Designing Hierarchical Edge Cloud for Mobile Computing

Wei Gao
The University of Tennessee, Knoxville

Supported by NSF Award No. CNS-1527612
Edge Cloud

DATA CENTERS

EDGE CLOUD

END USERS

Google

amazon

Microsoft Azure
Edge Cloud for Mobile Computing

- Reduced response latency
  - Delay-sensitive mobile applications
- Higher efficiency of resource utilization
  - Distributed processing
Applications of Edge Cloud

- Cloud Gaming
- Virtual Reality
- Internet of Things
- Smart Cities and Communities

NSF CSR Highlighted Area
Challenges

- **Adaptability**
  - Optimized performance?
  - Minimized cost?

- **User mobility**
  - Minimized cost?

Provisioning for the peak load

Complete move of data and program

User mobility
Our Solution: Hierarchical Edge Cloud

- Adaptability
  - Aggregation of peak load

- User mobility
  - Partial migration of data and program

Geo-distributed tree hierarchy
Task 1: Optimal Workload Placement

- **Our focus:** minimized response latency
  - Where to place a workload
  - How much capacity for a workload

- **Challenge**
  - Delay tradeoff

**A Hierarchical Edge Cloud Architecture for Mobile Computing, in IEEE INFOCOM’16.**
Task 1: Optimal Workload Placement

- Distributed optimization

\[
\min f = \sum_{i=1}^{m} \left( \frac{w_i}{\lambda_i, \gamma_i, C_{\gamma_i}} + \left( I(\gamma_i) - 1 \right) \frac{s_i}{B_{\gamma_i}} \right),
\]

s.t. \( \sum_{j \in J} \lambda_{i,j} = 1, j = 1, 2, \ldots, n \)

A Hierarchical Edge Cloud Architecture for Mobile Computing, in *IEEE INFOCOM'16.*
Task 1: Optimal Workload Placement

- Problem transformation

\[
\begin{align*}
\min f(\lambda, \gamma) \\
\text{s.t. } g(\lambda, \gamma) = 1
\end{align*}
\]

\[\gamma = \gamma^*\]

\[
\begin{align*}
\min f(\lambda | \gamma = \gamma^*) \\
\text{s.t. } g(\lambda | \gamma = \gamma^*) = 1
\end{align*}
\]

- Non-linear Mixed Integer Programming
- 1. Convex optimization with variable \(\lambda\)
- 2. Integer programming with variable \(\gamma\)

---

A Hierarchical Edge Cloud Architecture for Mobile Computing, in *IEEE INFOCOM’16.*
Task 2: Supporting User Mobility

- Remote program execution with least context migration

Task 2: Supporting User Mobility

Implementation

- Heterogeneous mobile and wearable platforms
  - Samsung Galaxy S4
  - LG Watch Urbane
  - Samsung Nexus 10 Tablet
  - 1,500 LoC in Java and 1,000 LoC in C++ over Android v5.1.1 OS kernel

- Edge cloud servers
  - x86-based instances of Dalvik VM
  - Executing ARM-based native programs
Ongoing Work and Future Directions

- Optimal resource provisioning
  - Minimizing both CapEx and OpEx

- Virtual reality over mobile platforms
Thank you

- Questions?