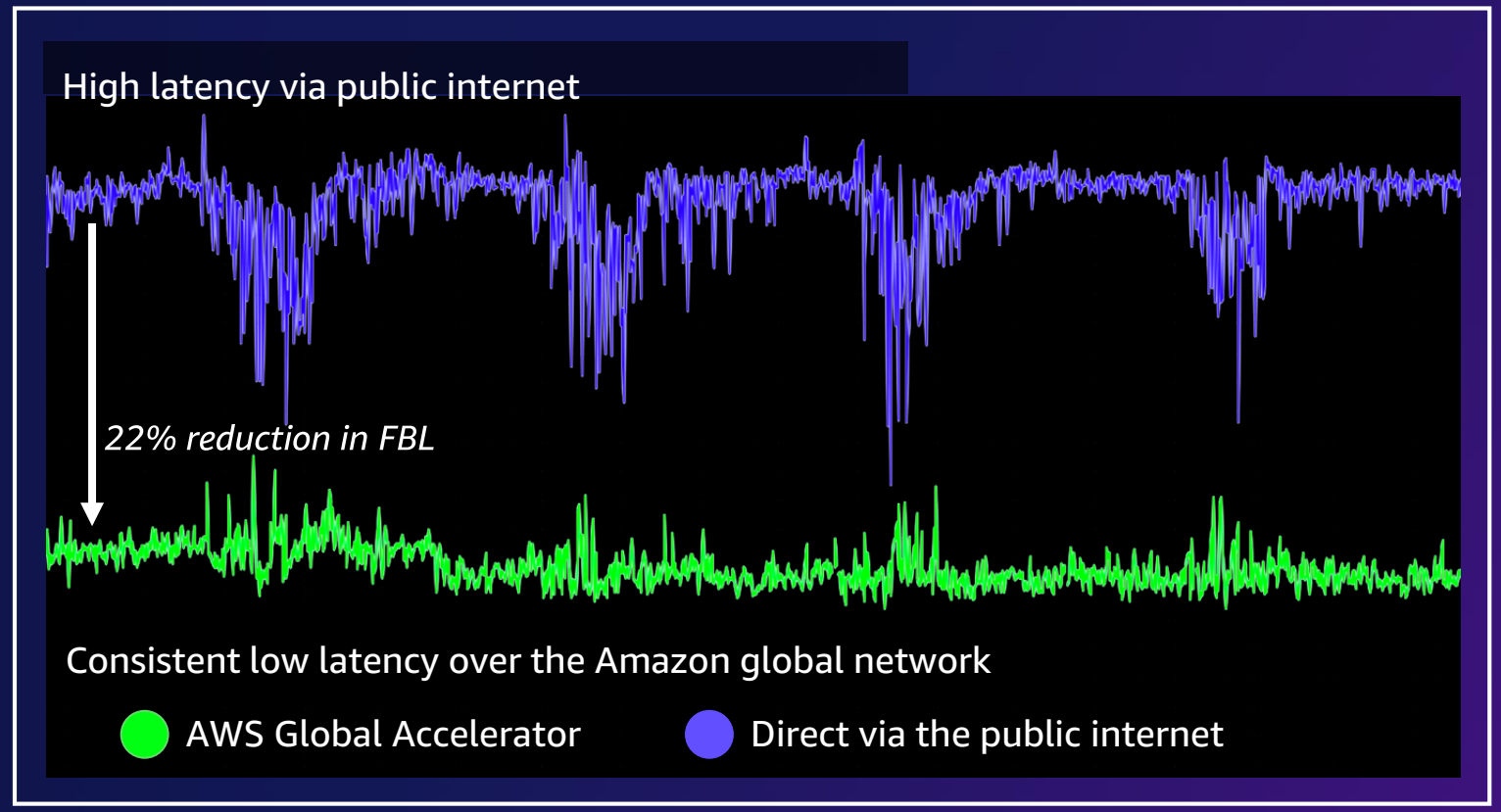
- Up to 49% better than direct *across continents*
- Up to 34% *within a continent*

Throughput

- Up to 60% better than direct *across continents*
- Up to 40% *within a continent*

Jitter

- Up to 58% better than direct *across continents*
- Up to 8% *within a continent*



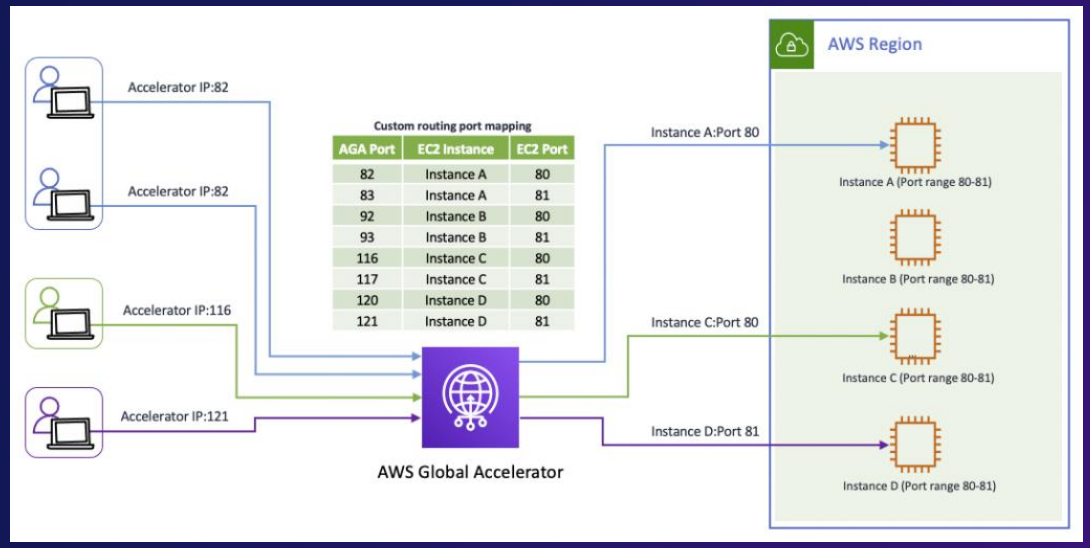
Third-party real-user measurement (p90) from users in Singapore to Ireland Region

Differences between CloudFront and Global Accelerator

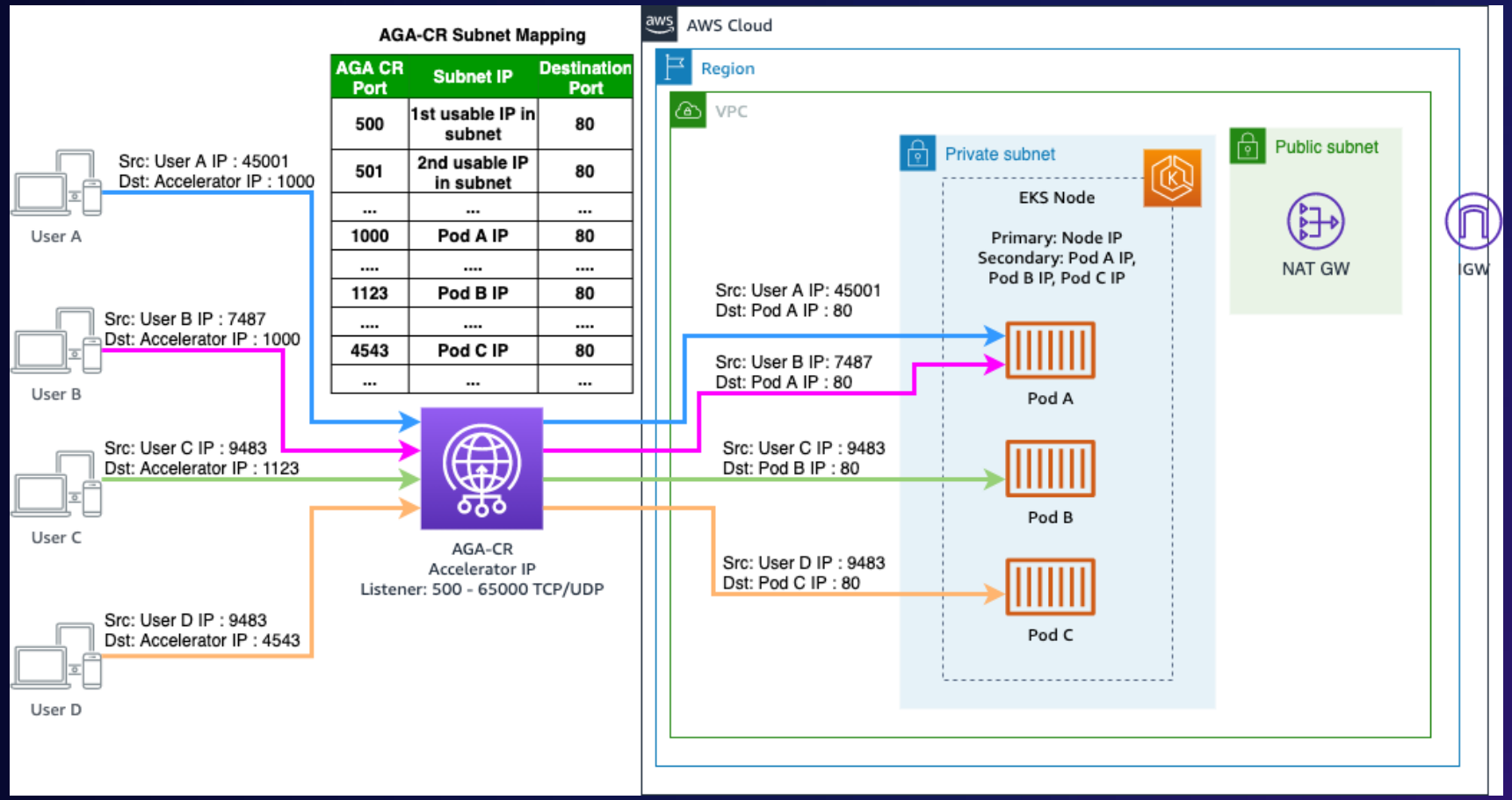
Key Features	Amazon CloudFront	AWS Global Accelerator
Description	Layer 7 HTTP/S content delivery network	Layer 4 TCP/UDP proxy OR Global traffic manager
Protocol Support	HTTP(S)	Any protocol running over TCP or UDP
Content caching	Yes	No
Routing	DNS-based	Anycast
IP addressing	Dynamic IP addresses, plus soon option to get fixed IP addresses (unicast IPs per PoP)	Two global static IP addresses, with ability to Bring Your Own IP address ranges
Failover	Native origin failover based on HTTP error codes or timeouts, or Route 53 DNS	Built-in origin failover in less than 30 seconds with no dependency on DNS TTLs.
Application hosting	Amazon S3 buckets, HTTP servers (for example, a web server), Amazon MediaStore, or other servers from which CloudFront gets your files	Application Load Balancers, Network Load Balancers, EC2 instances, and Elastic IPs

AWS Global Accelerator custom routing

- AWS Global Accelerator Customer Routing (AGA-CR) allows you to map AGA port to a specific VPC IP address and port
- On Elastic Kubernetes Service (EKS) the Pod Networking assigns a VPC IP to each Pod
- AGA-CR can connect to those Pod IP addresses as they appear as EC2 instances
- Need to disable EKS's default SNAT behavior
- The AGA-CR targets can be private only requirement is that the VPC has a IGW (doesn't need to be in any Route table)



AGA-CR and EKS

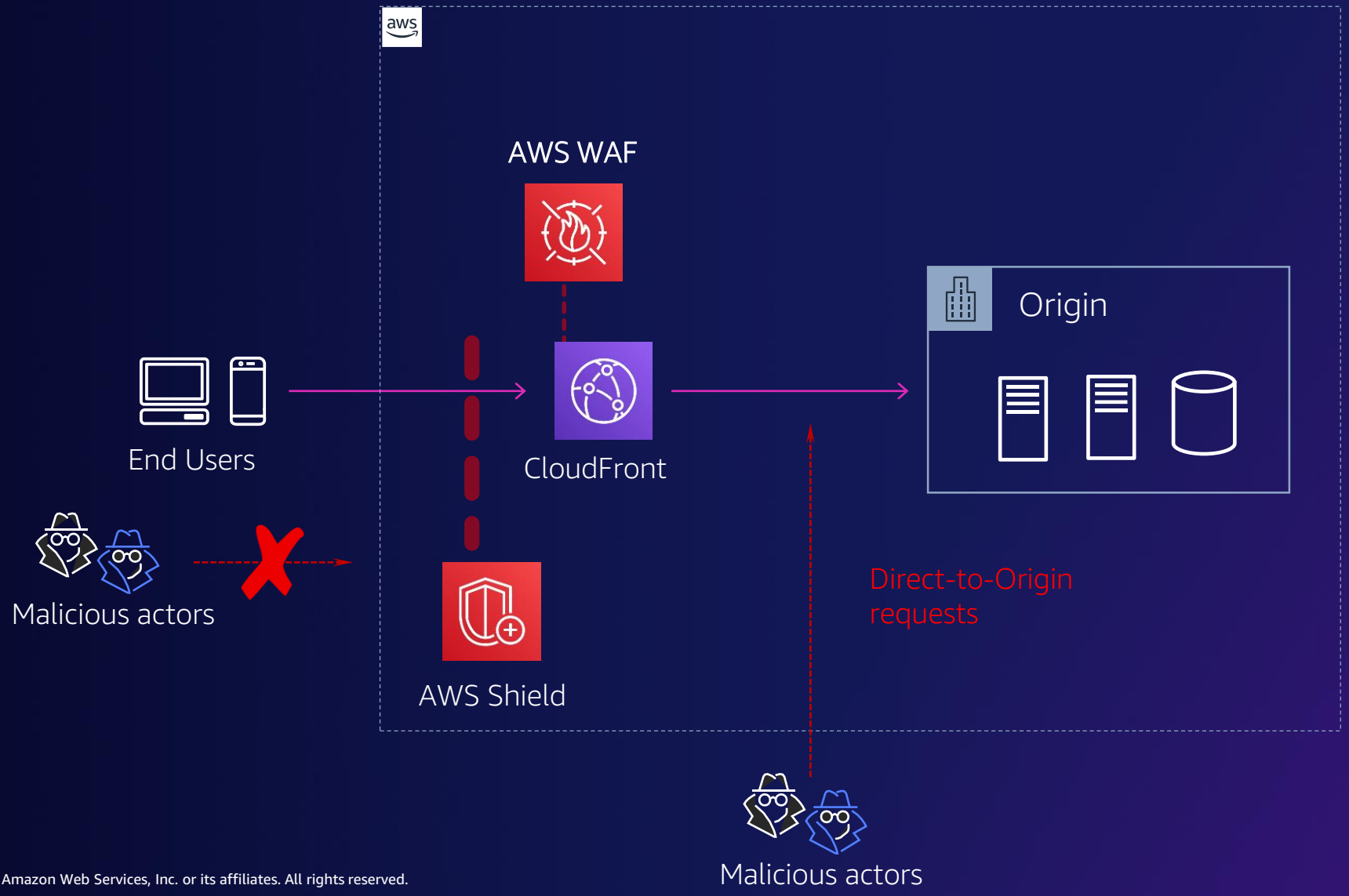


Blog Post Link

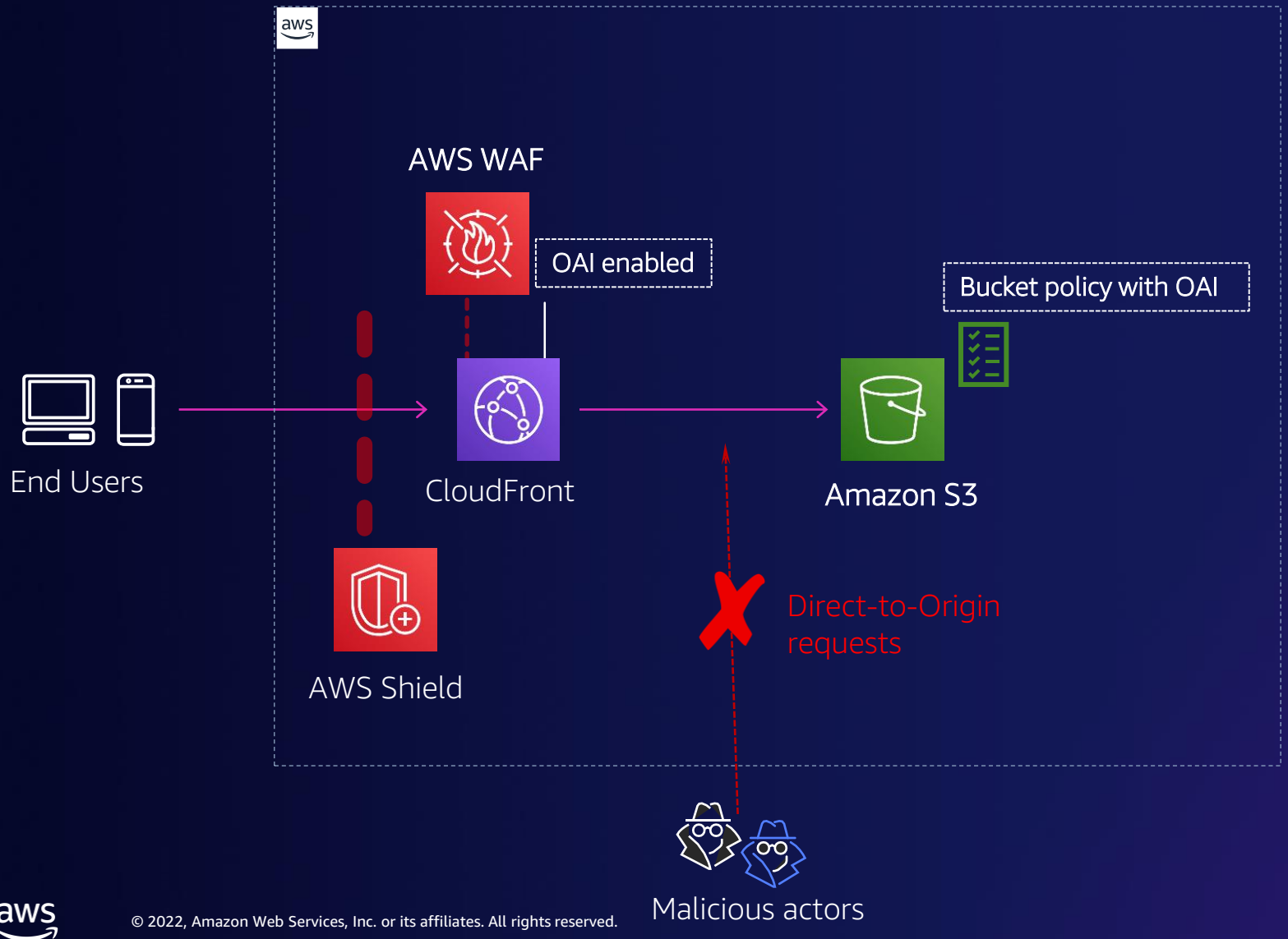
DEMO

Reducing your attack surface

AWS Well-Architected web applications



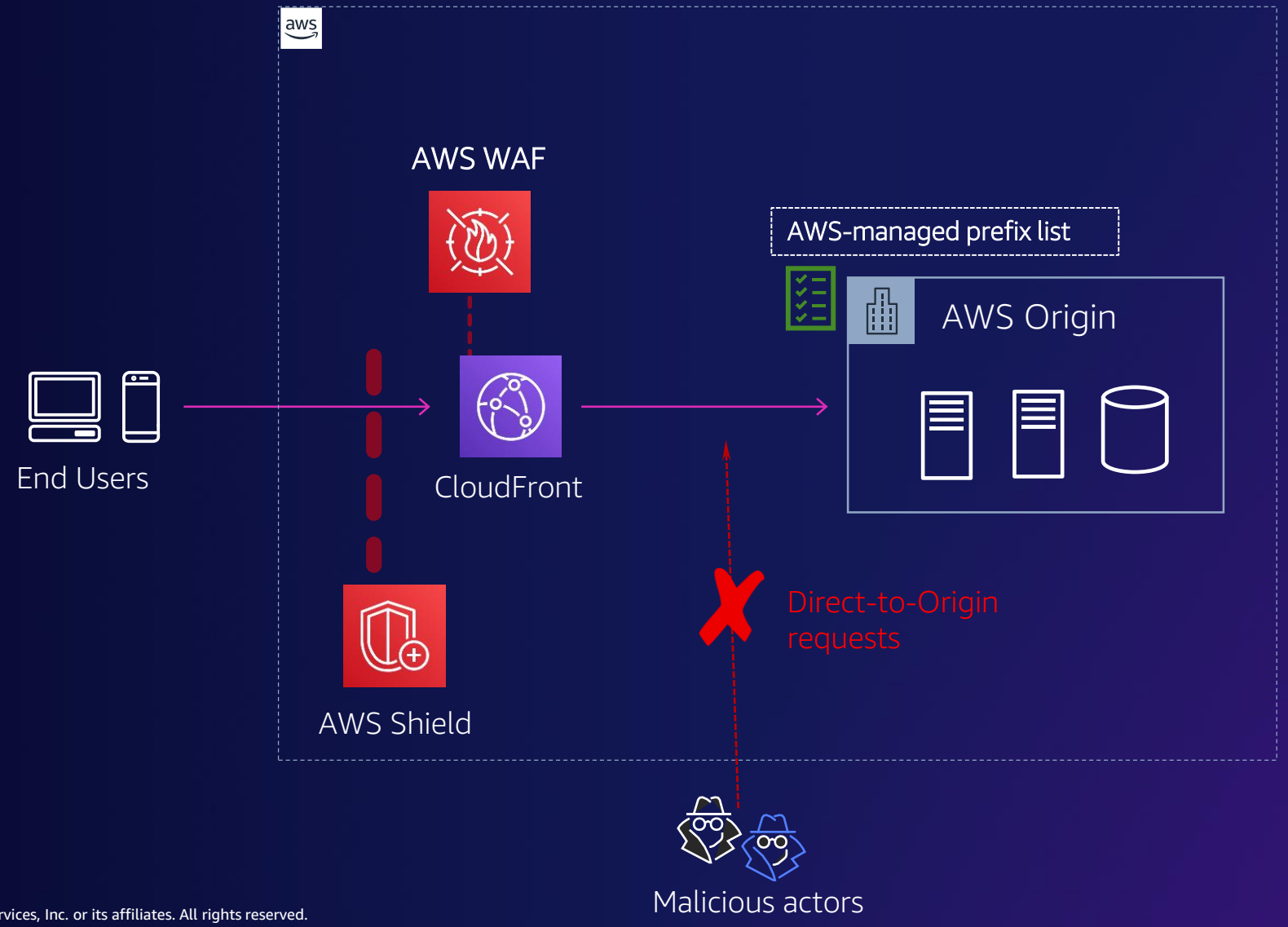
Securing access to S3 with OAI



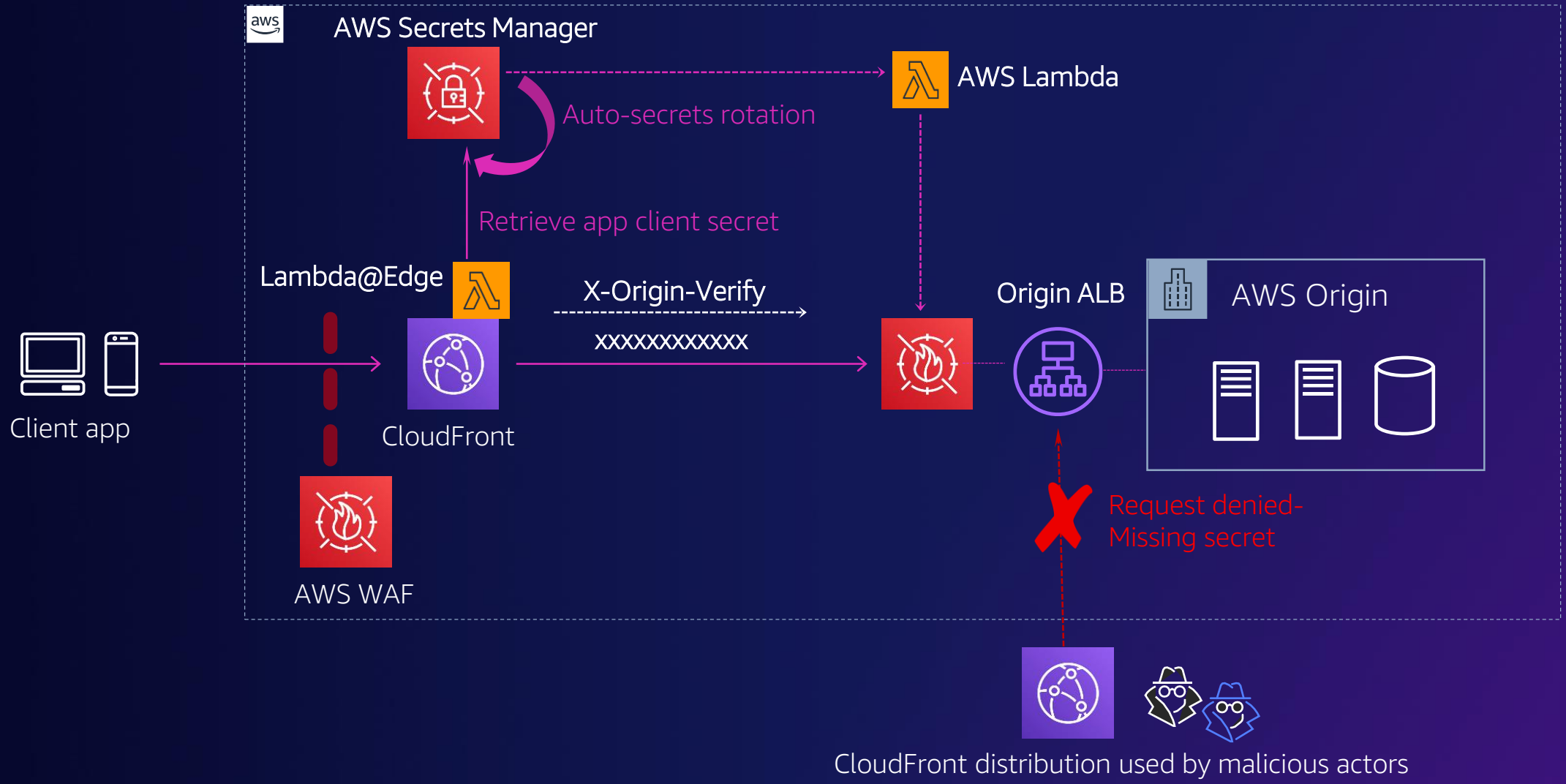
Origin Access Identity (OAI)

A virtual user identity that gives your CloudFront distribution permission to fetch a private object from Amazon S3.

Managed Prefix Lists for VPC based origins



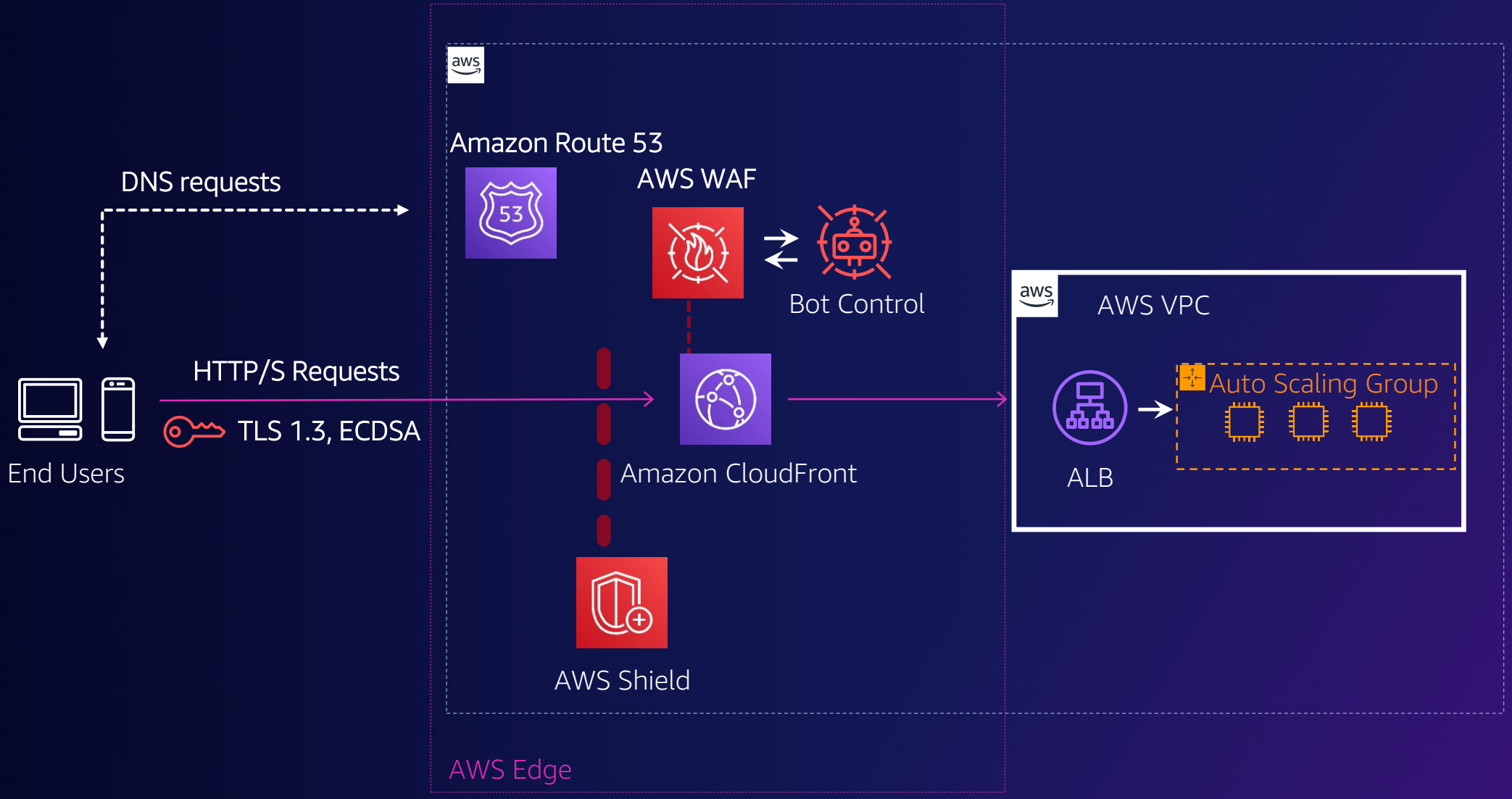
Hardening origin access control at layer 7



DEMO

Defending against DDOS attacks

Well Architected Web Application



Rate Limits

KEY PAGES

A large number of requests on a certain page could indicate a DDOS attack

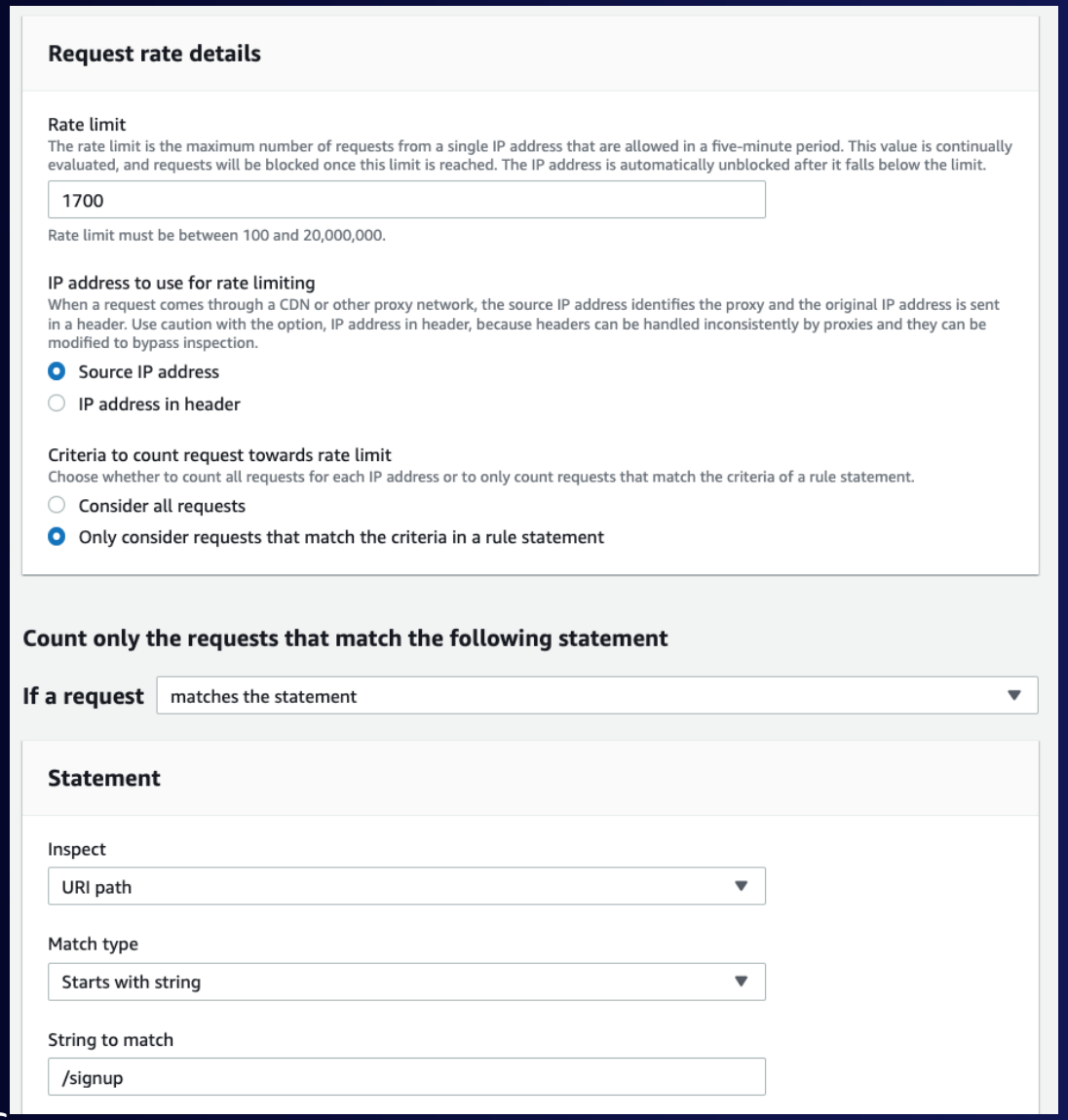
SUSPICIOUS IPs

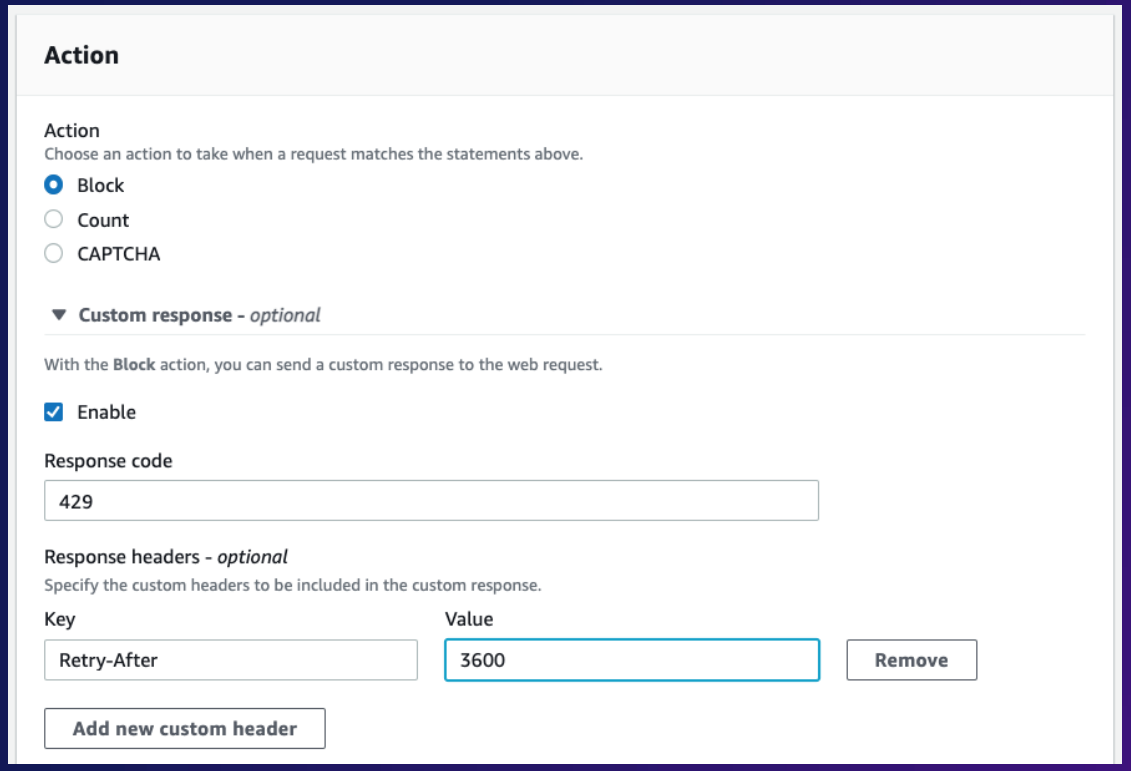
IPs that are identified in bad IP reputation lists are candidates for rate limiting with low thresholds if blocking is not desirable.

CATCH-ALL

A blanket rule that applies to the entire hostname to flag unusual spike in request activity

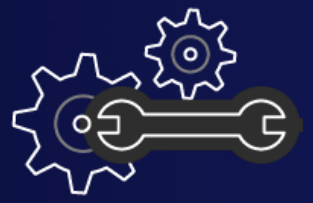
Conditioned Rate Limiting Rule Example





AWS Shield Advanced

Shield Advanced



Standard L3–L4 protection for AWS infrastructure

L3–L7 protection for your applications

Faster mitigation for your applications

Amazon CloudWatch event notification

DDoS Threat Environment Dashboard

24/7 access to AWS Shield Response Team (SRT)

Health-based detection

Adaptive L3–L4 protection

L7 anomaly detection with AWS WAF

Automatic application layer mitigation

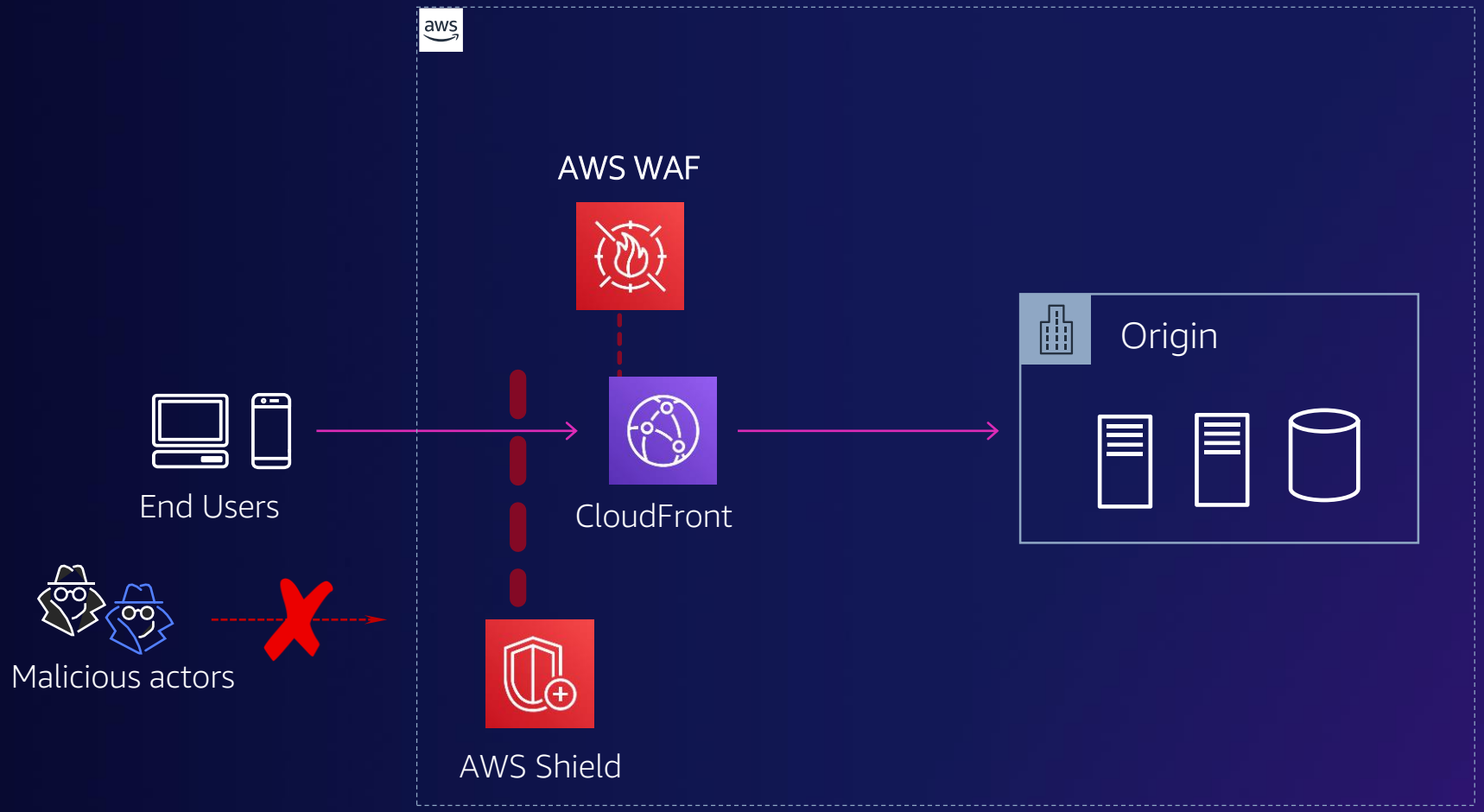
Proactive event response

No charge for AWS WAF
for Shield Advanced–protected resources

Central config and compliance
Firewall Manager cost included with Shield Advanced subscription

Cost protection for scaling during an attack

Automatic DDoS mitigation at layer 7



Take-aways

- Consider availability from different angles: infrastructure failure, malicious activity, software bugs, code changes…
- Architect your application for high availability, and be intentional about tradeoffs.
- Plan ahead your failures, because Everything fails, all the time
- Try out the origin failovers with AWS Disaster Recovery Workshop https://disaster-recovery.workshop.aws/

Learn in-demand AWS Cloud skills

AWS Skill Builder

Access **500+ free** digital courses and Learning Plans

Explore resources with a variety of skill levels and **16+** languages to meet your learning needs

Deepen your skills with digital learning on demand

Train now

AWS Certifications

Earn an industry-recognized credential

Receive Foundational, Associate, Professional, and Specialty certifications

Join the **AWS Certified community** and get exclusive benefits

Access **new** exam guides

Thank you!

Toni Syvänen

Please complete the session survey