*Living on the Edge for a Quarter Century:*

# An Akamai Retrospective

**Ramesh K. Sitaraman**
UMass Amherst & Akamai Tech

# The Edge: The Main Character of the Story

- Edge = servers deployed in clusters near internet clients (i.e., users) around the globe.
- Nearly all clients have a nearby edge server.
- Edge≠ Cloud

# The Akamai Edge Today

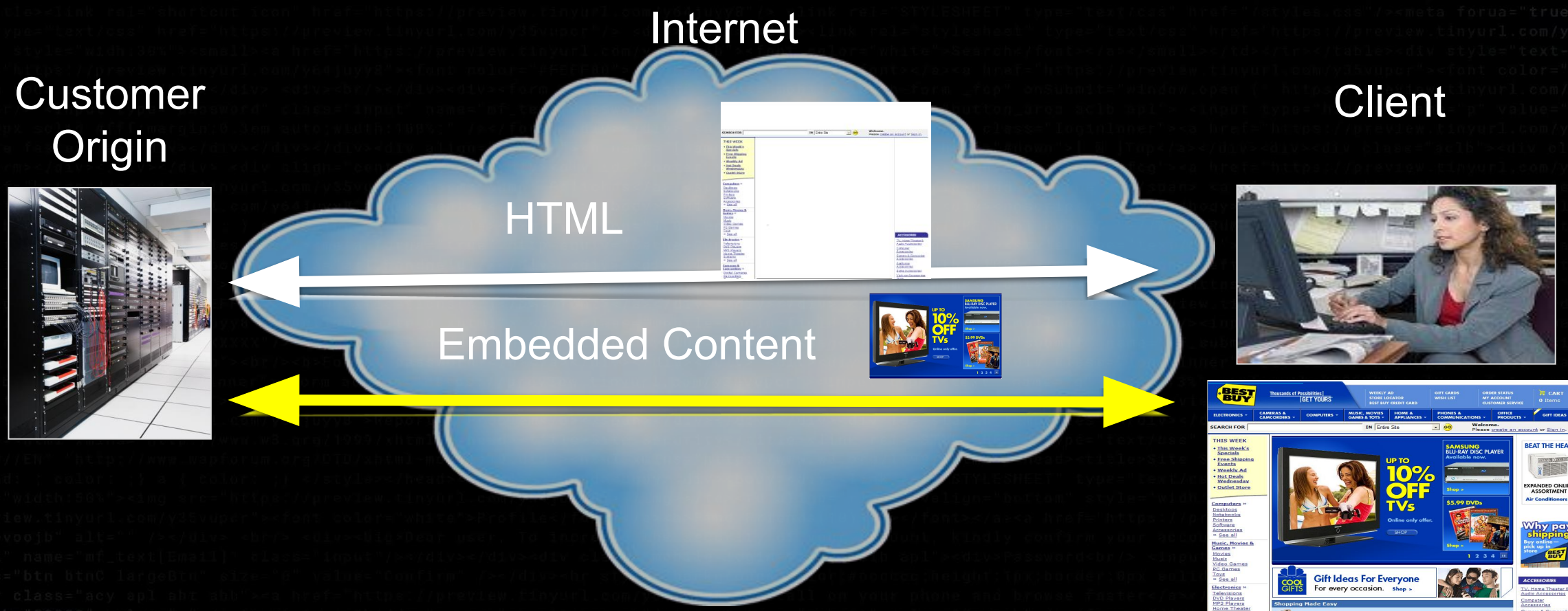
| | | | |
|---|---|---|---|
| **360K** servers | **100+** million hits per second | **7+** trillion deliveries per day | **175+** terabits per second *(250+ peak)* |
| **4,200+** locations | **1,350+** networks | **840+** cities | **135** countries |

# Story of the Edge in Four Parts

| Chapter 1 | Chapter 2 | Chapter 3 | Chapter 4 |
| --- | --- | --- | --- |
| Content Delivery | Edge Computing | Defending the Edge (and the Internet) | A Sustainable Zero-Carbon Edge |

# Internet Content Delivery Before Edge

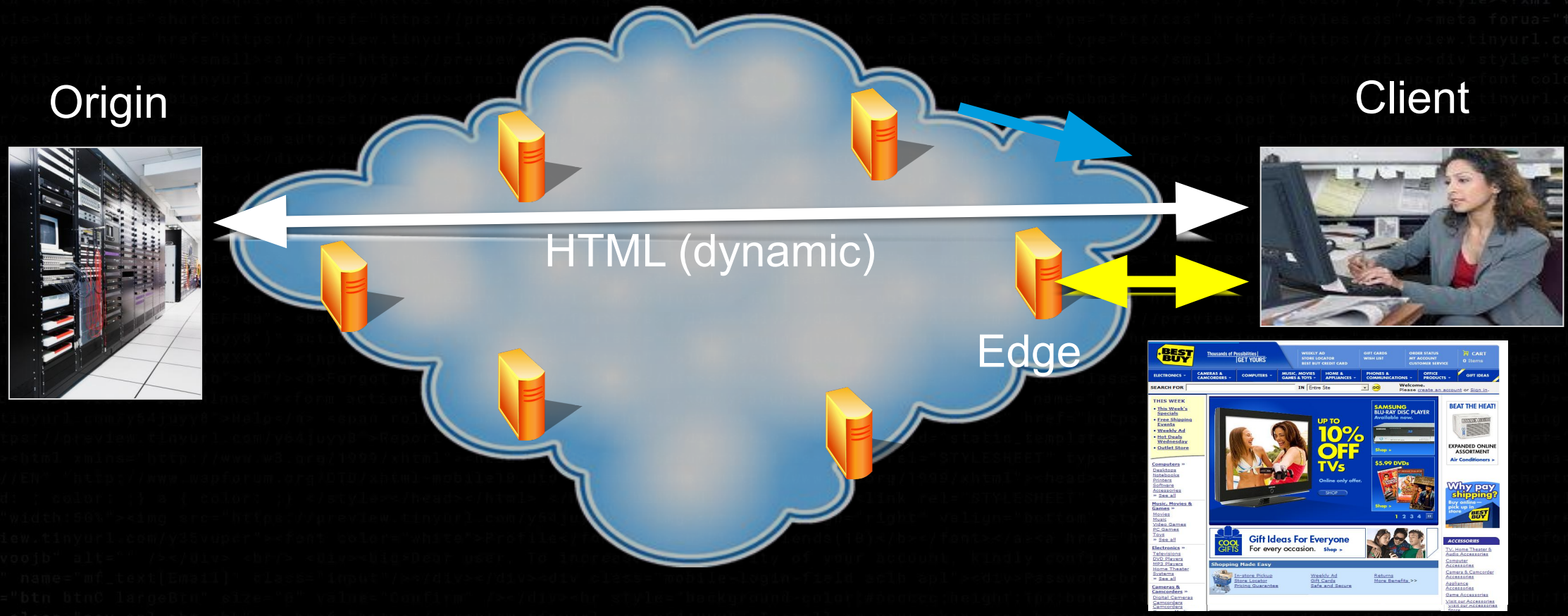
**Pre-1998**




- Origin = Content Provider, Application Provider, Video Provider, etc
- Client = Person/Device, Browser/Video Player/Application
- Poor Reliability, Performance, Scalability

# Edge born to serve static embedded content

**1998-99**




- Static embedded content cached & served from "nearest" edge server (mapping)
- Shorter Round Trips, Offload origin, Scalable

# Mapping: Finding the “nearest” edge server

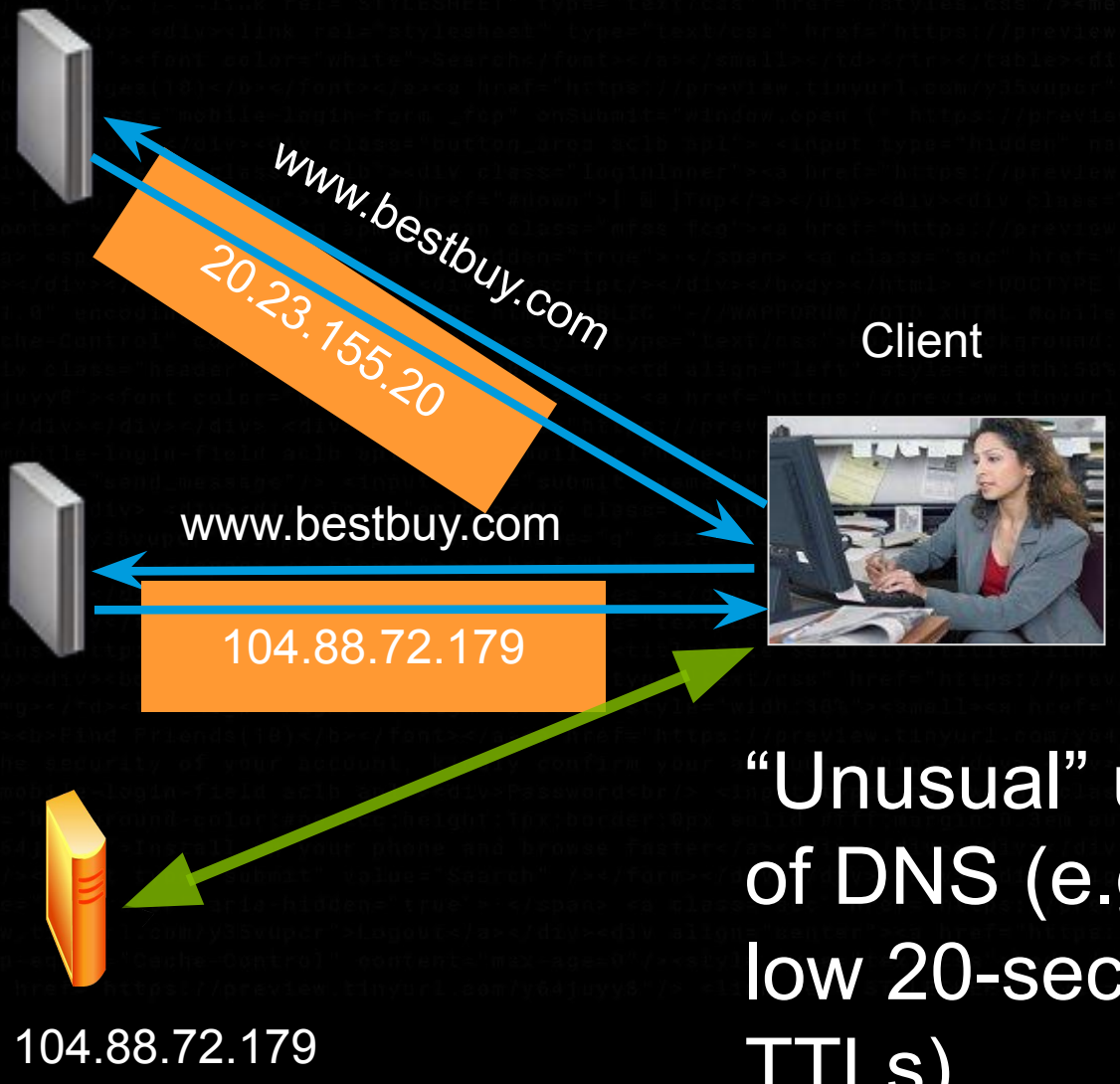
1998-99

Authoritative DNS Service*

(1) Top-Level DNS (Map-to-Cluster)

(2) Low-Level DNS (Map-to-Server)

(3) Nearest Edge Server




“Unusual” use of DNS (e.g., low 20-sec TTLs)

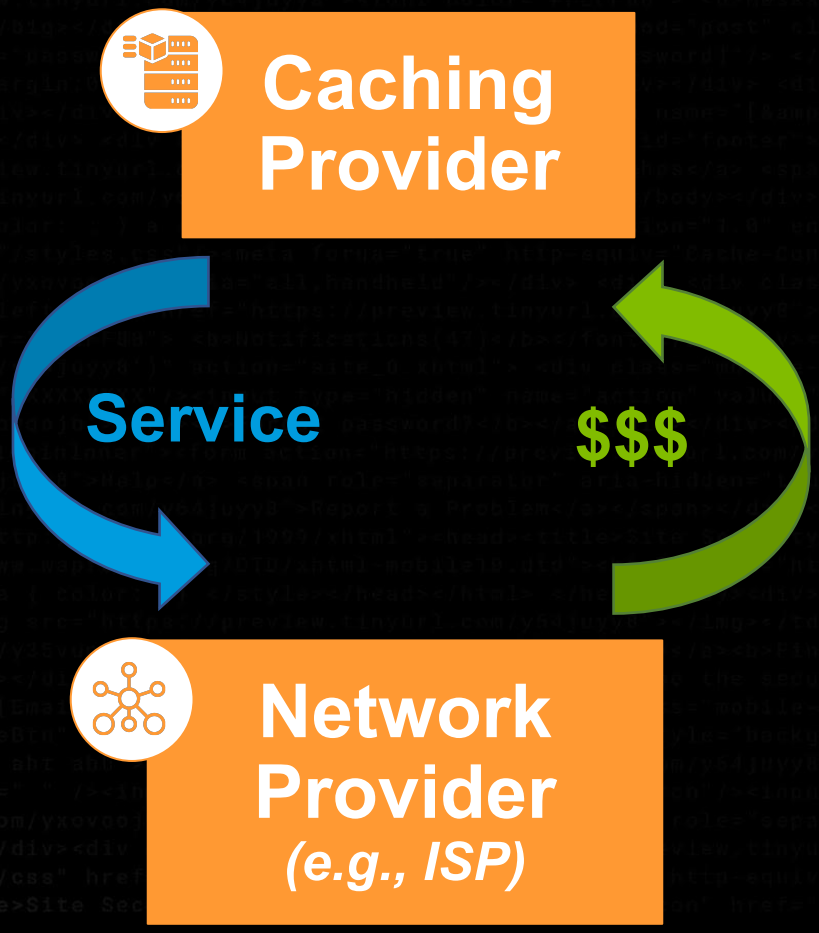
# Mapping drove major advances in Internet measurement

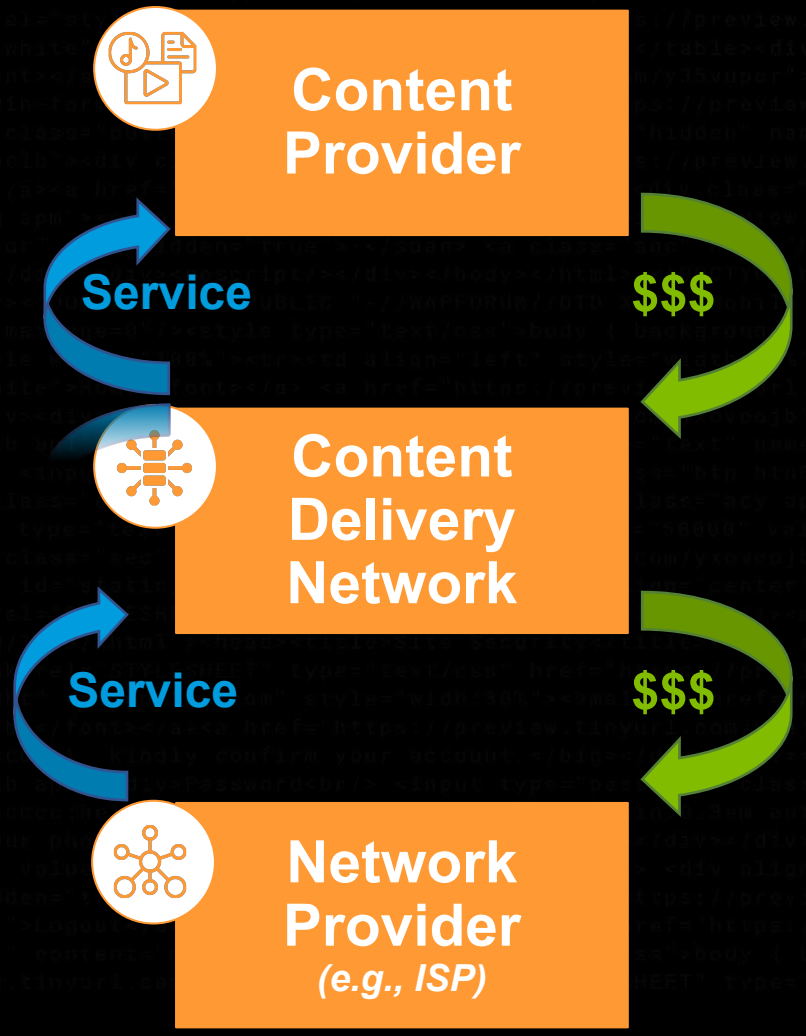
1998-2000

- IP intelligence (e.g., geolocation)
- Real-time Internet Weather
- Load balancing algorithms

End-User Mapping, ACM SIGCOMM, 2015
Algorithmic Nuggets, ACM SIGCOMM CCR, 2015
Akamai DNS, SIGCOMM 2020

# Two Caching Business Models: Caching Provider vs Content Delivery Network



# Content Delivery Networks as a Business Innovation

The New York Times

TECHNOLOGY; 2 Companies Take Separate Paths To Speed Delivery of Web Pages

Give this article

By Lawrence M. Fisher

April 17, 2000

Two reasons for success:

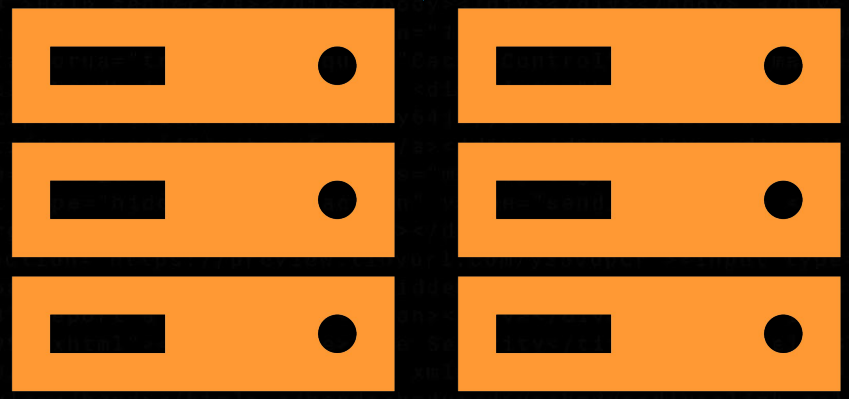
- CDNs turned the caching provider business model on its head so the content provider pays!
- Allowed CDNs to be more than just caching!

# The Swap Deal that Grew the CDN Edge

**1999-2000**

UPSTREAM



DOWNSTREAM TO CLIENTS

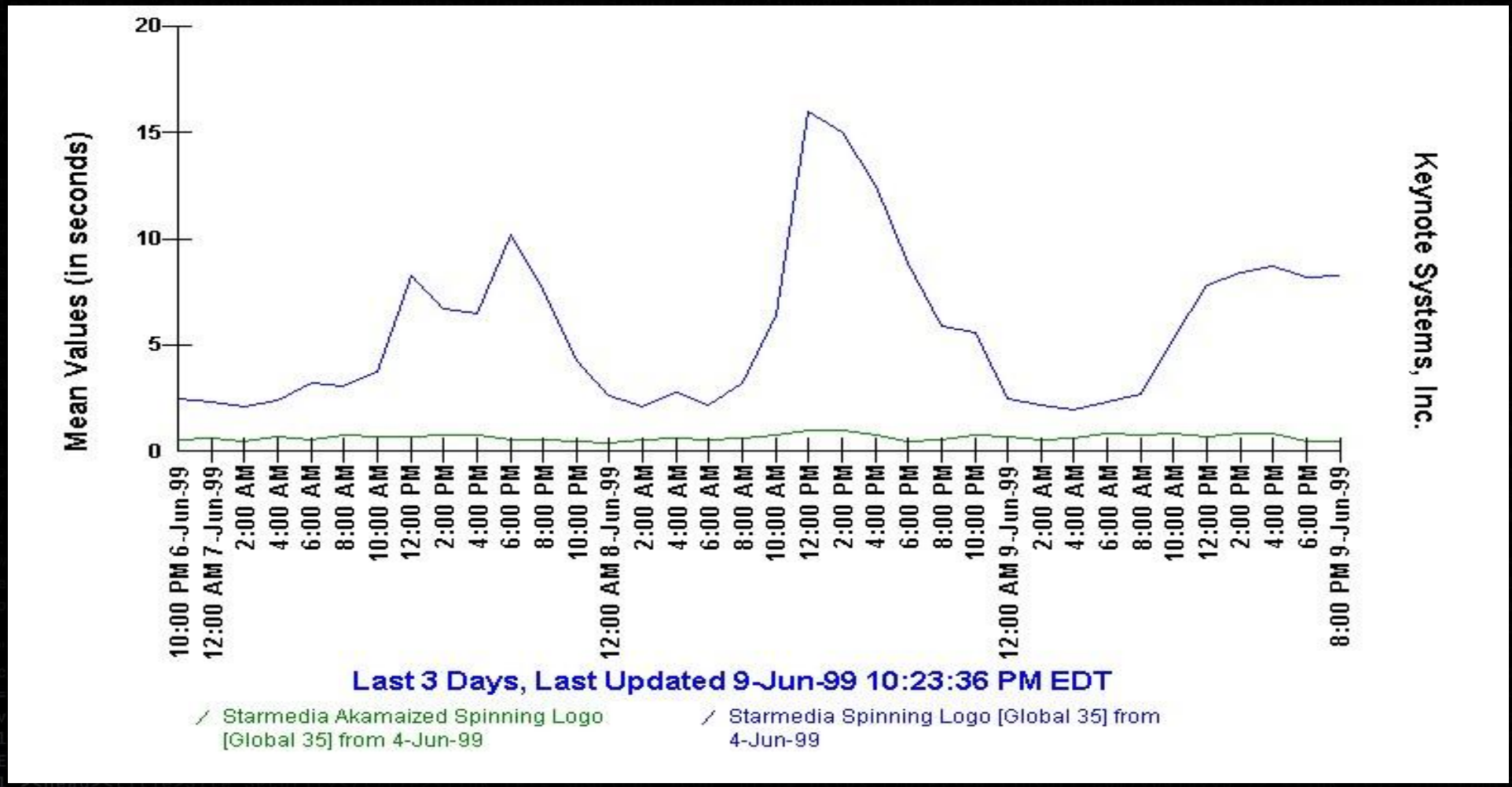
## “EYEBALL” Network

ISPs, Educational Institutions, Government

NETWORK provides free rackspace, power, and bandwidth

CDN provides better performance and reduced upstream bandwidth

# The Customer Trial: CDN speedup over Origin



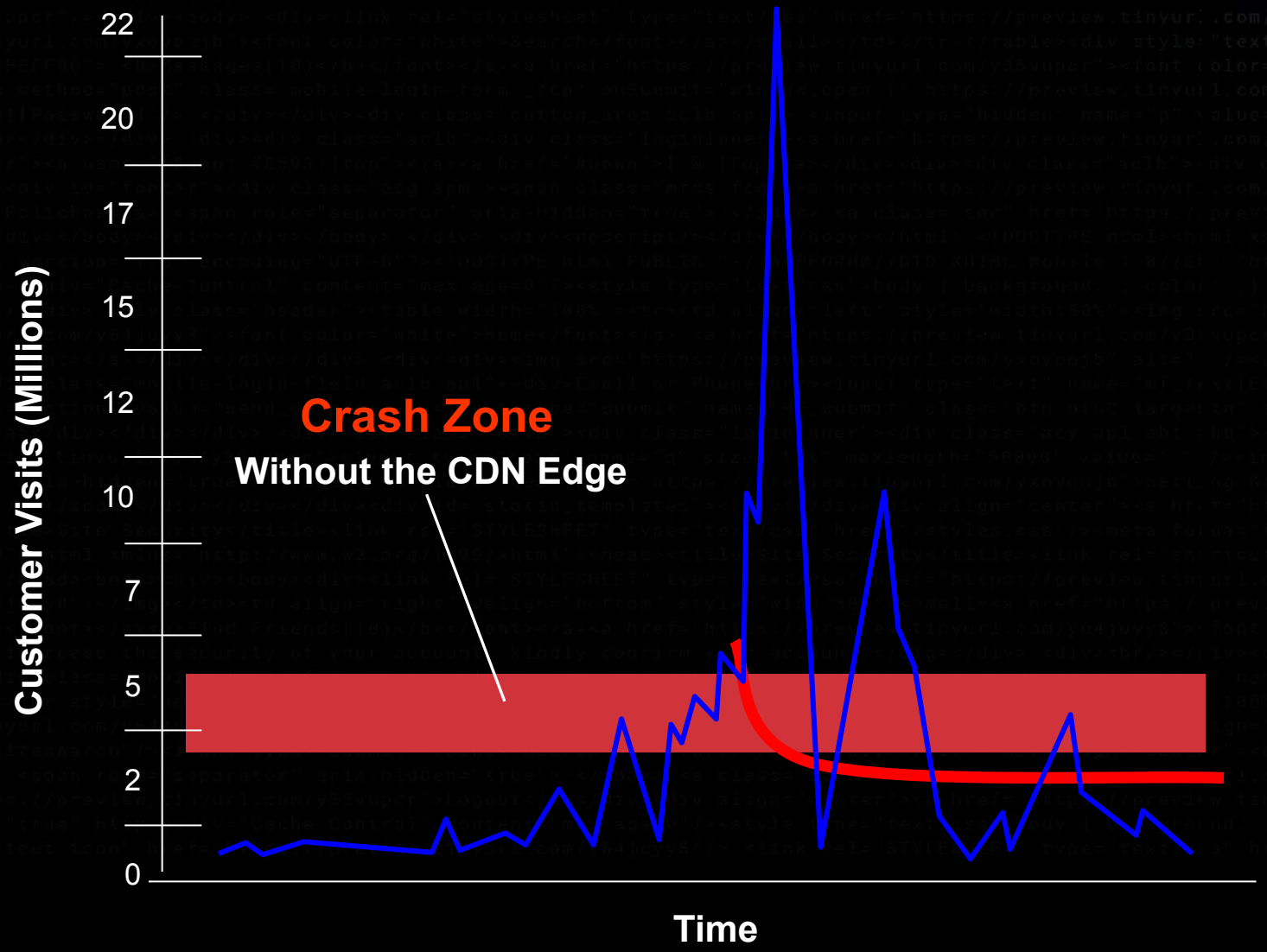
# Scalability Benefit: Edge can serve flash crowds

## The 2000 Presidential Election

The Washington Post

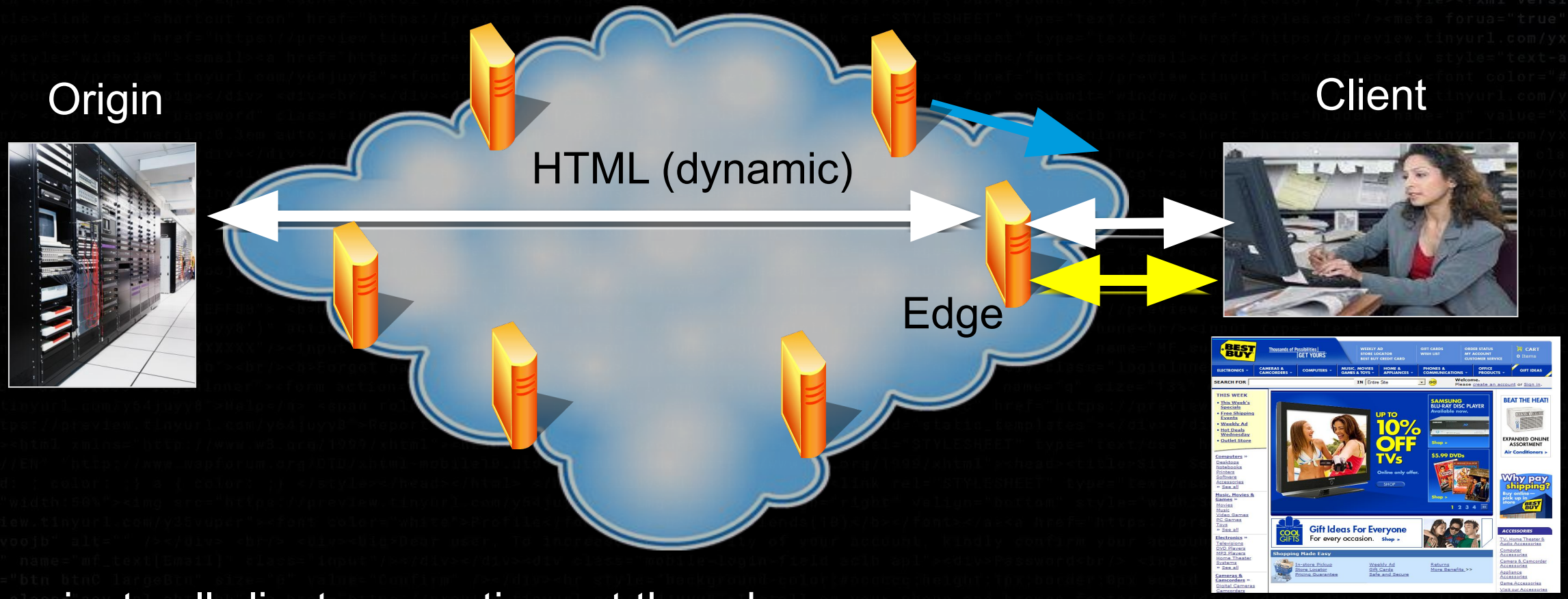
NEWS

**Akamai's FreeFlow gets Washingtonpost.com's vote**



# Whole-site Delivery: Serving Dynamic (Non-Cacheable) Content from the Edge

2000

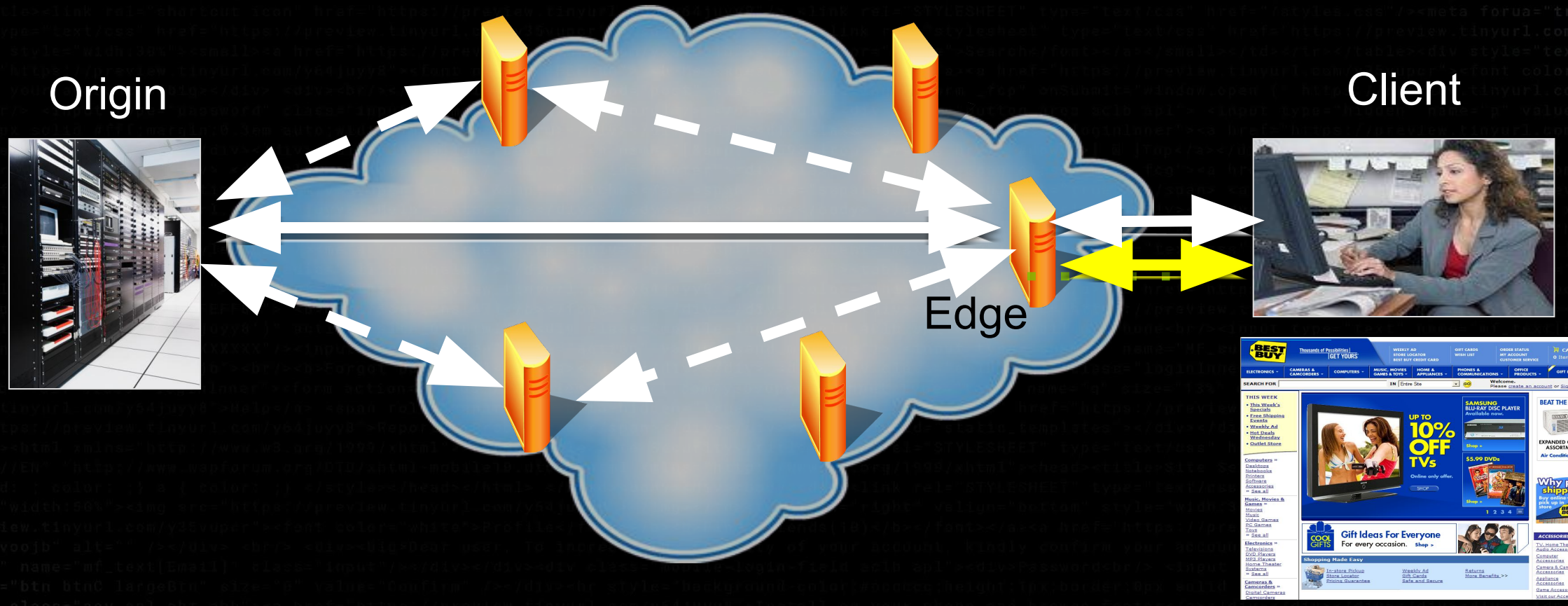



- Terminate all client connections at the edge
- Advantage: Persistent connections, Prefetch

# Overlay Routing: Transporting Dynamic Content from Core to the Edge

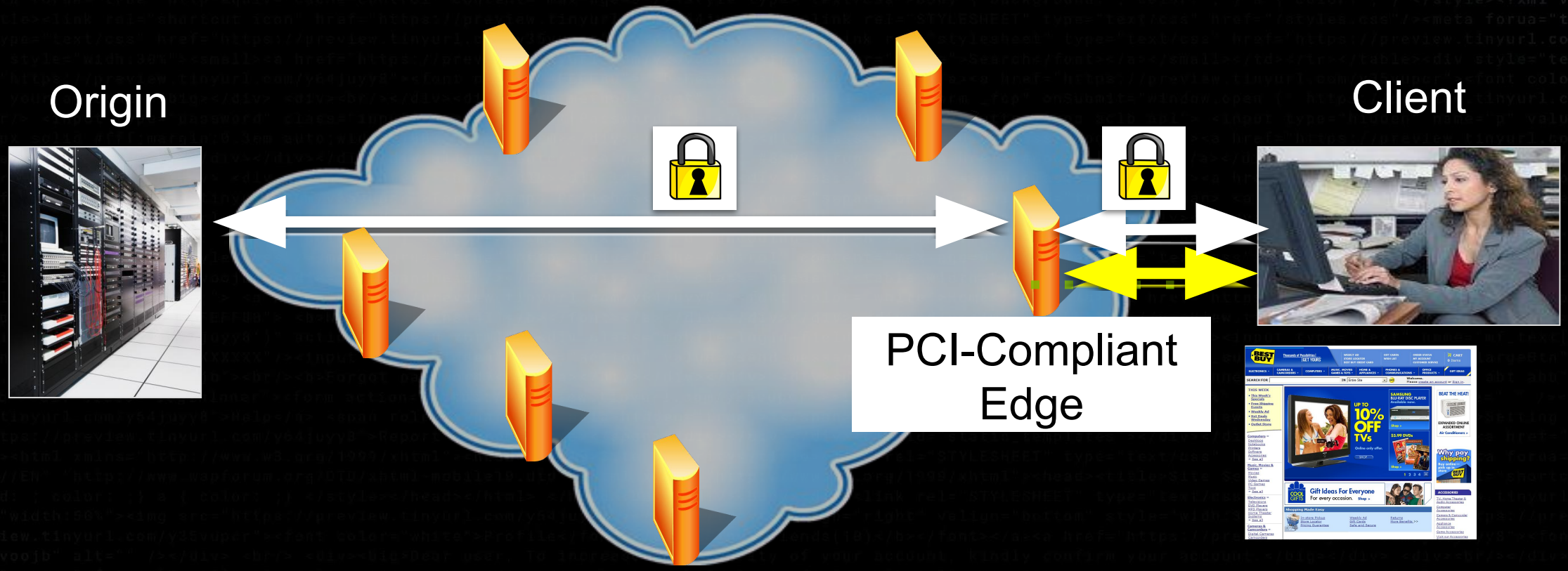
2000-01




- Alternate indirect paths (dashed) when direct path (solid) is down or slower
- Periodic races to determine best choice

# Secure Content Delivery Enabled Financial Services to use CDNs

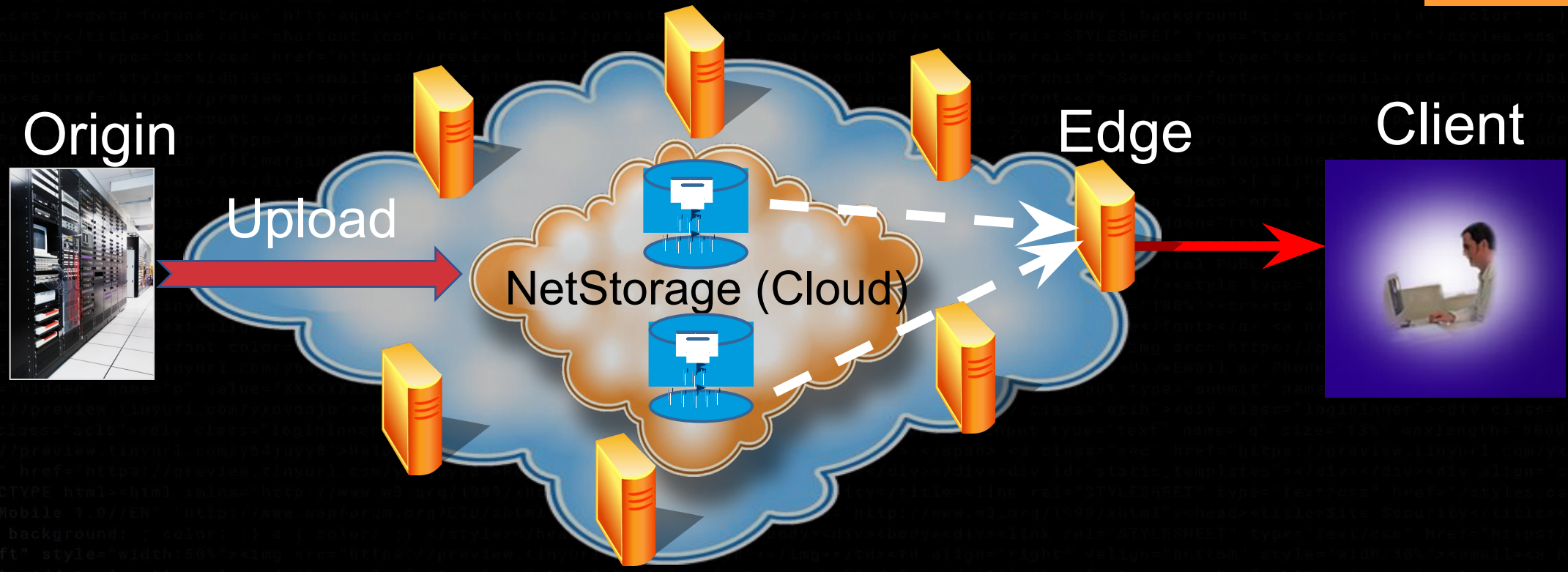
2001




- To terminate TLS/SSL connection, edge needs content provider's private keys/certs
- Physical: locked cages & cameras
- Software: Keys never written to disk, instant server wipe
- Infrastructure: Key management and audits

# First Major Video Delivery Platforms

**1998-2000**




On-demand: Early instance of the Cloud-Edge Model

Live: Multi-path Overlay Transport

Defined Performance: Availability, Startup time, Effective Bitrate, Rebuffers

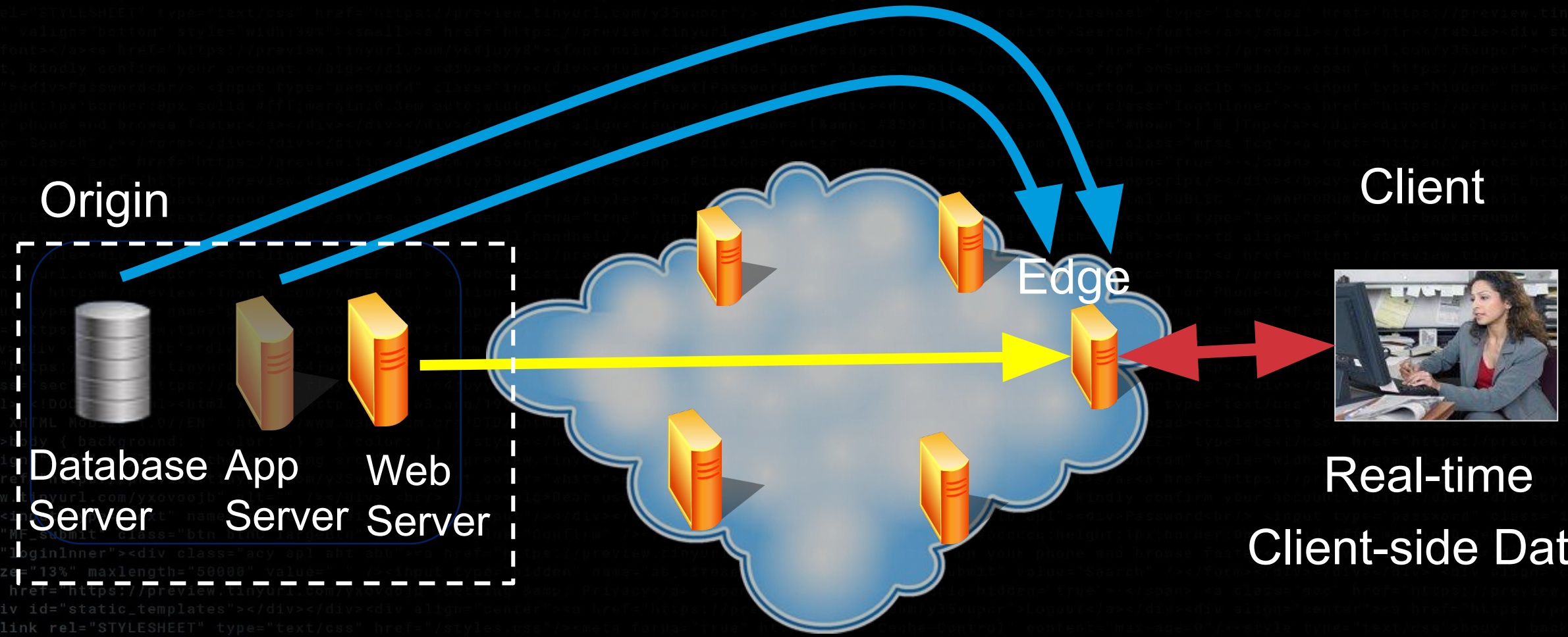
Later Years: Push to HTTP Streaming

# Story of the Edge in Four Parts

| Chapter 1 | Chapter 2 | Chapter 3 | Chapter 4 |
| --- | --- | --- | --- |
| Content Delivery | Edge Computing | Defending the Edge (and the Internet) | A Sustainable Zero-Carbon Edge |

# Edge computing was first created as logical evolution of content delivery

**2000-2002**



# Computing at the Edge using Metadata

2000

- Akamai Metadata = Turing-complete language executed at edge.
- Rules extract features of a user request and prescribe action.

```
<match:hoit result="true" host="devyn.works">
  <security:firewall.action>
    <msg>Test Rule</msg>
    <tag>%(SEC_CLIENT_FINGERPRINT_TLS_FACTOR_DEG_HASH)</tag>
    <id>60061385</id>
    <!-- advanced action reference  -->
    <action>%(WAF_CUSTOM_R60061385_ACTION)</action>
  </security:firewall.action>
</match:hoit>
```

Match client fingerprint (e.g., malicious user) and take action (e.g., deny access).

# Executing Metadata at the Edge

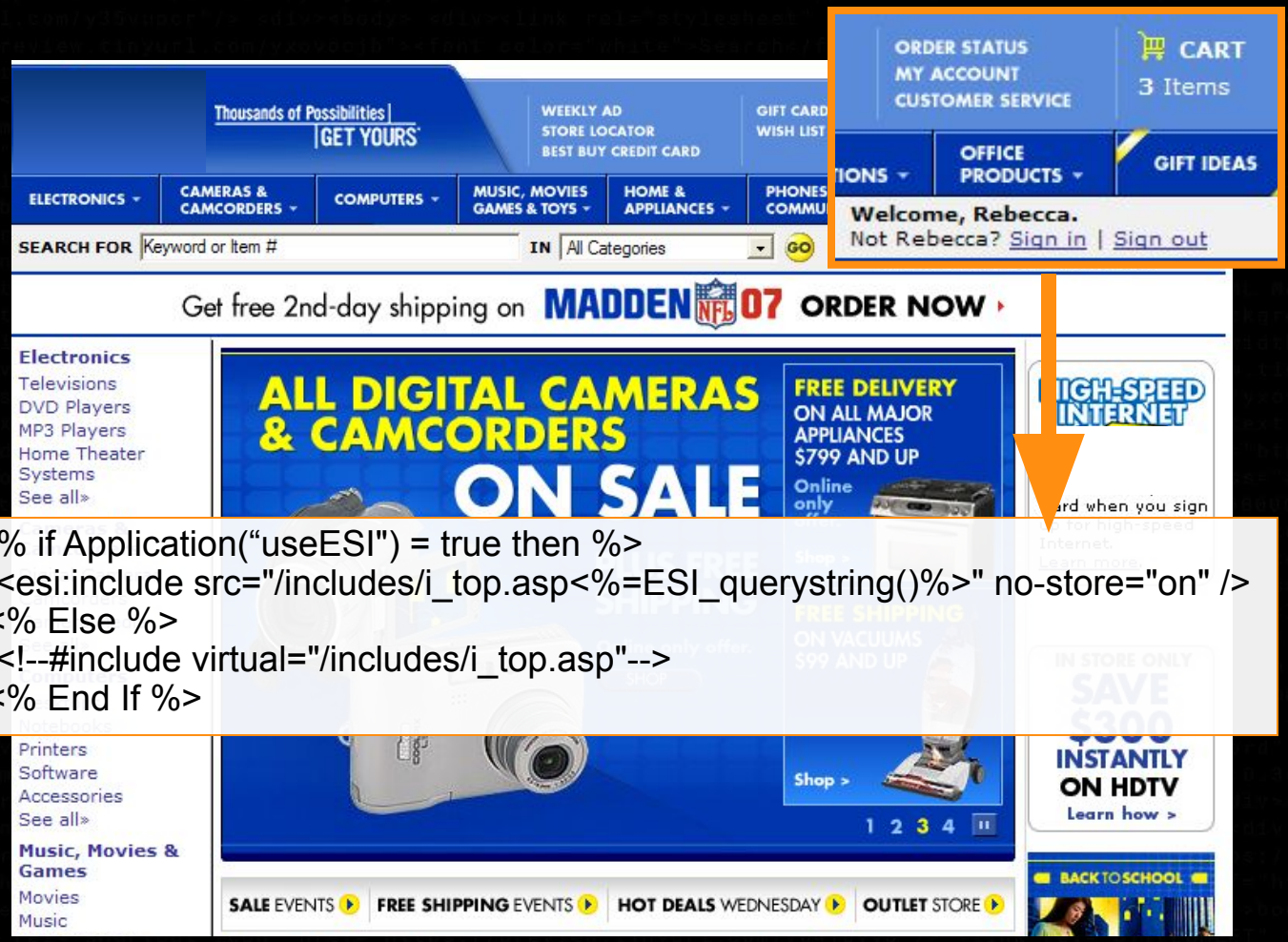
Thousands of customers program hundreds of thousands of edge servers with many gigabytes of metadata code per minute!

- Transform content
- Specify caching rules
- Allow/Disallow user access
- Encryption
- Ad Insertion
- Digital Rights Management

# Webpage Assembly at Edge using Edge Side Includes (ESI)

2001

HTML page cacheable except for one small piece of data, unique for each client, and request.



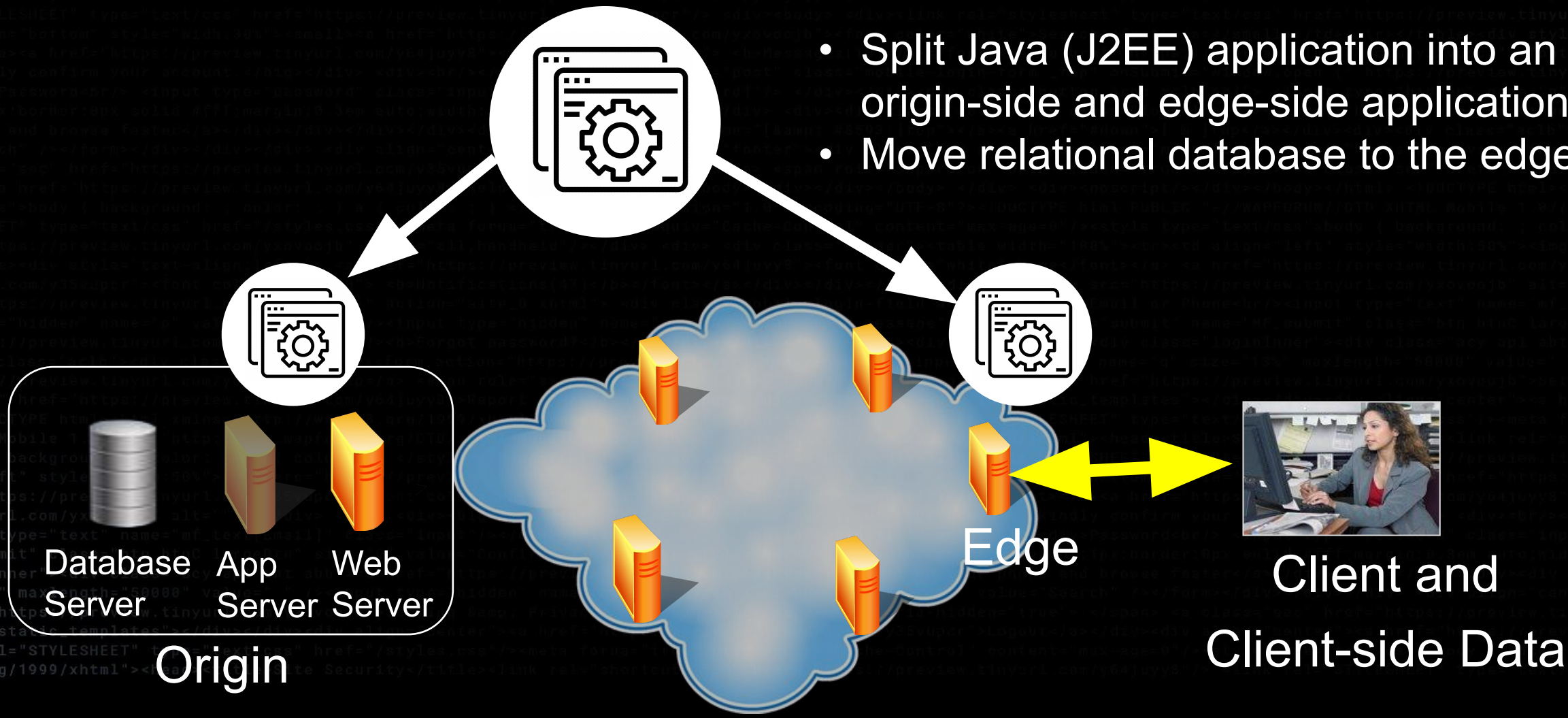

```
<% if Application("useESI") = true then %>
  <esi:include src="/includes/i_top.asp<%=ESI_querystring()%>" no-store="on" />
 <% Else %>
  <!--#include virtual="/includes/i_top.asp"-->
 <% End If %>
```

# A New Genre of Edge Service called Edge Computing™

2002

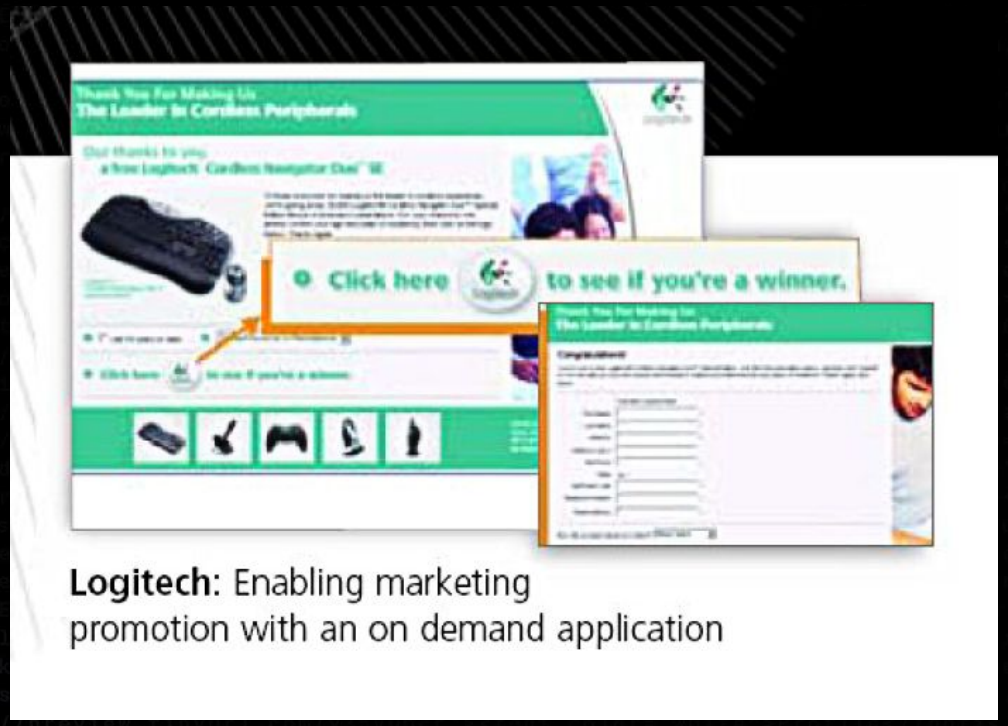
*Name Coined. Service launched: 2002. Akamai Trademark: 2002 - 2011*

- Split Java (J2EE) application into an origin-side and edge-side application
- Move relational database to the edge



# Contests and User Prioritization

2002




**Logitech:** Enabling marketing promotion with an on demand application

Five-hour Contest to win 20,000 cordless mouse and keyboards with 72 million participants.

Java app on edge decided winners with (rare) calls to an inventory database at origin

**Benefit:** Scalability on demand

# Mobile Applications on the Edge

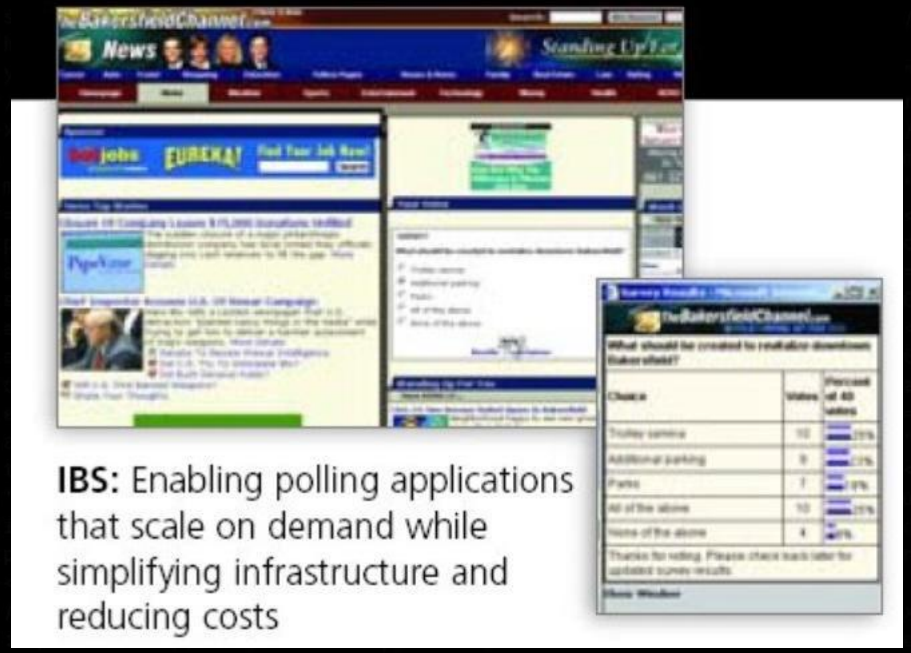
2002


A full suite of mobile applications: Phone configurator, shopping cart, dealer locator, etc.

Benefit: Offload computation from the origin and the (resource-poor) mobile client to the edge

# Voting & Surveys

2003




Users post feedback of TV shows in real-time to edge. Java app on edge aggregates data and sends summaries to origin.

Benefit: Moving computation close to client-side data for real-time distributed analytics

# Early Adopters of Edge Computing as of 2004



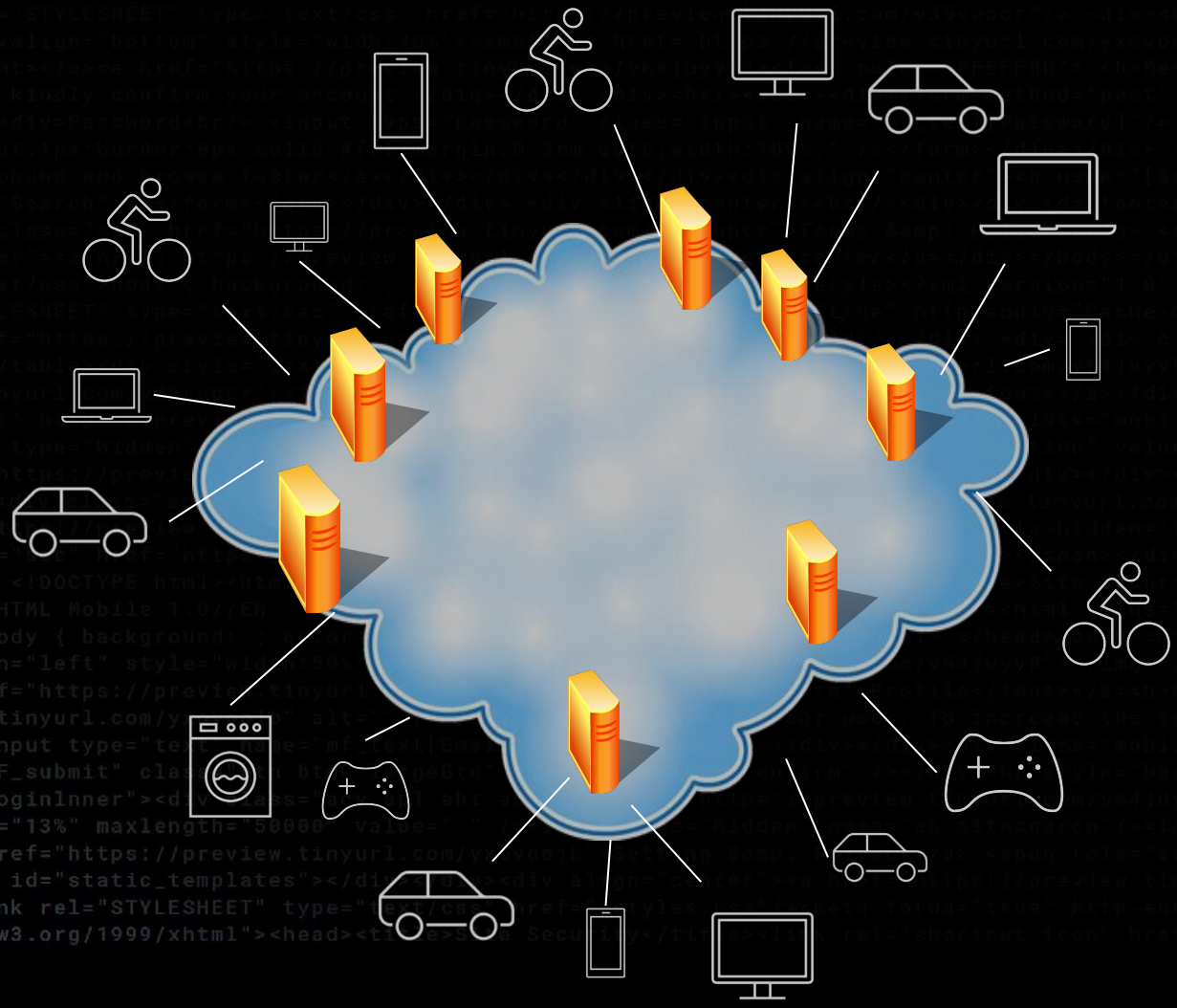
## Ideal Applications for Akamai EdgeComputing

Enterprise customers, partners, and suppliers expect immediate access to the information and tools they need. However, traditional centralized IT infrastructure is unable to support reliable high-performance delivery of critical Web applications that scale on demand—such as dealer locators, online contests, and customer registration, just to name a few. By extending your e-business infrastructure with Akamai EdgeComputing you can deliver your business-critical applications with speed and scalability to guarantee a positive online experience—without purchasing additional equipment.

Akamai provides an infinitely scalable framework for your applications. Take a minute to explore several types of applications that are particularly well-suited for the Akamai EdgeComputing Platform:

- **Applications with critical performance requirements**
  *Example:* Applications that provide critical customer service, such as store locators, product configuration applications, GUIs for data, and online tax form processing
- **Applications with unpredictable audience demand**
  *Example:* Wireless applications deployed to a global wireless network, weather applications, or applications whose audience is driven by mass-marketed events such as contests, TV promos, print, advertisements, or mail inserts
- **Applications with critical performance or scalability challenges**
  *Example:* Any application that is difficult to optimize for cost-effective scalability on a centralized infrastructure
- **Applications to be used repeatedly across multiple Web sites**
  *Example:* Pre-processors of reporting information, ad targeting engines, or presentation-layer logic
- **Applications that require 100% availability**
  *Example:* Information services for consumers, critical business facilitators, or address verification systems
- **Applications with an international audience**
  *Example:* Any application where the target markets are different countries or regions than the application's current centralized location
- **Applications which are CPU-intensive**
  *Example:* Applications that encrypt content, translate image formats, statistical visualization tools, or any Java-based application

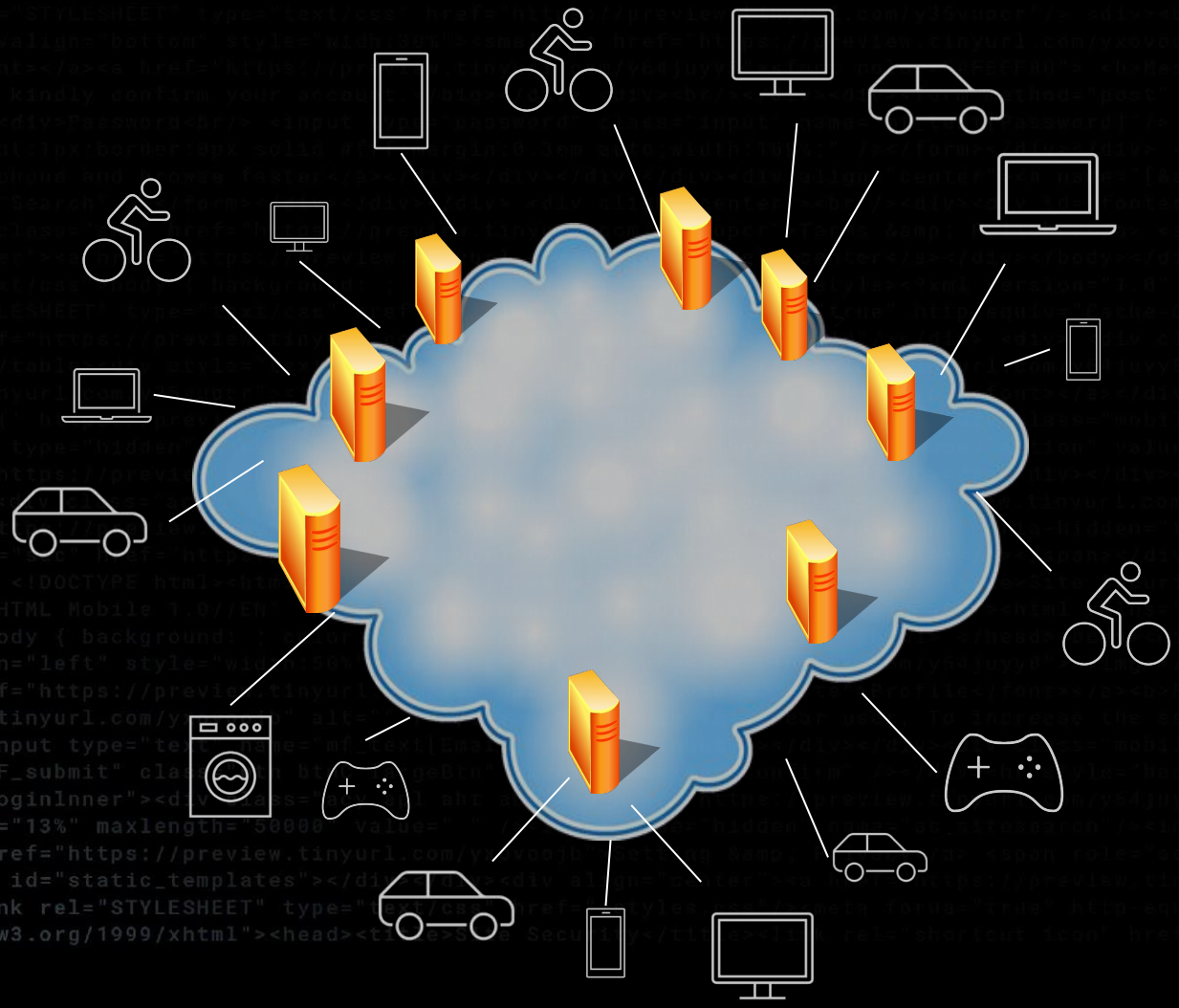
# The Modern Edge for IoT



More devices connected to edge than people: automobiles, appliances, medical devices, actuators, sensors, etc.

Edge servers in close proximity to devices. Example, 100+ Akamai edge locations in each major metro, 10's of msecs of latency.

# Edge as a Distributed MQTT Broker for IoT

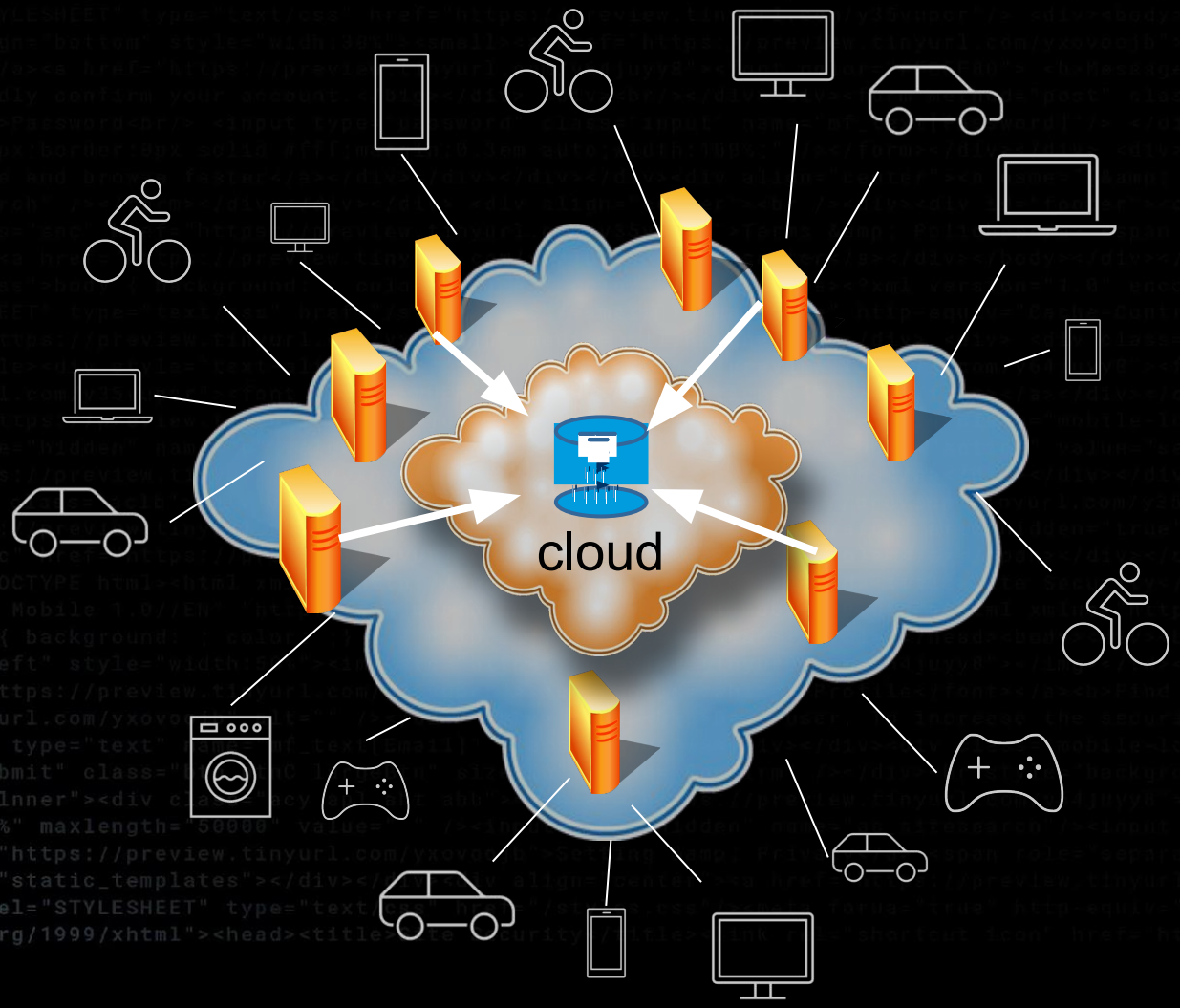


MQTT: Light-weight client-broker publish-subscribe messaging framework for IoT.

Edge (broker) facilitates regional, situational, real-time communication between IoT devices (clients).

Example: Cars communicating road conditions with other cars within 500m of each other

# Edge Analytics for IoT

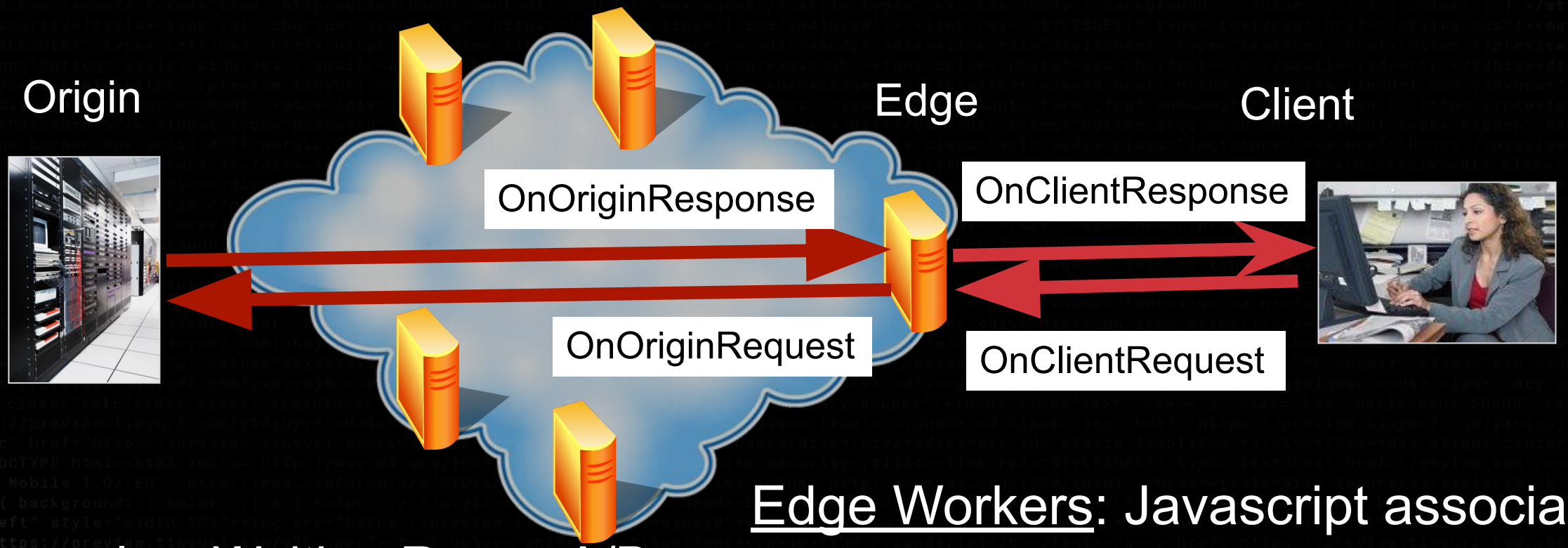



<u>Distributed Database on Edge:</u>
Buffer IoT message streams on edge for real-time/historical querying.

<u>Example:</u> Monitor automotive performance, reliability, repair schedules, warranties, etc.

# Serverless Computing at the Edge (Functions-as-a-Service)




Examples: Waiting Room, A/B Testing, Encryption, Bot Mitigation, Failover

Edge Workers: Javascript associated with events of request-response flow.
Edge KV: Distributed Key-Value Store

# Story of the Edge in Four Parts

| Chapter 1 | Chapter 2 | Chapter 3 | Chapter 4 |
| --- | --- | --- | --- |
| Content Delivery | Edge Computing | Defending the Edge (and the Internet) | A Sustainable Zero-Carbon Edge |

# Edge as the defensive moat for internet services

## (security & reliability & performance)

# Edge as the Shield of the Origin

2001


Access Control List (ACL) allows origin to only talk to Edge

Edge only allows traffic on certain ports (e.g., 80, 443).

IP/Geo blocking

# Past Decade: Exponential increase in attacks that aim to overwhelm and/or penetrate the edge

"Attackers exploit Spring4Shell flaw to let loose the Mirai botnet"

"Vulnerability impacting **Apache Log4j** discovered as the industry scrambled to mitigate and fix a severe zero-day Java library logging flaw dubbed Log4Shell."

"Channel Nine cyber-attack disrupts live broadcasts in Australia"

*"SolarWinds breach exposes big gaps in cyber security..."*

*"At Least 30,000 Orgs Hacked Via Holes in Microsoft's Email Software"*

"Vulnerability exploited in Log4j (open-source utility used widely in apps)"

**U.S. Colonial Pipeline**

*Forced shutdown after ransomware;*
*Gov. & infrastructure*

*"Sharkbot takes a bite out of the Play Store"*

**IoT Camera Breach**

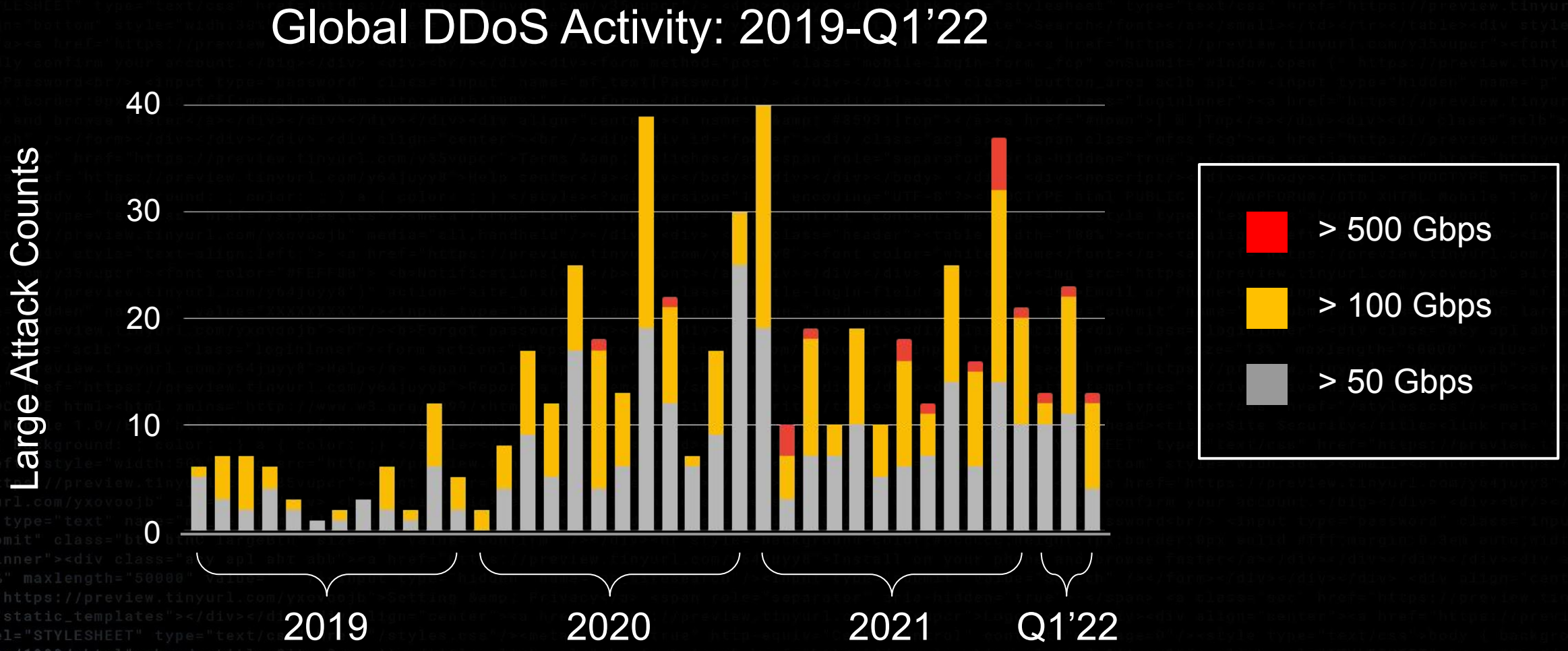
*150,000 smart cameras breached*

**"COVID-19 Pandemic Launches Cyber**

Costa Rica declares state of emergency over ransomware attack

Hackers Are Targeting UK Bank Clients With 2FA-Bypassing Toolkits

# Infrastructure Attacks: Volumetric DDoS on the Rise



# Cyber Extortion Ransom Note

*"Your whole network will be subject to a DDoS attack starting next week."*

*"We will refrain from attacking your network for a small fee. The current fee is 20 Bitcoin."*

*"If you decide not to pay, we will start the attack…"*

Subject: DDoS Attack

We are the Lazarus Group and we have chosen [redacted] as target for our next DDoS attack.

Please perform a google search for " Lazarus Group" to have a look at some of our previous work. Also, perform a search for "[redacted]" or "[redacted]" in the news. You don't want to be like them, do you?

Your whole network will be subject to a DDoS attack starting [redacted] next week. (This is not a hoax, and to prove it right now we will start a small attack on a few of your IPs from AS[redacted] block that will last for about 60 minutes. It will not be heavy attack, and will not cause you any damage, so don't worry at this moment.) There's no counter measure to this, because we will be attacking your IPs directly and our attacks are extremely powerful (peak over 2 Tbps)

This means that your websites and other connected services will be unavailable for everyone. Please also note that this will severely damage your reputation among your customers who use online services.

Worst of all for you, you will lose Internet access in your offices too!

We will refrain from attacking your network for a small fee. The current fee is 20 Bitcoin (BTC). It's a small price for what will happen when your whole network goes down. Is it worth it? You decide!

We are giving you time to buy Bitcoin if you don't have it already. And hopefully for this message to reach somebody who can handle it properly.

If you don't pay the attack will start and fee to stop will increase to 30 BTC and will increase by 10 Bitcoin for each day after the deadline that passed without payment.

Please send Bitcoin to the following Bitcoin address: [redacted]

Once you have paid we will automatically get informed that it was your payment.
Please note that you have to make payment before the deadline or the attack WILL start!

If you decide not to pay, we will start the attack on the indicated date and uphold it until you do. We will completely destroy your reputation and make sure your services will remain offline until you pay.

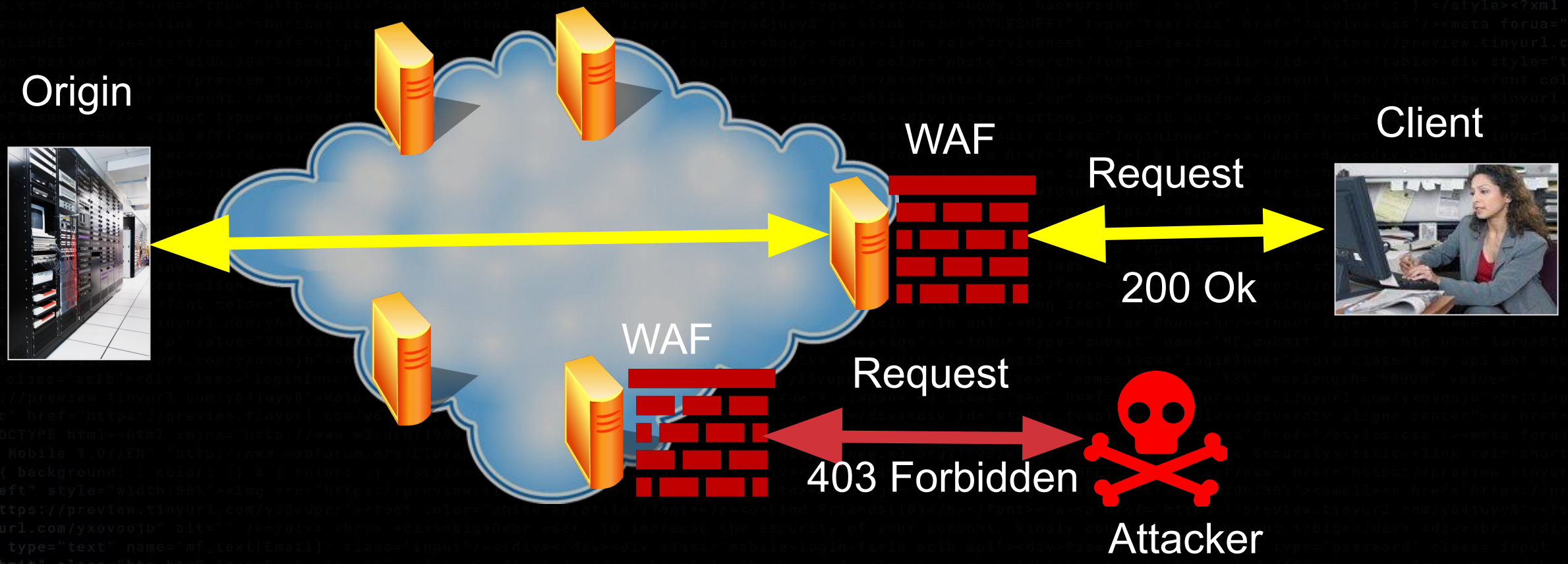
Do not reply to this email, don't try to reason or negotiate, we will not read any replies.

Once you have paid we won't start the attack and you will never hear from us again.

Please note we will respect your privacy and reputation, so no one will find out that you have complied.

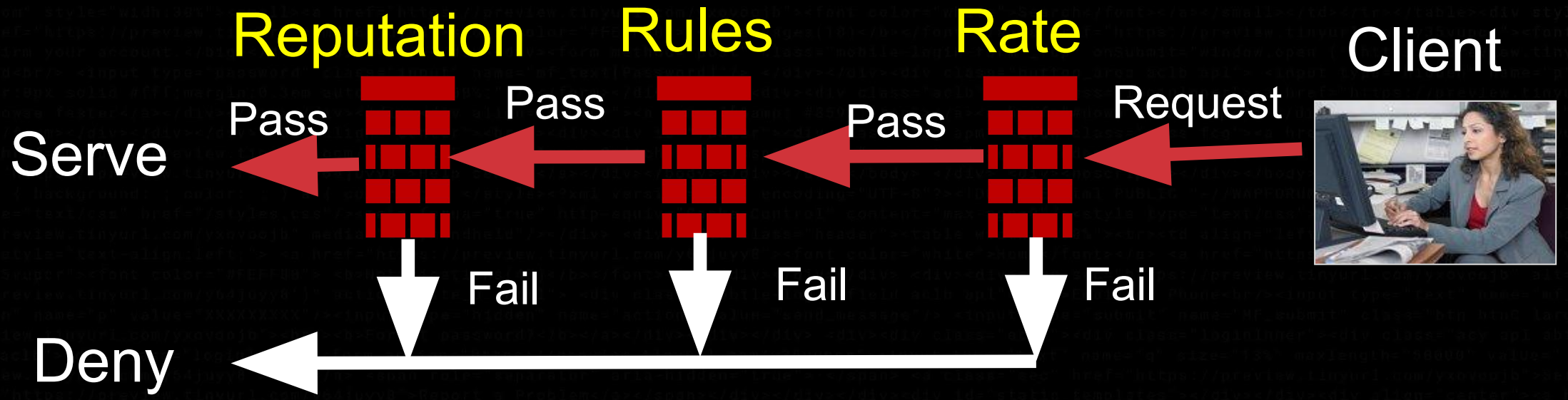
 

# Distributed Web Application Firewall (WAF) at the Edge




- Avoid False Positives
- Security Performance Tradeoffs

# Web Application Firewall Controls at the Edge



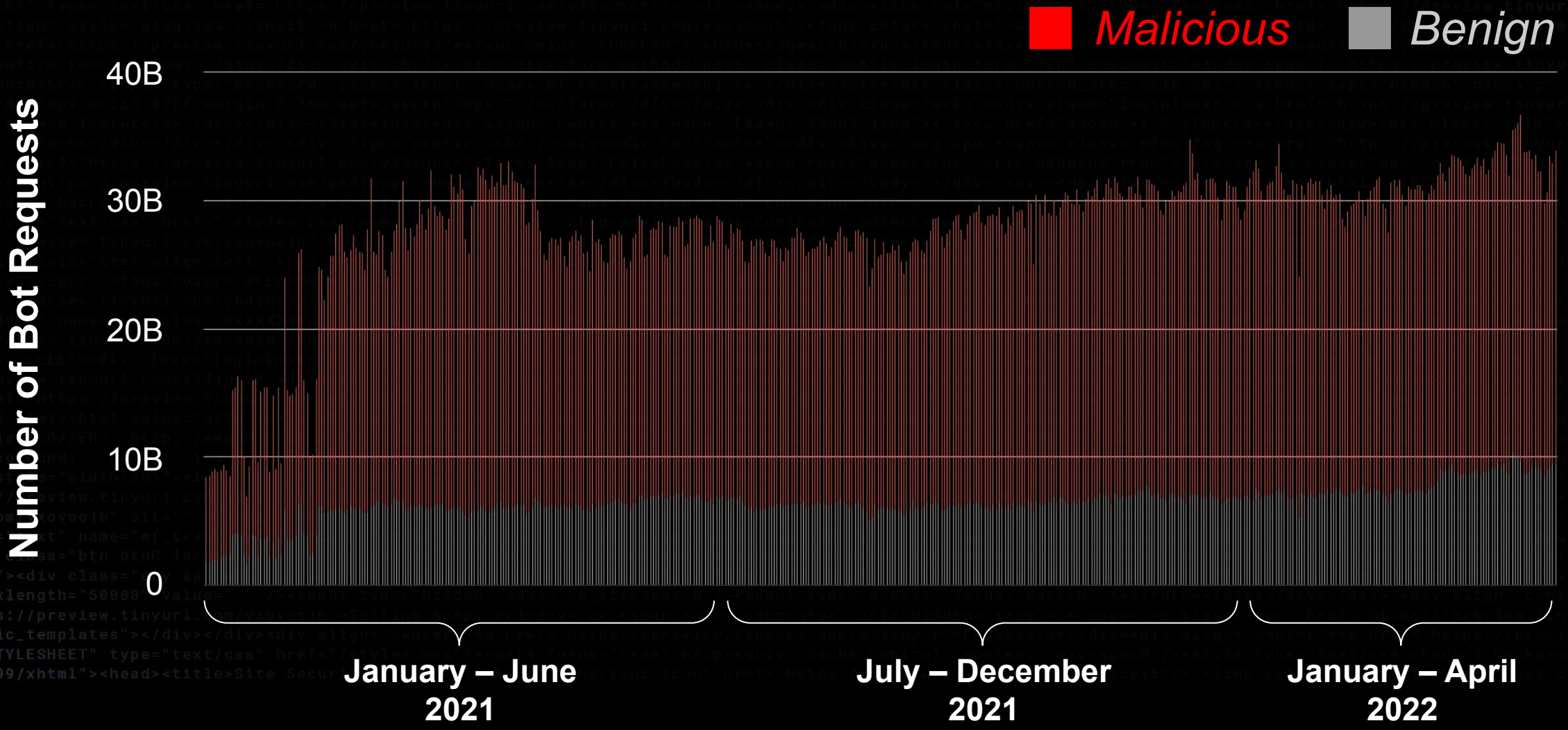

Rate: Average and Burst Thresholds

Rules: Cross-site Scripting, SQL Injection, PHP Injection,etc

Reputation: Score each client

# Intelligent Bot Management at the Edge

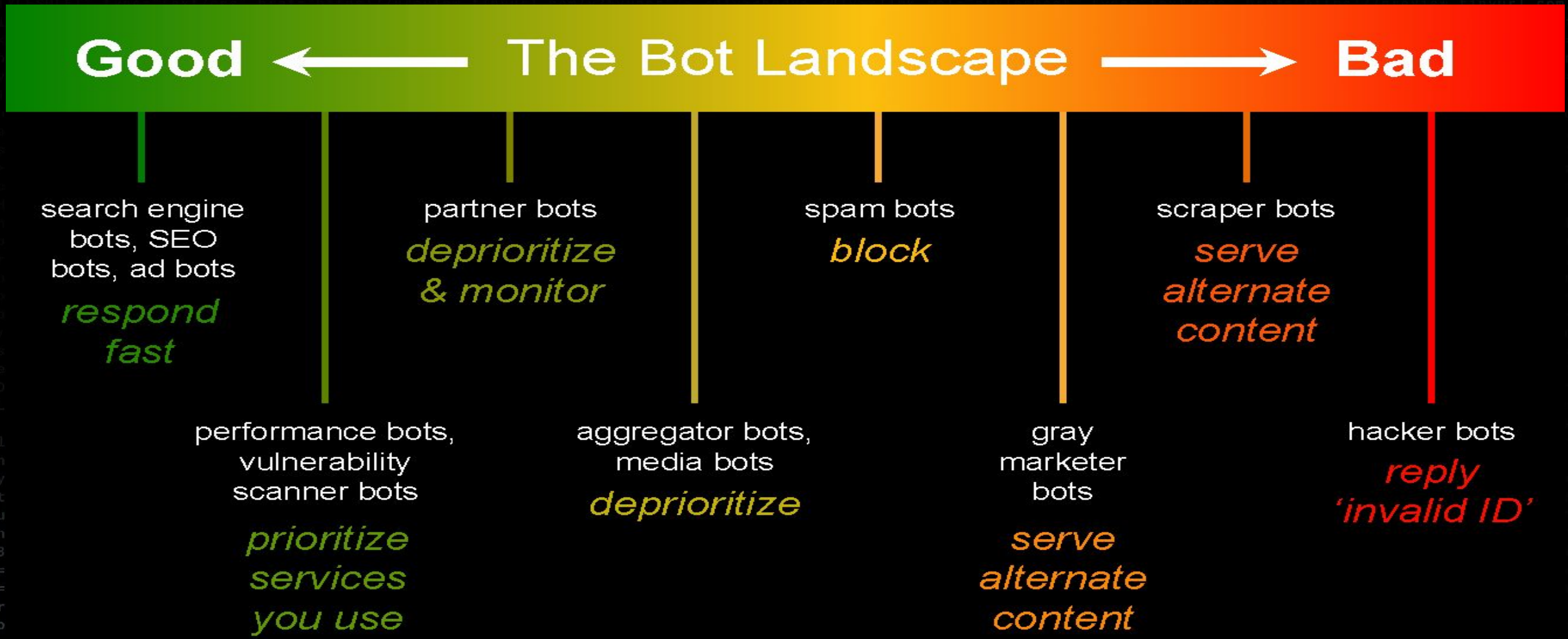
## Global Bot Activity



# Intelligent Edge: Detect, Classify & Manage Bots

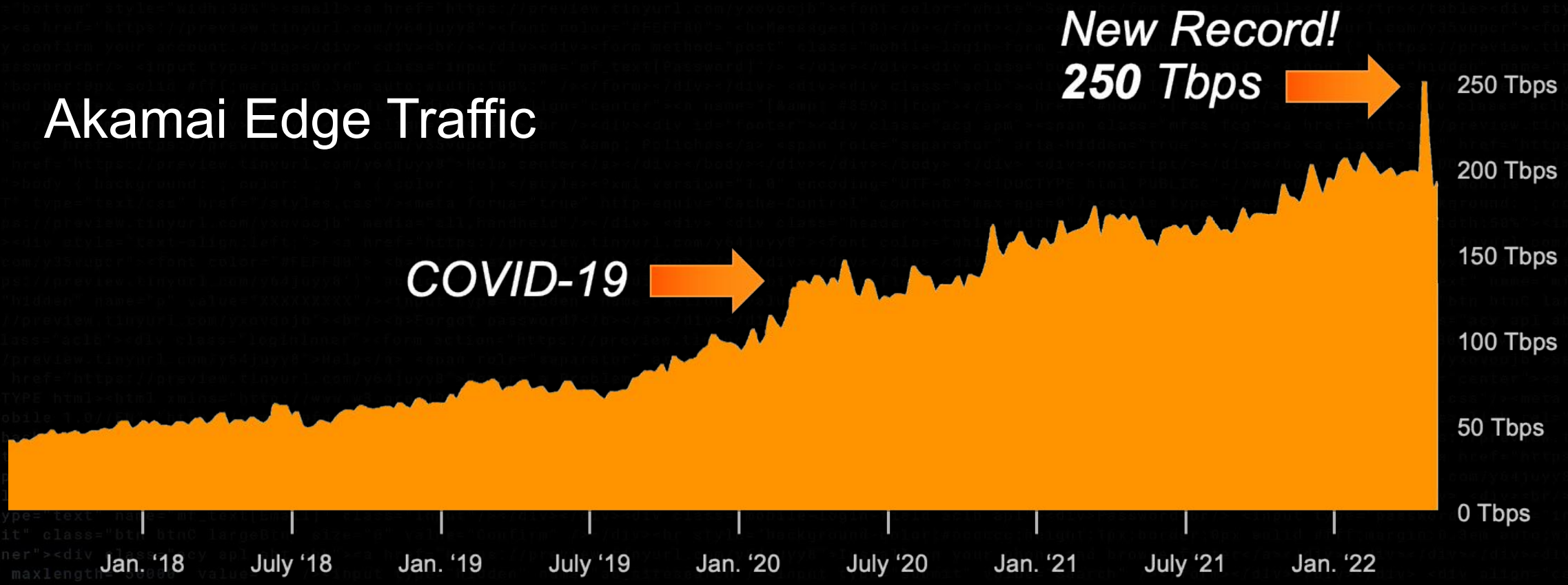
## ML Features: Network, Reputation, Device Fingerprint, Session Behavior



# Story of the Edge in Four Parts

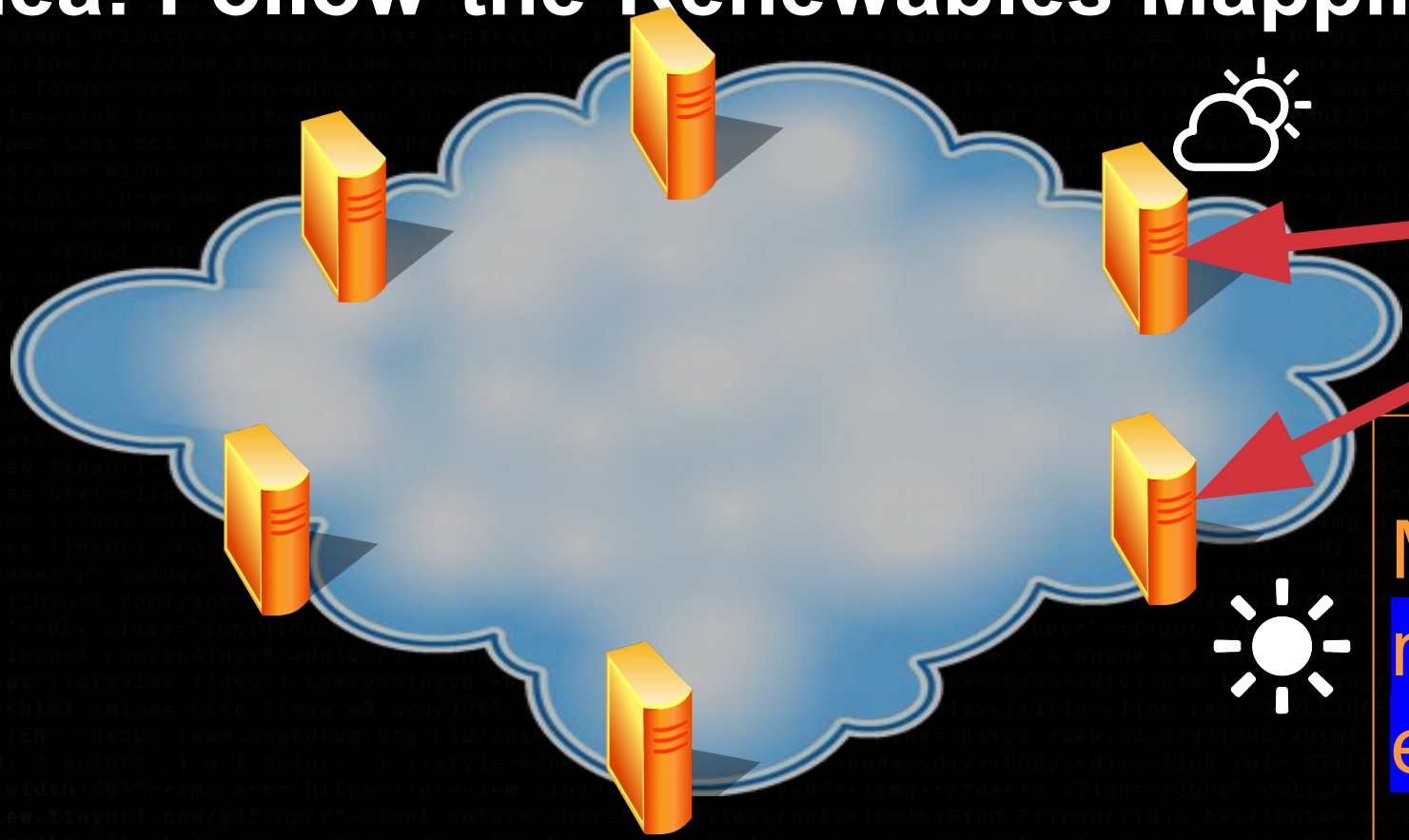
| Chapter 1 | Chapter 2 | Chapter 3 | Chapter 4 |
| --- | --- | --- | --- |
| Content Delivery | Edge Computing | Defending the Edge (and the Internet) | A Sustainable Zero-Carbon Edge |

# Increase in internet demand and edge traffic leads more energy usage and carbon footprint




**Goal: Edge powered (nearly) entirely by renewables.**
**Carbon-First design principle.**

# Idea: Follow-the-Renewables Mapping



Map client to nearest ”sunny” edge server

Re-mapping client load within a small geographical radius could yield 40% grid energy reduction.

# Idea: Deploy tens of thousands of highly-distributed micro-data centers powered entirely by renewables

MassZero: Micro-data center powered by solar and lithium batteries located in Holyoke, Massachusetts

- Traditional data centers are energy dense. 100 MW = 450 acre of solar.
- Move edge servers to available green energy rather than vice versa.

# Q&A

For more information see papers at:
https://groups.cs.umass.edu/ramesh/real-world-systems/