
Niraj Kumar, Arijit Mondal

Department of Computer Science & Engineering
Indian Institute of Technology Patna
India

RTSS 2019 WiP Presentation

April 24, 2020
Introduction

- The successive (however, overlapping) phases of computing paradigm
  - Mainframe computers → Personal computers → Network/Internet computing → Grid computing → Cloud computing

- Cloud Computing
  - Widely deployed
  - Huge number of devices
  - It is expected that by 2020
    - nearly 50 billion devices will be connected to the Internet
    - generating an economy of exceeding 3 trillion
    - data volume of more than 43 trillion gigabytes

- Issues
  - Huge amount of data → tremendous network bandwidth
  - Large latency
  - Energy Consumption
Fog Computing

- Addresses the inherent issues of cloud computing
- Pushes applications, services, computing, and decision making near to the devices where data is being generated
- Offers the benefits of the cloud computing systems to the real-time applications
- Complementary not a replacement
Problem Formulation

- Cloud computing: a few centralized servers
- Fog computing systems: a large number of geographically separated fog devices
- A major challenge: offloading of tasks with various constraints
- Another major challenge: set up a pricing mechanism for the usage of resources/services
  - as usually the fog devices are owned by different parties
- Cloud computing
  - well-accepted and established computing model
  - pricing problem has been widely studied
Only a few elementary works deal with the pricing for fog computing systems.

Common practice: either of the two problems
- pricing and offloading

The pricing strategy of service providers $\rightarrow$ objective is to maximize the profit.

Users $\rightarrow$ timely execution of the workload (with certain constraints), however with minimum cost.

*social welfare* $\rightarrow$ an inclusive parameter.

Assumption: service providers fix the prices independently, the two problems becomes strongly related.
- which is realistic for highly distributed systems such as fog computing systems.
Architecture

Smart Gateway

CS_1 \quad \cdots \quad CS_l

FI_1 \quad FI_2 \quad \cdots \quad FI_m

TN_1 \quad TN_2 \quad TN_3 \quad TN_4 \quad \cdots \quad TN_n

private fog

private fog
3-layer fog architecture

Each terminal node is a device generating a workload

A fog instance consists of one or multiple fog devices that acts as a unit

Each workload is to be scheduled on a fog instance that promises timely execution with the least cost

The fog devices are geographically separated
  - not every fog instance is reachable from each terminal node

Smart gateway selects the most appropriate fog/cloud node to execute the workload
Model

All the workloads generated in an interval are offloaded at the beginning of next interval

\[ \gamma_j(\tilde{t}) \] and \( \delta_j(\tilde{t}) \) → computational capacity and cost of execution per unit time during \( \tilde{t} \) at \( F|j(\tilde{t}) \in \mathbb{F}(\tilde{t}) \)

\( TN_i(\tilde{t}) \in \mathbb{T}(\tilde{t}) \) generates a workload \( w_i(\tilde{t})\langle in(i, \tilde{t}), out(i, \tilde{t}), C_i(\tilde{t}), D_i(\tilde{t}) \rangle \)
- input and output data \( in(i, \tilde{t}) \) bytes and \( out(i, \tilde{t}) \) bytes, respectively
- \( C_i(\tilde{t}) \) → required number of computation cycles
- \( D_i(\tilde{t}) \) → deadline

\( \alpha_{ij}(\tilde{t}) \) → transmission rate
Formulation

- $\beta_{ij}(\tilde{t}) \rightarrow$ connected
- $\Theta_i(\tilde{t}) \rightarrow$ reachable
- $\sigma_{ij}(\tilde{t}) \rightarrow$ allocated
- *accepted workload*
  \[
  \sum_{\forall j | \text{Fl}_j(\tilde{t}) \in \Theta_i(\tilde{t})} \sigma_{ij}(\tilde{t}) = 1
  \]

- *communication time*
  \[
  \lambda_i(\tilde{t}) = \sum_{\forall j | \text{Fl}_j(\tilde{t}) \in \Theta_i(\tilde{t})} \sigma_{ij}(\tilde{t}) \left( \frac{\text{in}(i, \tilde{t}) + \text{out}(i, \tilde{t})}{\alpha_{ij}(\tilde{t})} \right)
  \]

- *computation time*
  \[
  \mu_i(\tilde{t}) = \sum_{\forall j | \text{Fl}_j(\tilde{t}) \in \Theta_i(\tilde{t})} \sigma_{ij}(\tilde{t}) \cdot \frac{C_i(\tilde{t})}{\gamma_j(\tilde{t})}
  \]
Formulation

- Output available time OAT\((i, \tilde{t})\)
  \[
  \lambda_i(\tilde{t}) + \mu_i(\tilde{t}) + \text{wait}(i, \tilde{t}) \leq D_i(\tilde{t})
  \]

- Cost of execution
  \[
  \text{cost}(i, \tilde{t}) = \sum_{\forall \text{FI}_j(\tilde{t}) \in \Theta_i(\tilde{t})} \sigma_{ij}(\tilde{t}) \cdot \frac{C_i(\tilde{t})}{\gamma_j(\tilde{t})} \cdot \delta_j(\tilde{t})
  \]

- User surplus at TN\(_i(\tilde{t})\) during \(\tilde{t}\) is
  \[
  \gamma_i(\tilde{t}) = \text{util}(i, \tilde{t}) - \text{cost}(i, \tilde{t})
  \]

- \(P_j(\tilde{t})\) → profit at Fl\(_j(\tilde{t})\) in interval \(\tilde{t}\)

- Social welfare during \(\tilde{t}\) is
  \[
  SW(\tilde{t}) = \sum_{\forall j|\text{Fl}_j(\tilde{t}) \in F(\tilde{t})} P_j(\tilde{t}) + \sum_{\forall i|\text{TN}_i(\tilde{t}) \in T(\tilde{t})} \gamma_i(\tilde{t})
  \]

- Compute \(SW(p)\)
Algorithm 1: $h\text{Cost}$

1. for each $w_i(t) \in \mathbb{W}(t)$ do
2.     $proximity(i, t) \leftarrow \emptyset$
3.     for each $FI_j(t) \in \Theta_i(t)$ do
4.         if $\frac{in(i,t)}{\alpha_{ij}(t)} + \frac{C_i(t)}{y_j(t)} + \frac{out(i,t)}{\alpha_{ij}(t)} \leq D_i(t)$ then
5.             $proximity(i, t) = proximity(i, t) \cup FI_j(t)$
6.     end
7. end
8. for each $w_i(t) \in \mathbb{W}(t)$ do
9.     $cand(i, t) \leftarrow \emptyset$
10.    for each $FI_j(t) \in proximity(i, t)$ do
11.        if Eq. (5) holds for $w_i(t)$ and existing workloads on $FI_j(t)$ then add $FI_j(t)$ to $cand(i, t)$
12.    end
13. Allocate $w_i(t)$ to the fog instance $FI_j(t) \in cand(i, t)$ with least execution cost
14. end
## Proposed Approach

| cfg | #f | %Impr | %Reject  
|-----|----|-------|---------
|     |    |       | hUtil   | hCost   |
| 1   | 100| 0.81  | 4.23    | 4.61    |
| 2   | 150| 14.86 | 1.99    | 0.21    |
| 3   | 200| 24.36 | 1.14    | 0.11    |
| 4   | 250| 31.72 | 0.86    | 0.09    |
| 5   | 300| 35.97 | 0.75    | 0.07    |
Conclusion and Future Work

- Addresses **pricing** and the **offloading** problem in an integrated manner for the real-time tasks.
- Objective is to **maximize the social welfare**, whereas **minimize the cost**.
- **Future Works**
  - Include the cloud layer
  - Obtaining an optimal solution
  - Pricing mechanism
    - devise the price at the beginning of each interval
    - must examine the interplay of revenue and profit with other parameters
    - further exploration on computing $util(i, \bar{t})$
Thanks