Network Optimization via Path Selection in a Heterogeneous Cloud-Based Internet of Things

Eric Sturzinger
Group Meeting
19AUG 2016

Outline

• Purpose/Motivation
• Problem Statement
• Functional Based Application Requirements
• Performance & Cost Application Requirements
• Topology/Flow Scenarios
• ILP Formulation
• Dynamic Simulation
• Extending over WSNs
• Conclusion
### Purpose/Motivation

- **2019**: Cisco predicts > 500 zettabytes of Internet traffic/year
  - 1 zettabyte = $10^{21}$ bytes = 1 billion TB = 1 trillion GB
- **2020**: > 50 billion connected devices, cloud consumption will increase 4x by 2020 to 8.6 zettabytes, enterprises plan to transfer 56% of applications and data to cloud
- **M2M Devices**: 30% of all connected devices today to 46% in 2020, but only ~2-4% of traffic
- Will IoT simply require more physical layer core capacity? Or will traffic nature and increased heterogeneity require more robust traffic engineering and/or policy based/constraint-based routing?


### Problem Statement

1st Priority: Optimize performance metrics of a set of traffic flows with heterogeneous application & traffic profiles via path selection within a core layer internetwork.

Solution form: Unique path for each node pair & application profile, intended to generate diverse set of SLAs for each node pair

2nd Priority: Given above path solutions by node pair, minimize operational cost to MAN service provider while maintaining end to end performance requirements. Simulate various application profile mixtures and loads; determine at what loads performance is substantially degraded. Repeat for various link capacities (higher costs).

3rd Priority: Inject a multi-homed redundant 4G/5G connection from specific sensors/sink node to specific nodes in access/metro layer
## Functional Based Application Requirements

- Each application requires various tiers of data processing capability prior to delivery to final destination
- Applications may require storage at DC
- Interactive applications: location of the client (and thus the path metrics) will affect the computation location
- Traffic may be aggregated from multiple sources for a single event

<table>
<thead>
<tr>
<th>Storage Required</th>
<th>Processing Required</th>
<th>Cause</th>
<th>Source</th>
<th>Destination</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>None</td>
<td>Primary Event</td>
<td>Single Sensor</td>
<td>Single Sensor</td>
</tr>
<tr>
<td>Data Center</td>
<td>Tier 1- Sink/GW</td>
<td>Secondary Event</td>
<td>Multiple Sensors</td>
<td>Multiple Sensors</td>
</tr>
<tr>
<td>Tier 2</td>
<td>Client Demand</td>
<td>DC</td>
<td>DC</td>
<td></td>
</tr>
<tr>
<td>Tier 3 - DC</td>
<td>Periodic to DC</td>
<td>Client</td>
<td>Client</td>
<td></td>
</tr>
</tbody>
</table>

### Application Performance/Cost Requirements

<table>
<thead>
<tr>
<th>Latency</th>
<th>Bandwidth</th>
<th>Reliability</th>
<th>Jitter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uni-directional: &lt; 50 ms</td>
<td>Min 50 Mbps per link</td>
<td>Prob. Delivery: &gt; 99.9%</td>
<td>Latency +/- 10%</td>
</tr>
<tr>
<td>Bi-directional: &lt; 100ms</td>
<td>Path of Least Congestion</td>
<td>Latency +/- 20%</td>
<td></td>
</tr>
</tbody>
</table>

### Upstream to Internet Downstream from Internet Storage Processing (Tier I, II, III)

- X dollars per Mbps
- X dollars per Mbps
- X $ per Mbit
- I: X dollars per Mbps (high)
- II: X dollars per Mbps (moderate)
- III: X dollars per Mbps (low)
Flow Scenarios (cont.)

Unidirectional – w/ Storage and/or Tier 3 Processing
Sensor node → Client

Tier 3 Tier 2 Tier 1 N/A Processing
1
1-1
1-2
1-3
Sink/GW Tier 1 Proc.
WSN
Tier 2 Proc.

12-1
12
12-2
DC/Cloud Tier 3 Proc.

Flow Scenarios (cont.)

Interactive (Through DC or direct)
Client → Sensor node → Client

Tier 3 Tier 2 Tier 1 N/A Processing
1
1-1
1-2
1-3
Sink/GW Tier 1 Proc.
WSN
Tier 2 Proc.

12-1
12
12-2
DC/Cloud Tier 3 Proc.

Client
Mathematical Formulation

Variables:

\( v_{c,a} \): Offered traffic of application profile \( a \), between node pair \( c \).

\( r_{c,k,a} \): Traffic of application profile \( a \), routed over the \( k^{th} \) admissible path between node pair \( c \).

\( S_{a,i} = 1 \), if traffic of application profile \( a \) requires storage at DC, located behind node \( i \)

\( d_{t,i,j} \): Transmission delay at node \( i \) on link \( i,j \)

\( d_{p,i,j} \): Propagation delay on link \( i,j \)

Mathematical Formulation (cont.)

Variables:

\( d_{p,i} \): Processing delay at node \( i \)

\( K_{c,a} \): Set of all admissible paths between node pair \( c \) for appl \( a \)

\( k_{c,a}^{k} \): \( k^{th} \) admissible path between node pair \( c \) for appl \( a \)

\( h_{c,k,a} \): Hop count of \( k^{th} \) admissible path for node pair \( c \), appl \( a \)

\( P_{a,t}^{c} = 1 \), if traffic of application profile \( a \) between node pair \( c \) is processed at tier \( t \).
Mathematical Formulation (cont.)

Variables:

\(\kappa_{a,t,c}\): Tier \(t\) processing cost per unit of traffic of application profile \(a\) between node pair \(c\)

\(\beta_a\): Parameter that defines relationship between computational power required and traffic of application profile \(a\)

\(\nu_a\): Cost per unit of traffic of application profile \(a\) for data center storage

---

Objectives Function:

\[
\max \prod_{c=1}^{N(N-1)} \prod_{a=1}^{A} \frac{Perf_a^c}{Cost_a^c}
\]

\(Perf_a^c\) and \(Cost_a^c\) are Performance and Cost functions of \(A\) application profiles between each node pair \(c\).

Subject to:

\[\sum_k r_{c,k,a} = v_{c,a} \forall (c,a)\] \[\text{Solenoidality}\]

\[\sum_a \sum_{r_{c,k,a} \in R_{i,j}} r_{c,k,a} \leq F_{i,j} \forall (i,j)\] \[\text{Capacity}\]
Mathematical Formulation (cont.)

\[
\sum_{t=0}^{3} P^e_{a,t} = 1 \quad \forall (a, c) \\
\]

Processing

\[
i \in k_{c,a} \text{ if } S_{a,i} = 1 \quad \forall (a, c) \\
\]

Storage

Performance Functions:

Total Latency (uni-directional):

\[
Perf_1^e = \sum_{(i,j) \in k_c} d_{p,i,j} + \sum_{(i,j) \in k_c} d_{t,i,j} + \sum_{(i) \in k_c} d_{n,i}
\]

Total Latency (bi-directional/interactive):

\[
Perf_2^e = 2(\sum_{(i,j) \in k_c} d_{p,i,j} + \sum_{(i,j) \in k_c} d_{t,i,j} + \sum_{(i) \in k_c} d_{n,i})
\]

Min Throughput Link x Total Latency

\[
Perf_3^e = (\min_{r_{c,k} \in R_{i,j}} r_{c,k}) \left( \sum_{(i,j) \in k_c} d_{p,i,j} + \sum_{(i,j) \in k_c} d_{t,i,j} + \sum_{(i) \in k_c} d_{n,i} \right)
\]
Cost Function:

\[ \text{Cost}^c_{a} = \alpha_{u,a}^c v_a^c + \alpha_{d,a}^c v_a^c + v_{c,a} \sum_{t=0}^{3} \kappa_{a,t}^c p_{a,t}^c + \nu_a s_{a,i} v_{c,a} \]

- \( \alpha_{u,a}^c \) = Cost per unit of upstream traffic of application profile \( a \) from source of node pair \( c \)
- \( \alpha_{d,a}^c \) = Cost per unit of downstream traffic of application profile \( a \) to destination of node pair \( c \)

Mathematical Explanation

Inputs:
- Performance functions based on the application profiles
- Cost functions
- Single/multiple cloud storage/DC locations
- Multiple data processing locations
- Topology, profile proportions at each source node (1/4, ¼, ½)
- Link Capacities

Objective function: Maximize product of performance/cost ratios via ideal paths of heterogeneous application traffic flows across all possible node pairs.

Outputs:
- For each node pair and application profile:
  - Designated path through core nodes
  - Intermediate processing location, if necessary
Logical (tree) topology in a WSN computed based on RPL
Optimized path selection per application profile can be based on the core network ILP solution
Future Work

ILP Solution: Ideal paths that maximize performance/cost ratio for all node pairs and application profiles

Formulate similar ILP for internal WSN path selection; determine feasibility of integration

Compute integrated (complete end-to-end) best path solution

Dynamic simulation: Given ILP solutions, simulate dynamic traffic to determine time varying delay (jitter) and other performance metrics in order to formulate strategy for further optimization.

References


