BITS Pilani presentation

BITS Pilani
Pilani Campus

Paramananda Barik
CS&IS Department

BITS Pilani
Pilani Campus

SEZG586/SSZG586, Edge Computing Systems

Lecture No.15_16

Edge Computing Systems- Classification

- Push from cloud

  Cloud providers push services and computation to the edge in order to leverage locality, reduce response time and improve user experience

- Pull from IoT

  Internet of Things (IoT) applications pull services and computation from the faraway cloud to the near edge to handle the huge amount of data generated by IoT devices

- Hybrid cloud-edge analytics

  Integration of advantages of cloud and edge provides a solution to facilitateboth global optimal results and minimum response time in modern advanced services and applications

Edge Computing Systems- Classification

- Push from cloud

  Cloudlet, Cachier, AirBox, CloudPath

- Pull from IoT

  PCloud, ParaDrop, FocusStack and SpanEdge

- Hybrid cloud-edge analytics

  Firework, Cloud-Sea Computing Systems, LinkEdge, AWS GreenGrass

Cloudlet -PFC

Started in 2009, Carnegie Mellon University

Cloudlet is a trusted,

- resource-rich computer
- cluster of computers
- connected to the Internet
- available to nearby mobile devices

Upgrades the two-tier architecture “Mobile Device-Cloud”

to a three tier architecture “Mobile Device-Cloudlet-Cloud”

Cloudlet -PFC

Cloudlet has three main features:

- Soft State
- Rich Resources
- Close to Users

Cloudlet supports application mobility:

- Cloudlet Discovery
- VM Provisioning
- VM Handoff

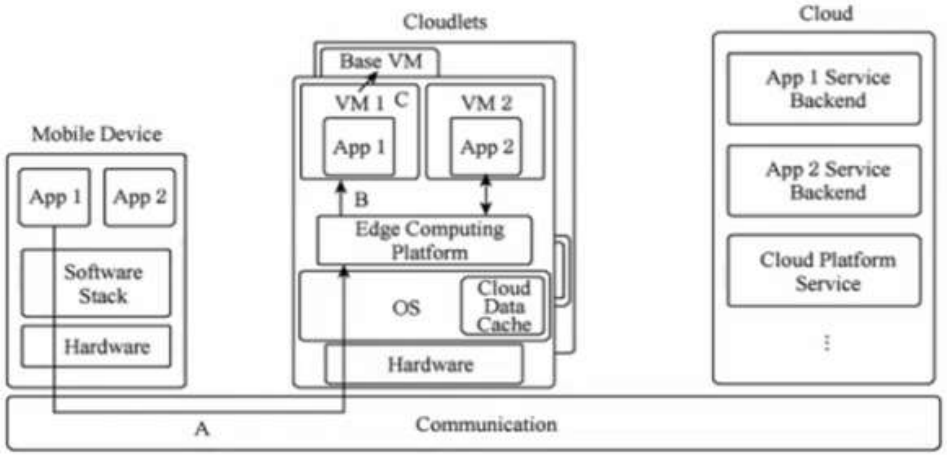


Fig. 3. Cloudlet Component Overview and Functions that support application mobility. A: Cloudlet Discovery, B: VM Provisioning, C: VM Handoff.

CloudPath - PFC

CloudPath, an edge computing system

University of Toronto

computing and storage are provided along the path from the user device to the cloud data center

Supports

on-demand allocation

dynamic deployment of the multi-level architecture

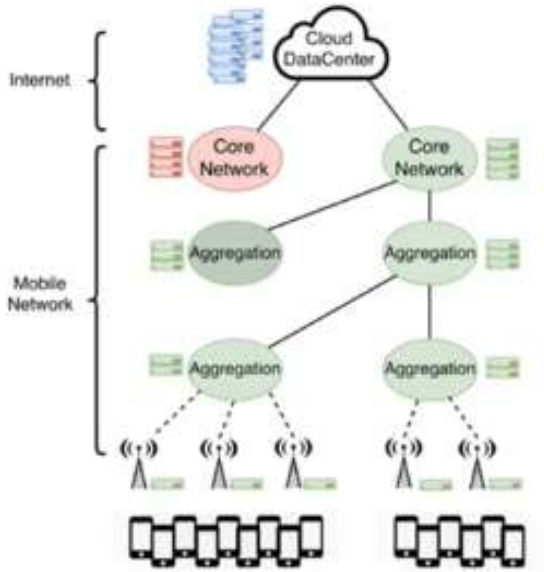


Fig. 4. CloudPath Architecture [28].

CloudPath

The system reassigns following tasks of the data centers along the path (for path computing) to support different types of applications:

- IoT data aggregation
- Data caching services
- Data processing services

Developers can select an optimal hierarchical deployment plan

- cost
- delay
- resource availability
- geographic coverage

CloudPath

Each CloudPath node usually consists of the following six modules:

1) PathExecute
2) PathStore
3) PathRoute
4) PathDeploy
5) PathMonitor
6) PathInit

Pcloud - FPI

The resources are virtualized through a special virtualization layer named STRATUS

Form a distributed resource pool that can discover new resources and monitor resource changes

The run-time mechanism

-Selects and combines appropriate resources

-Generates a new instance

-Provide corresponding services for external applications, according to the resource access control policy

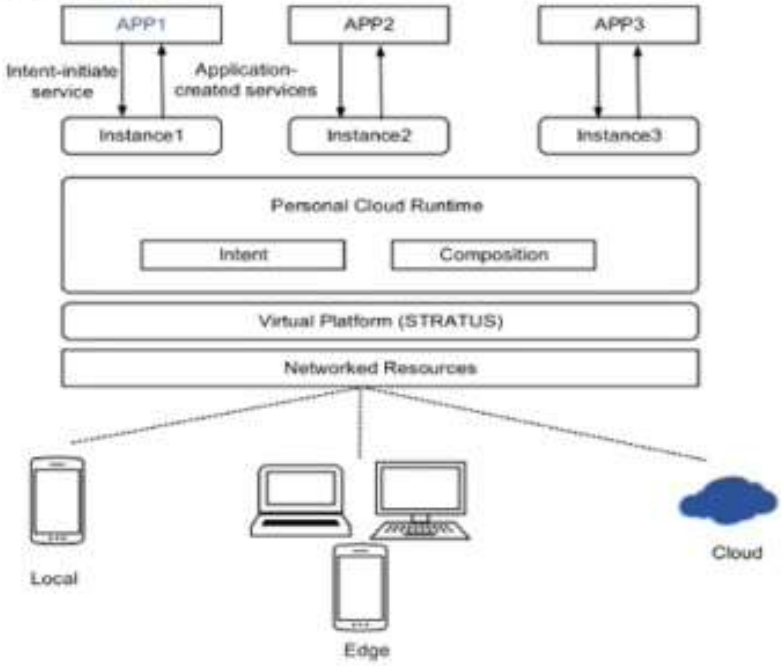


Fig. 5. PCloud Architecture [10].

PCloud

Actual operation:

1. Mobile application describes the required resources to the PCloud through interfaces
2. The PCloud will find out the optimal resource configuration by analyzing the description and the currently available resources
3. Generates an instance to provide corresponding services for the application.

ParaDrop - PFI

ParaDrop: University of Wisconsin-Madison

ParaDrop upgrades the existing access point to an edge computing system

ParaDrop leverages the lightweight container virtualization technique

ParaDrop server(in the cloud) controls:

- -deployment
- -starting
- -deletion of Apps

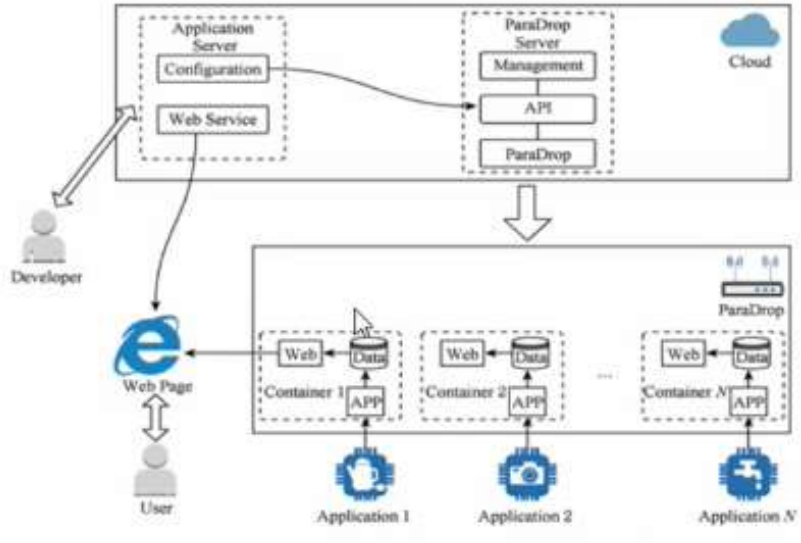


Fig. 6. ParaDrop System [12].

ParaDrop

ParaDrop is mainly used for IoT applications - IoT data analysis

Advantages:

a) Sensitive data can be processed locally, it protects the users privacy

b) WiFi access point is only one hop away from the data source, leading to low network delay and stable connection

c) Only user requested data are transmitted to the equipment through the Internet

d) Gateway can obtain the location information of the edge devices through radio signals (e.g., the distance between devices, and the location of the specific device), which facilitates the location-aware services

e) when edge devices cannot be connected to the Internet, the edge service can still work

SpanEdge - PFI

SpanEdge: Royal Institute of Technology in Sweden

Developed to handle huge amount of data generated by IoT devices

The data center in SpanEdge is composed of two levels:

- -The cloud data center
- -The edge data center

Partial streaming processing tasks run on the edge central nodes to reduce latency and boost performance

SpanEdge uses the master-worker architecture: one manager and multiple workers

The manager collects the streaming processing requests and assigns tasks to the workers

SpanEdge

Workers consist of cluster nodes whose primary responsibility is to execute tasks

Two types of workers:

- hub-worker (first level)
- spoke-worker (second level)

The communication in SpanEdge is also divided into:

- System management communication (worker-manager)
- Data transmission communication (worker-worker)

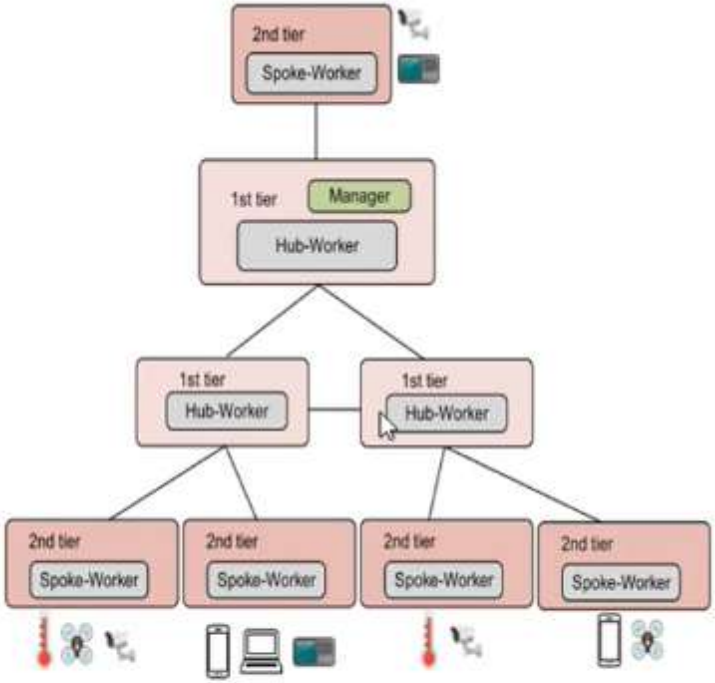


Fig. 7. SpanEdge Architecture [13].

Cloud-Sea Computing Systems - Hybrid

Cloud-Sea Computing Systems: Next Generation Information and Communication Technology initiative

launched by the Chinese Academy of Science in 2012

The NICT initiative aims to address the three major technology challenges:

i) improving the performance per watt by 1000 times

ii) supporting more applications from the human cyber-physical ternary computing

iii) enabling transformative innovations in devices, systems and applications, while without polluting beneficial IT ecosystems.

Cloud-Sea Computing Model

In the cloud-sea computing system, “cloud” refers to the datacenters and “sea” refers to the terminal side

The project contains four research components:

i. a computing model called REST 2.0

ii. a three-tier storage system architecture capable of managing ZBs of data

iii. a billion-thread datacenter server with high energy efficiency

iv. an elastic processor aiming at energy efficiency of one trillion operations per second per watt

Cloud-Sea Computing Model

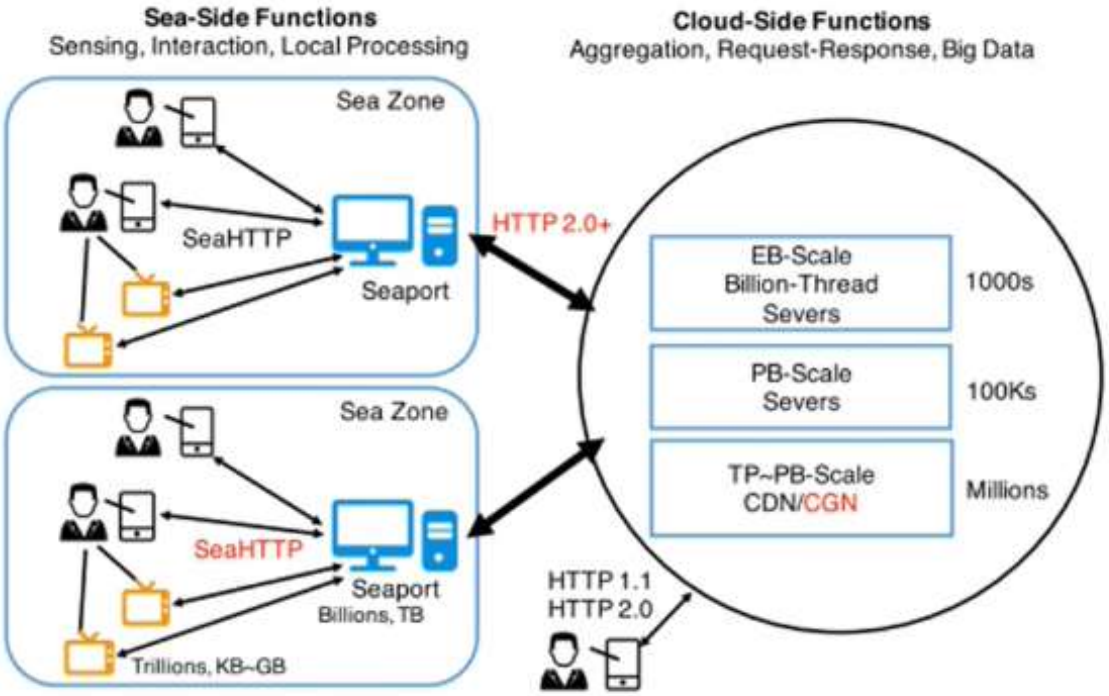
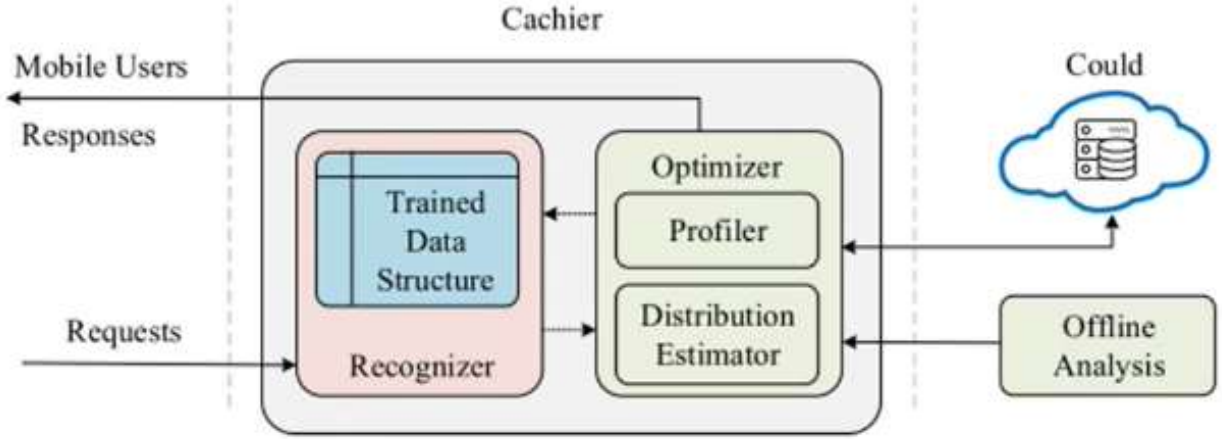


Fig. 8. Cloud-Sea Computing Model.

Cachier and Precog

Cachier and Precog: edge caching systems that proposed by the researchers from Carnegie Mellon University for **image recognition**

Recognition applications have **strict response time** requirements, while the computation is huge due to the **mode complexity** and **dataset size**



Cachier

Cachier consists of:

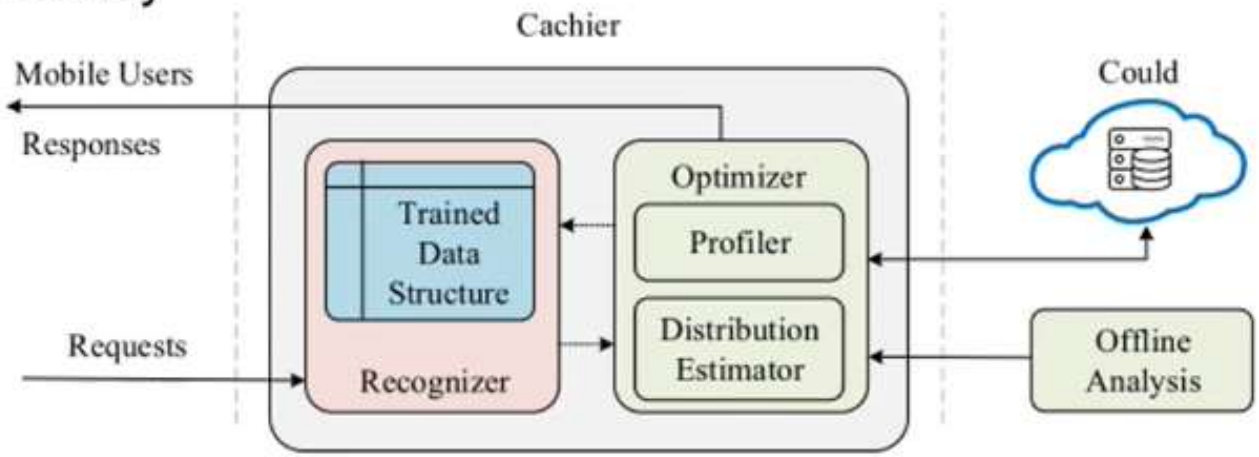
- Recognizer Module
- Optimizer Module
- Offline Analysis Module

The Recognizer Module is responsible for analyzing and matching the received figures according to the cached training data and model. If there is a match, the result will be directly returned to the user, otherwise the figure will be transmitted to the cloud for recognition.

Cachier

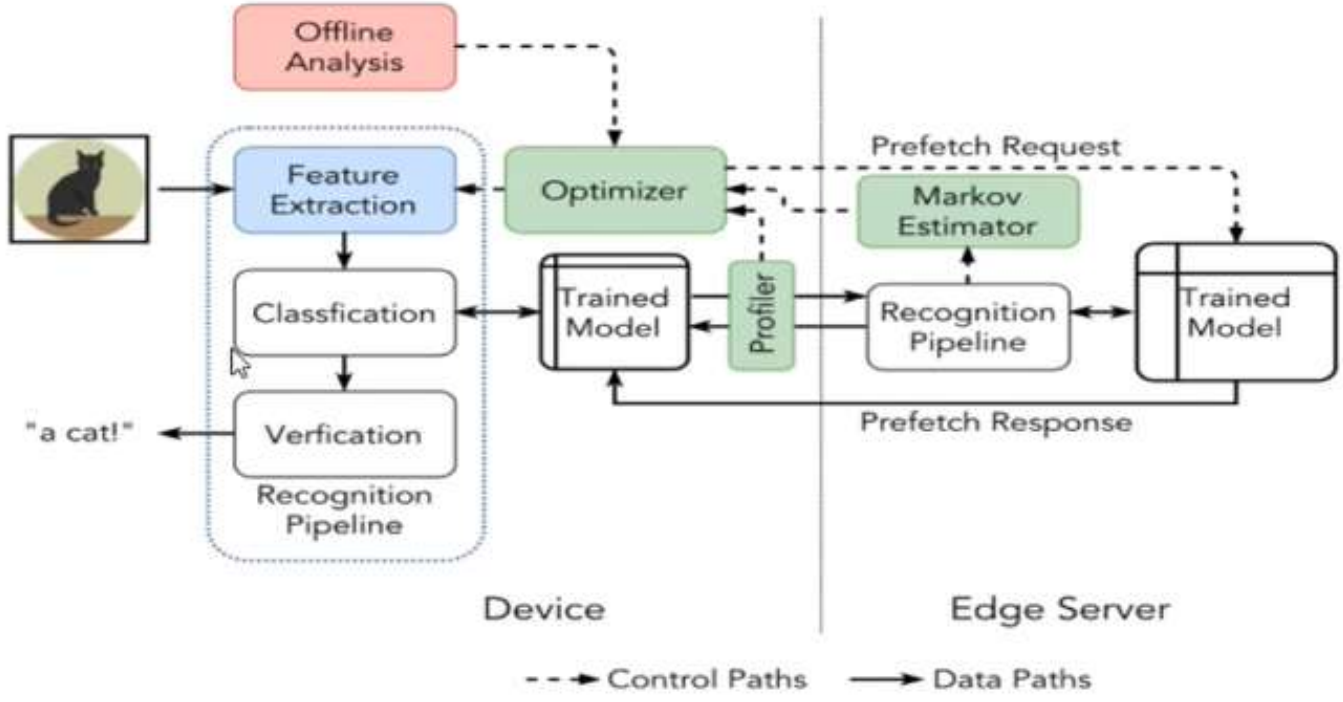
Optimizer Module:

- Distribution estimator: predict the request distribution
- Profiler submodule: estimates the network delay and cloud latency incurred by cache misses
- Offline Analysis: calculate cache searching delay and the cache accuracy



Precog

Precog is an extension of Cachier



FocusStack

FocusStack: developed by AT&T Labs

FocusStack consists of two parts :

i) Geocast system, which provides location-based situational awareness (LSA) information

ii) OpenStack extension (OSE), responsible for deploying, executing, and managing the containers on the edge devices.

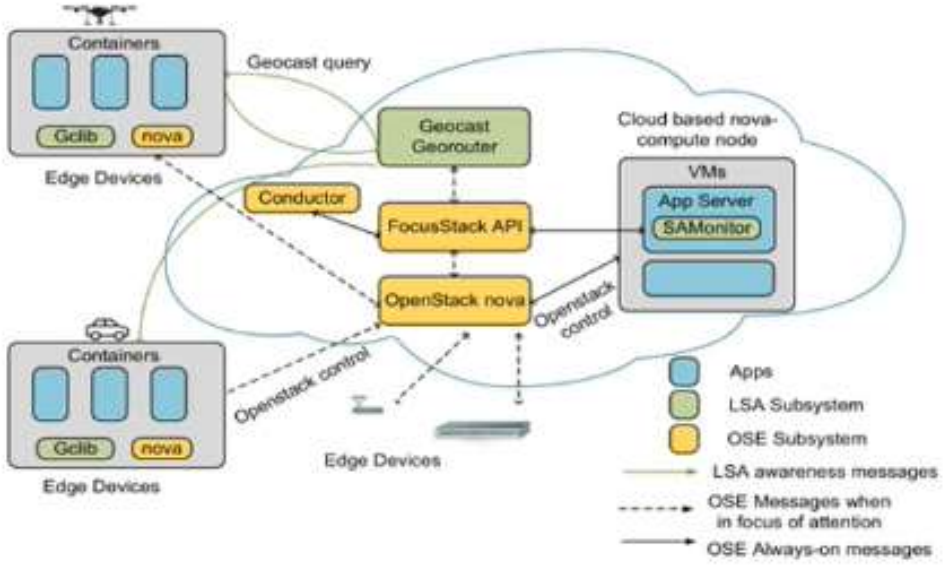


Fig. 11. FocusStack System [18].

Other Edge Computing Systems

Moving vehicles: an open full-stack edge computing-based platform OpenVDAP is proposed for the data analytics of connected and autonomous vehicles (CAVs)

SafeShareRide: an edge-based attack detection system addressing in-vehicle security for ridesharing services

Vigilia system: is proposed to strengthen the smart home systems by restricting the network access of devices

LAVEA: is an edge-based system built for latency-aware video analytics nearby the end users

VideoEdge: video analytics implementation across a 3-tier hierarchy of clusters in the city environment


thank you