QoS-aware and status-aware adaptive resource allocation framework in SDN-based IoT middleware

Atefeh Meshinchi

Supervisors:

Alejandro Quintero

Yann-Gaël Guéhéneuc

Université de Montréal-Ecole Polytechnique de Montréal

May 2018

Outline I

1. Problem definition
   - Internet of Things(IoT) definition and vision
   - QoS management in IoT
   - Objective
2. Related works and background
   - Related works
   - Software-Defined Networking
3. Proposed QoS framework
   - Architecture
   - Modules and interaction between components
   - Workflow sequence
4. Proposed QoS Model
   - Making-decision framework

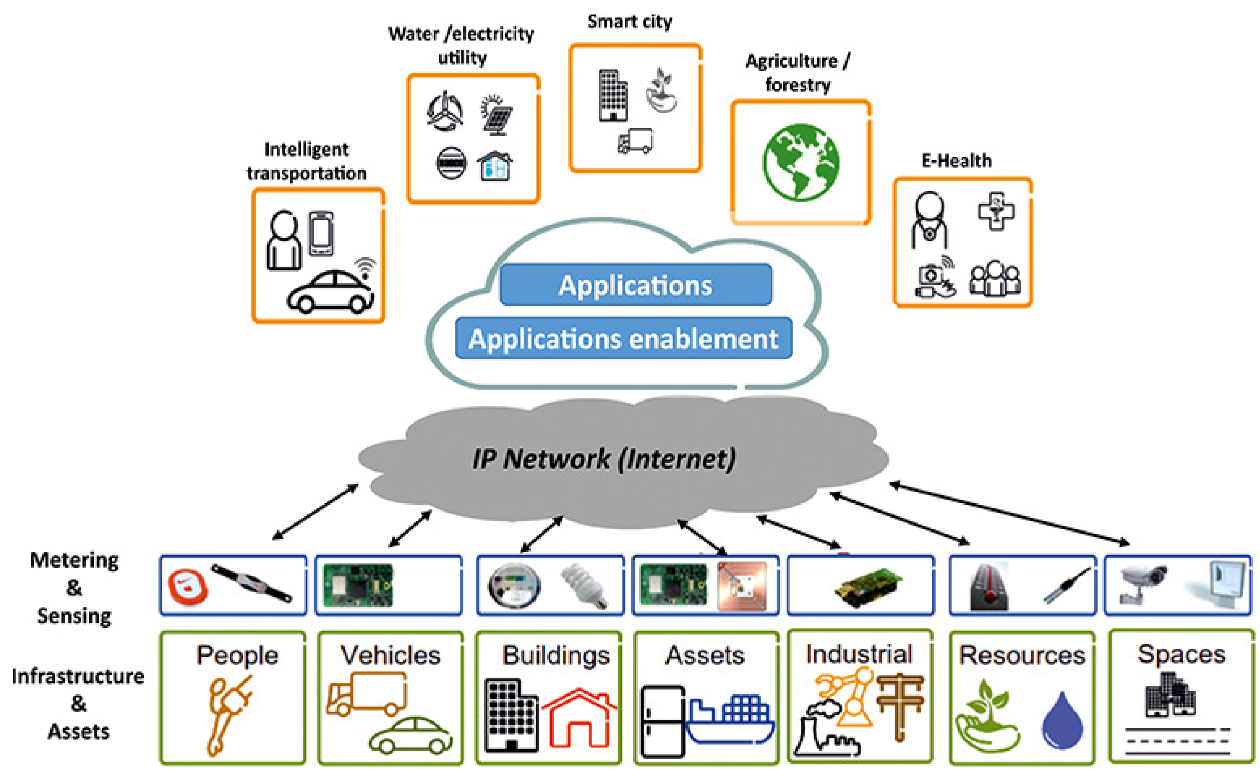
Outline II

- Mathematical model
- Normalization

5. Experiments and Results
   - Model Implementation
   - Experiment scenario
   - Result Analysis

6. Conclusion
   - Work summary and contribution
   - Limitation and Future Work

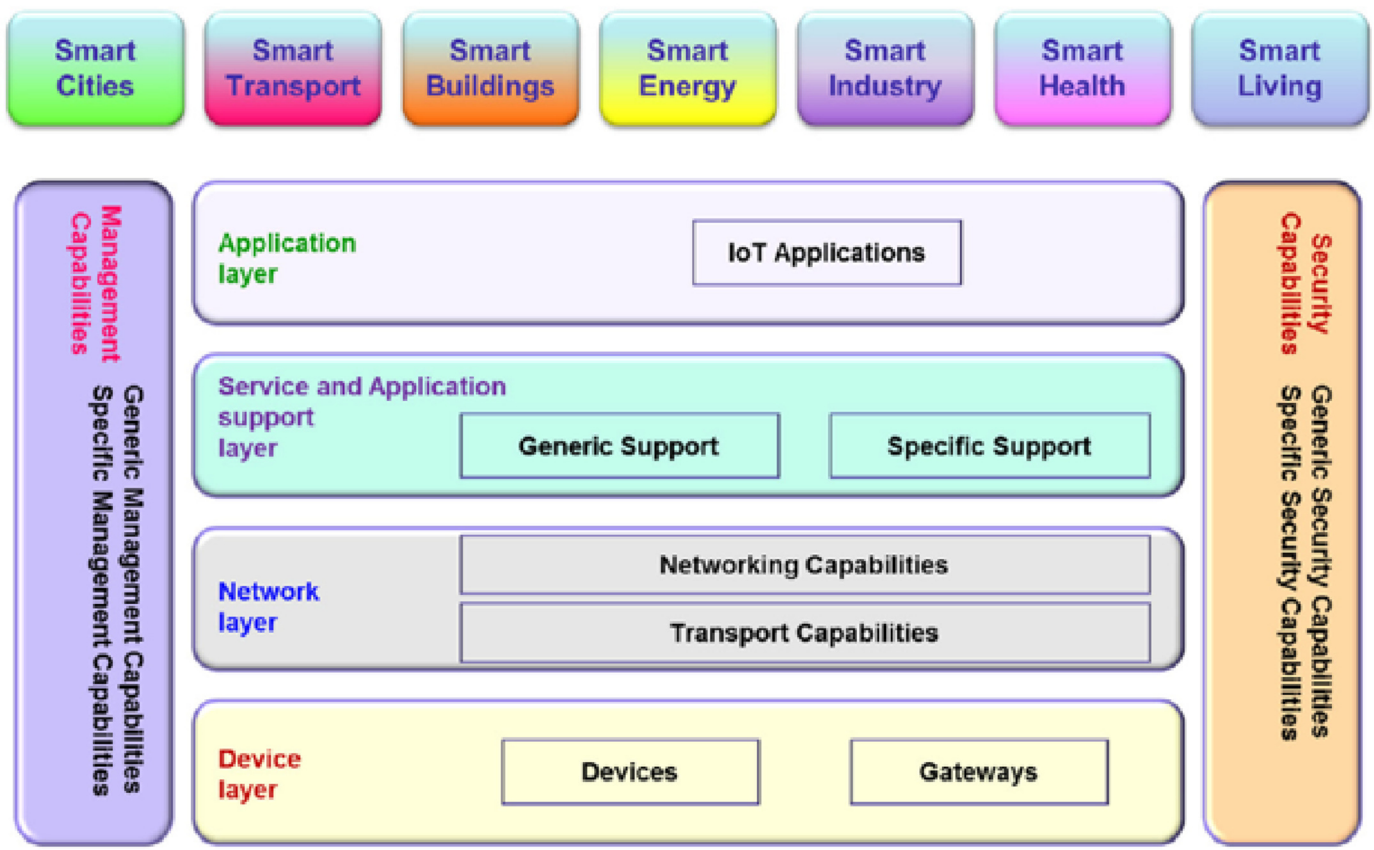
Internet of Things(IoT)



source : The university of Arizona

IoT is an ecosystem of physical objects that are accessible through the Internet.

Internet of Things(IoT) reference model



source :Overview of the Internet of things by Tatiana Kurakova

Aspects of problem

- Large numbers of heterogeneous devices

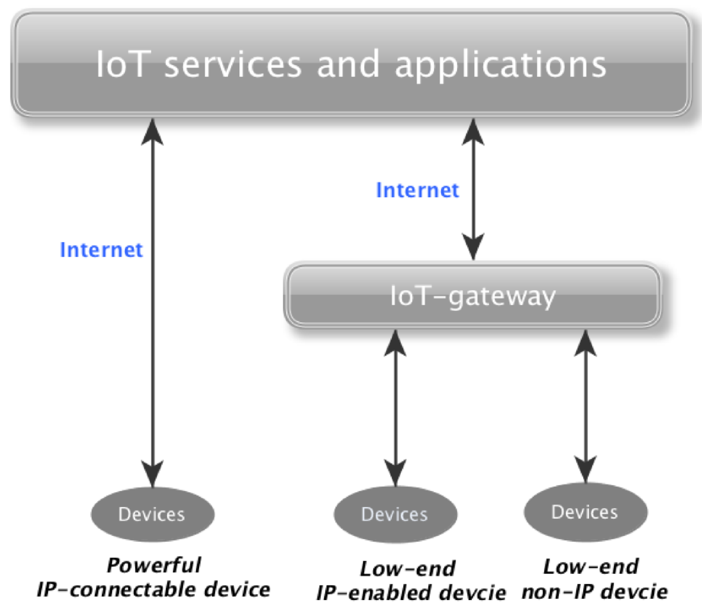
> It is predicted that the number of Internet-connected things will reach 50 billion by 2020. *Cisco/Ericsson*

- Diversity in application domain and application ranges : data-centric, innovative, and stochastic nature
- Multi-system environment and diverse SLA

QoS Indicator	
Transport Network	**Sensing Network**
Bandwidth Packet loss Jitter Delay	Data accuracy Data collection delay Sampling rate WSN lifetime WSN coverage

Aspects of problem

- Low-end IoT devices : Battery-powered and non-IP sensors



- Internet/current transport network limitation in adapting new ideas
- Shared and multi-service environment

Objective

General objective

Design a QoS management framework for IoT system which focuses on IoT infrastructure resource allocation based on the application QoS needs

Specific objective

- Propose a generic QoS support framework to enforce the diverse IoT application SLA in the resource allocation process
- Model the QoS-aware resource allocation control in core transport network/Internet
- Analyze and evaluate the performance of the proposed QoS model

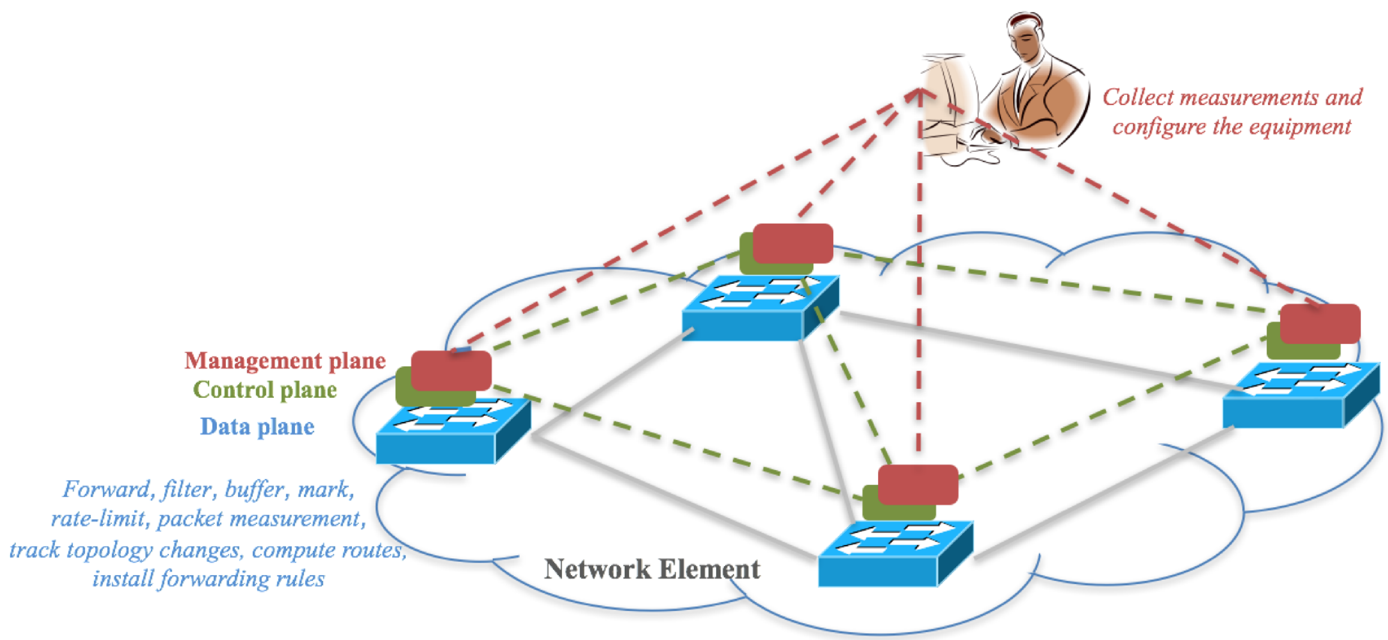
Related works

Different study aspects and approaches :

- QoS management within WSNs
  - ✓ *QoS Routing in Wireless Sensor Networks : A Survey* by R. Annie Uthra
- IoT QoS schemes and architecture
  - ✓ *Quality of service approaches in IoT : A systematic mapping* by Gary White et al.
- QoS-aware cross-layer resource allocation
  - ✓ *QoS-aware scheduling of services-oriented Internet of things by Li Ling et al.*
  - ✓ *Middleware to support sensor network applications(MiLAN)* by Heinzelman et al.

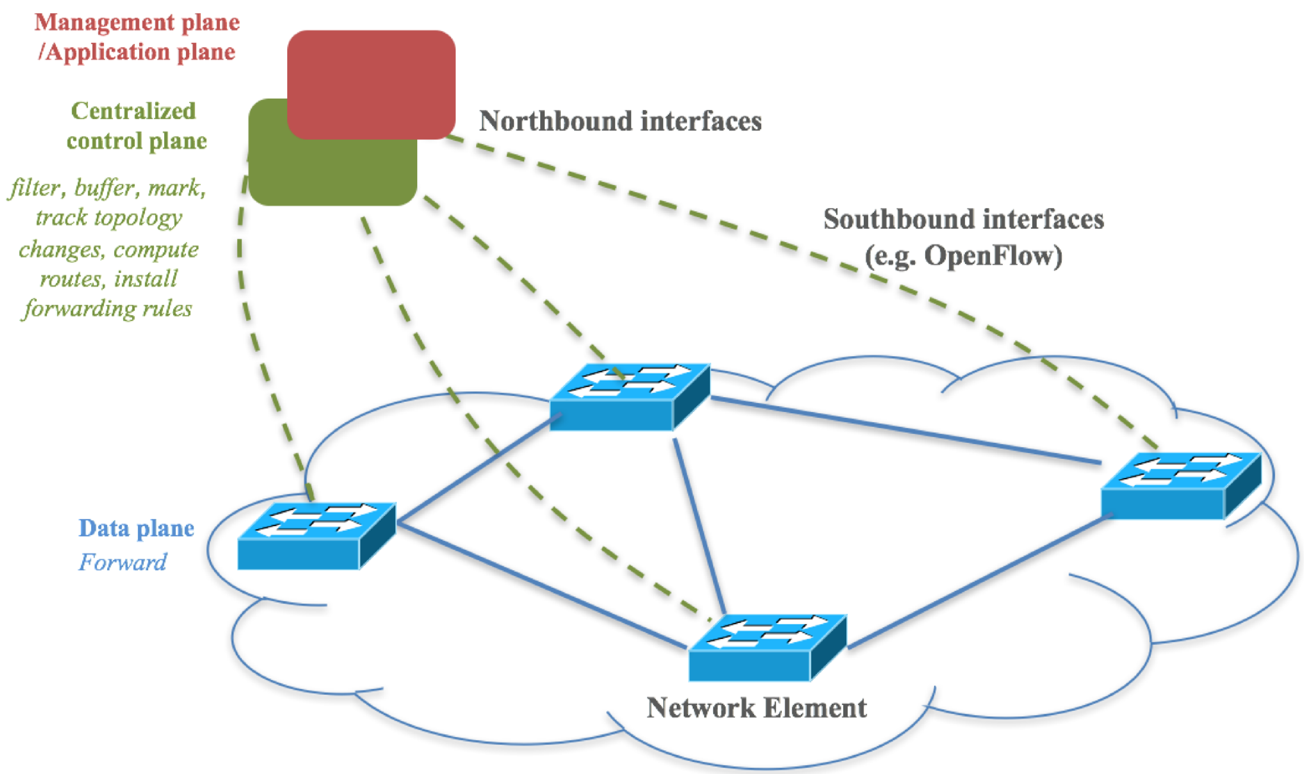
Software-defined system
SDIoT : a software defined based Internet of things framework by Yaser Jararweh

Traditional network architecture



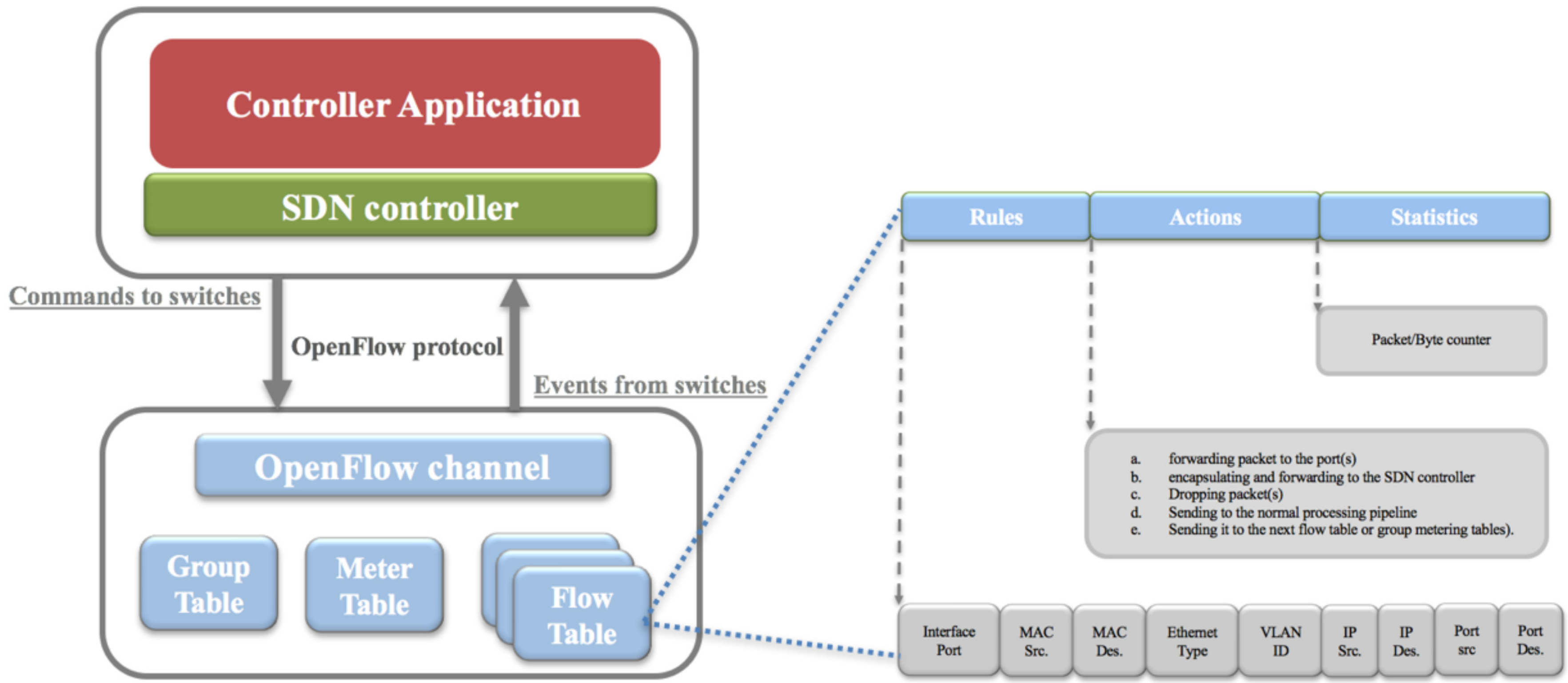
- Closed equipment
- Distributed control
- High equipment and operation cost

Software Defined Networking(SDN)



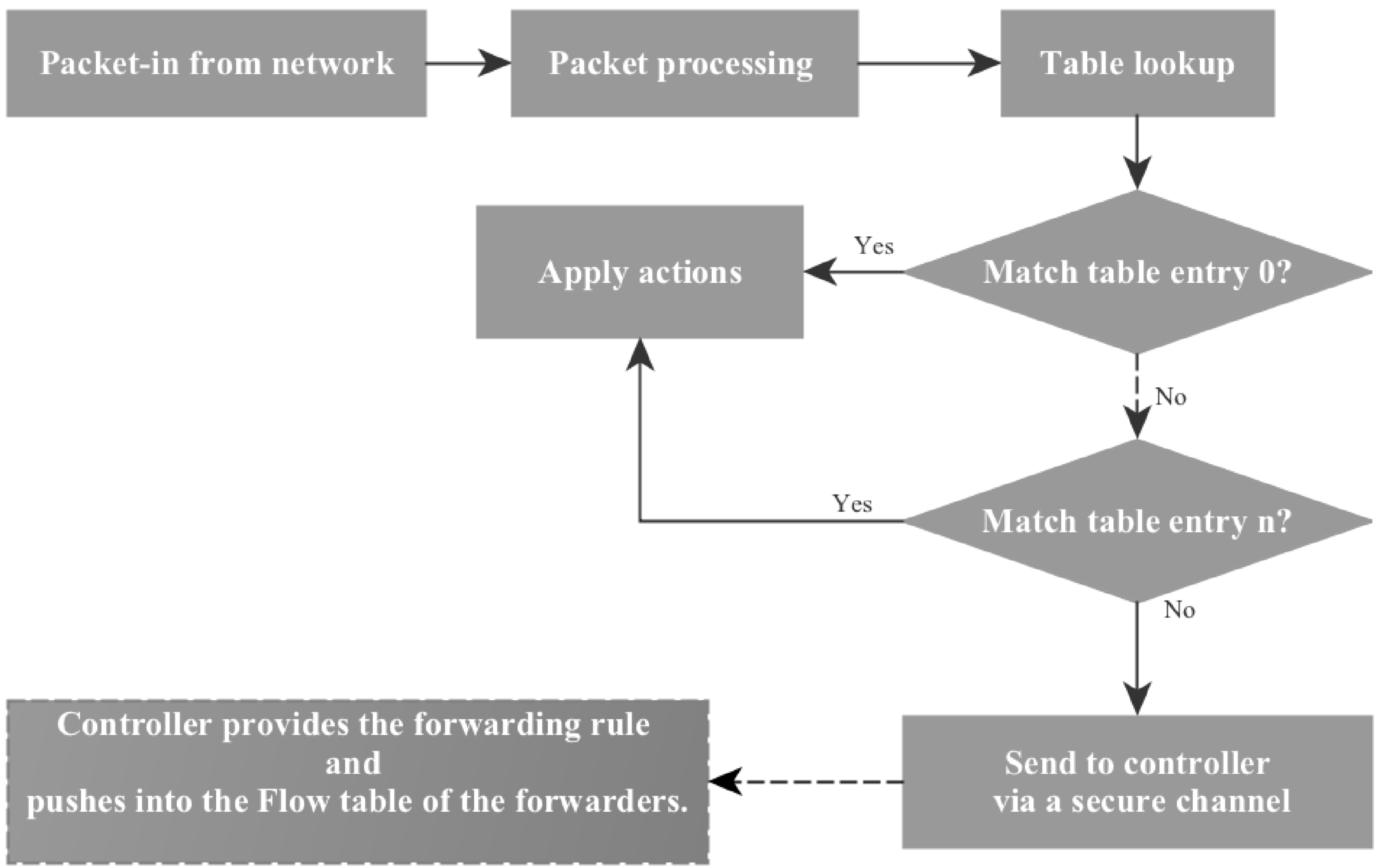
- Centralized management
- Network programmability
- CAPEX and OPEX efficiency

OpenFlow protocol

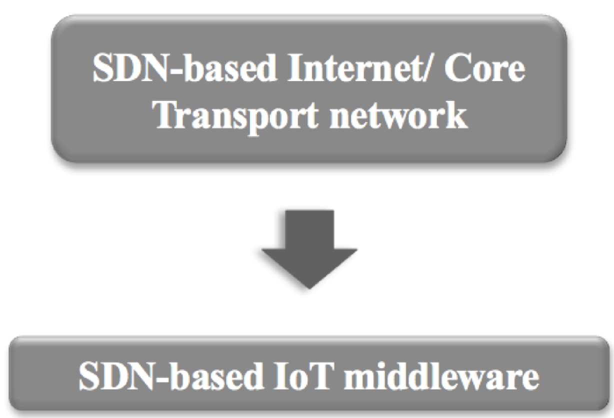


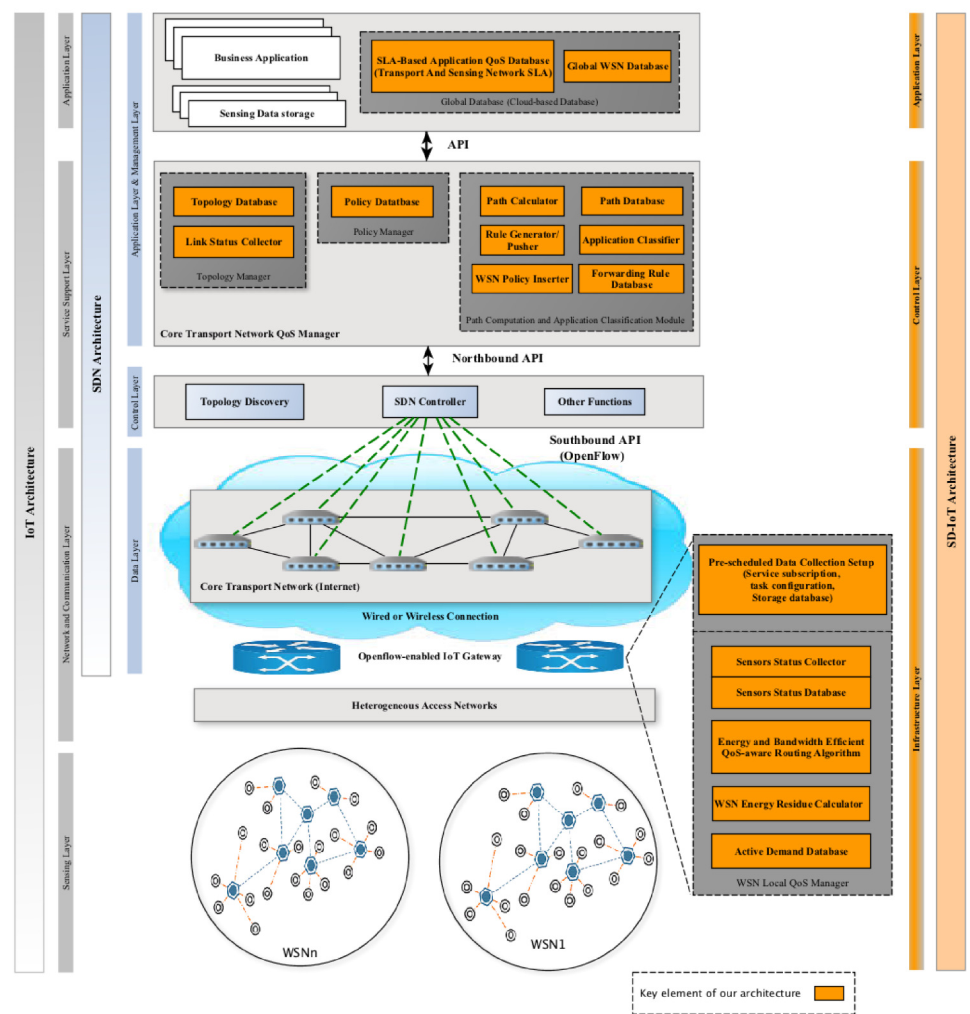
Direct over TCP/ Secure SSL channel

SDN operation mechanism



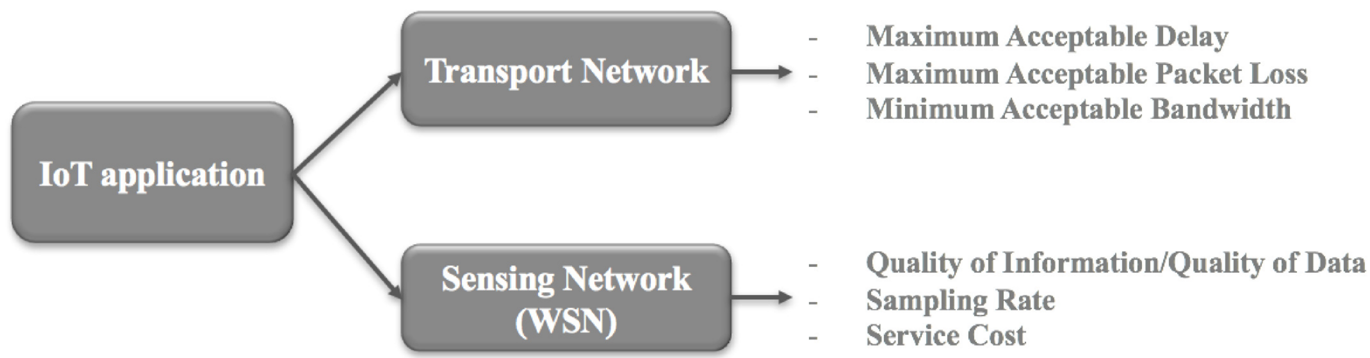
QoS support framework in IoT



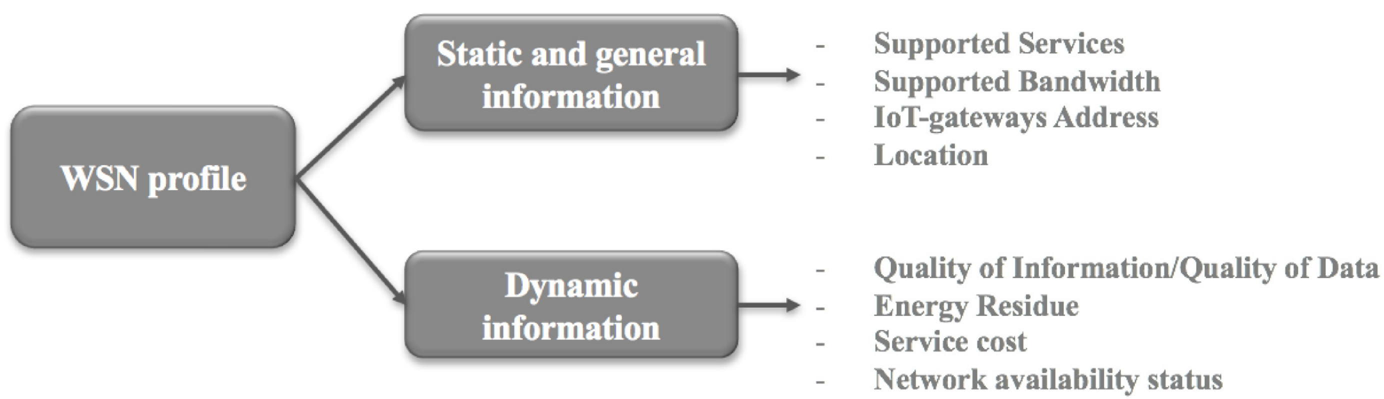


Global Database

SLA-based application QoS Database :

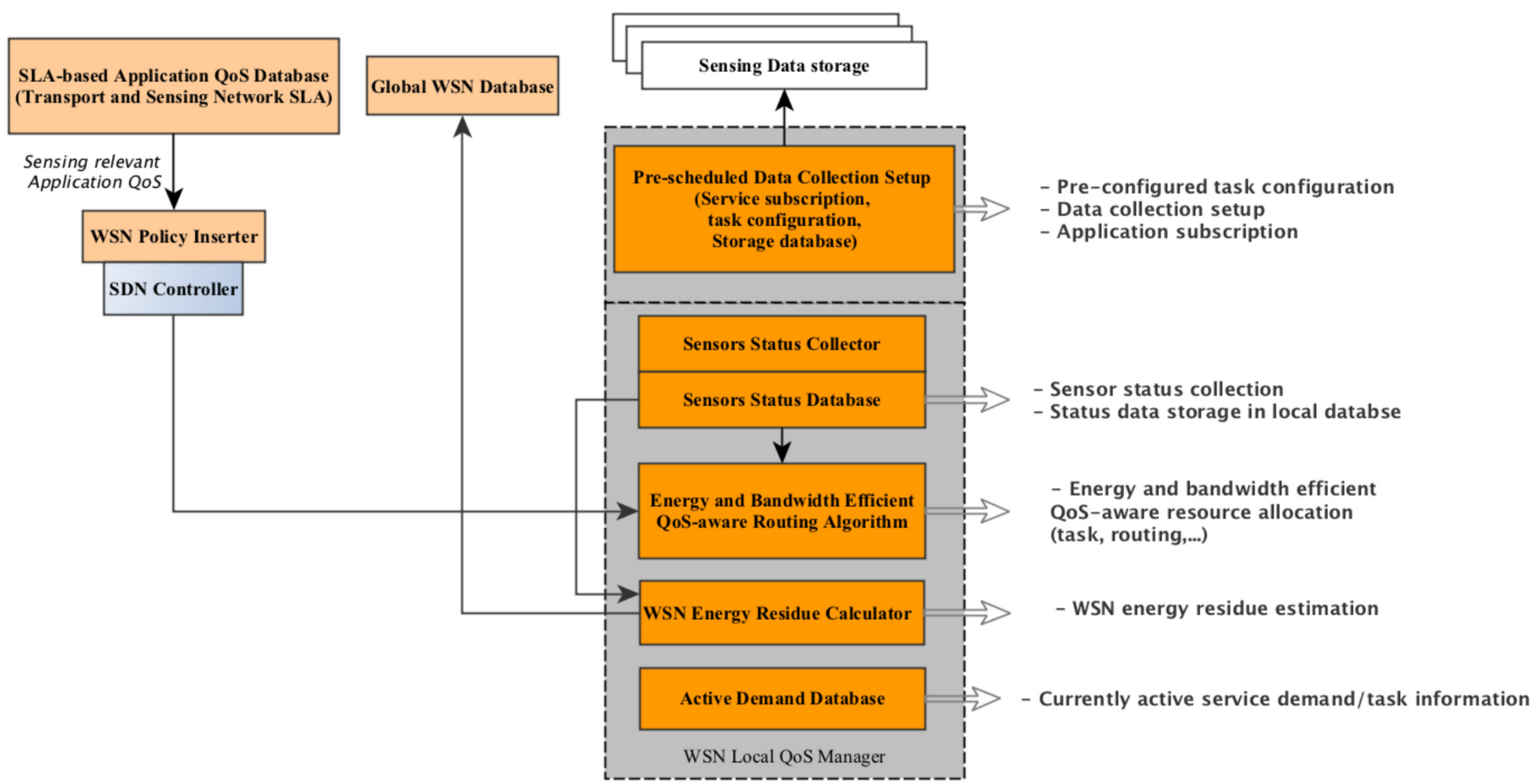


Global WSN Database :



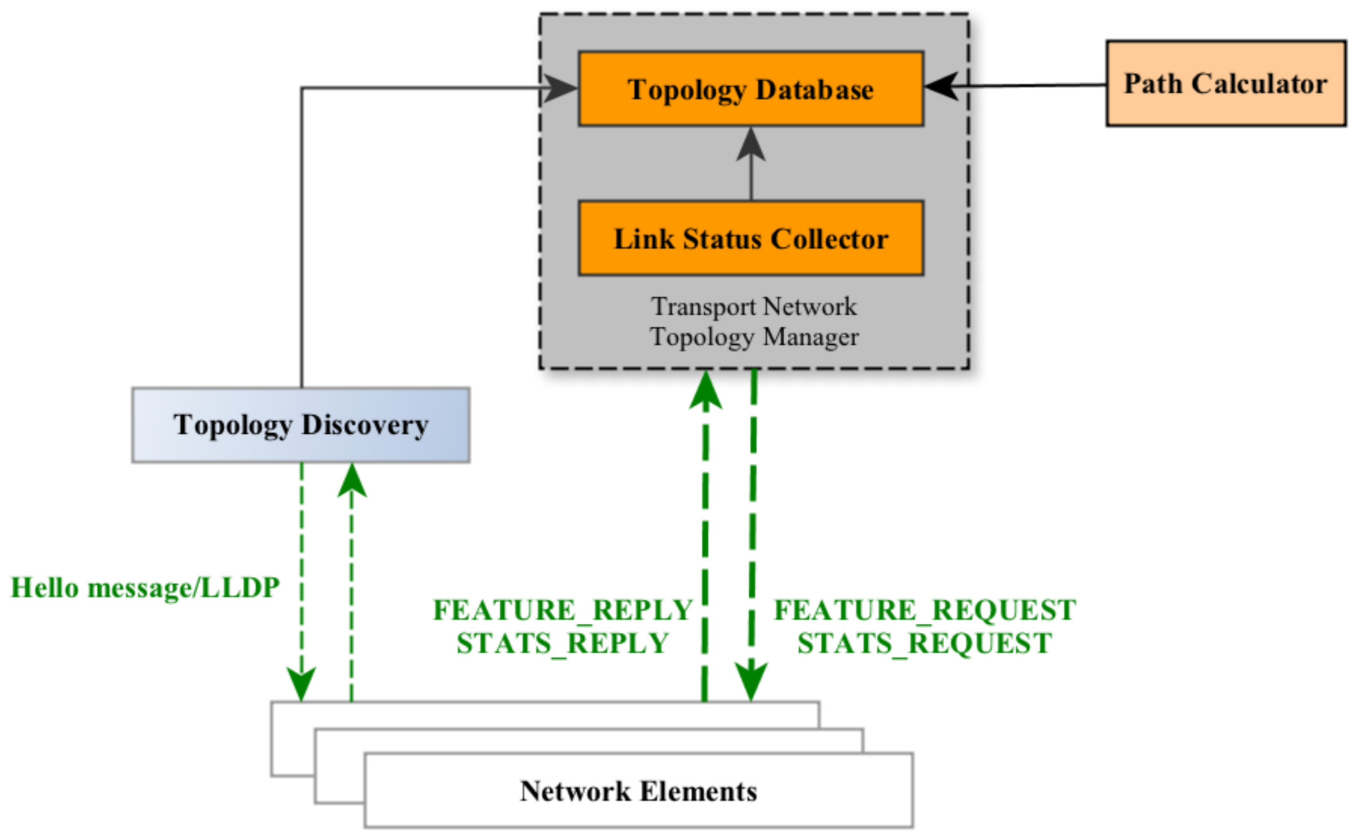
WSN Local QoS Management Module

- ✓ QoS-aware sensing resource allocation and efficient task management
- ✓ Sensor status tracking
- ✓ Sensor energy residue estimation



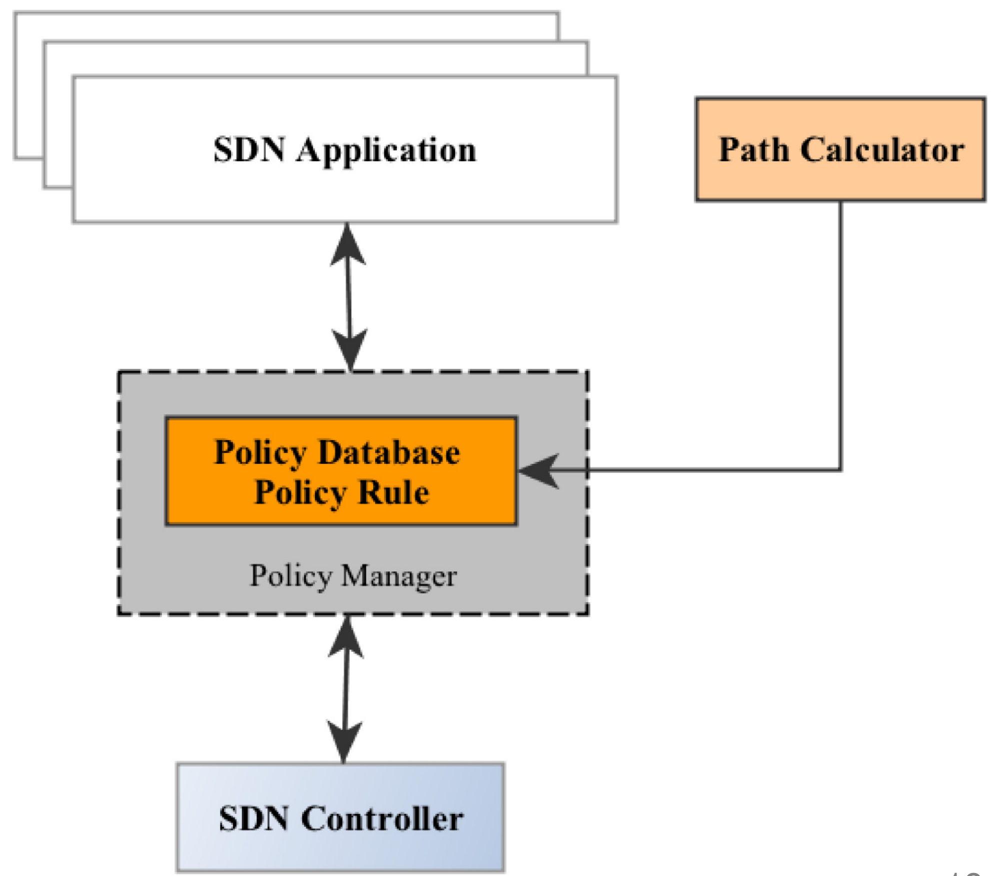
Core Transport Network Topology Management Module

- ✓ Network topology discovery
- ✓ Network link status collection in terms of QoS parameters
- ✓ Updated database



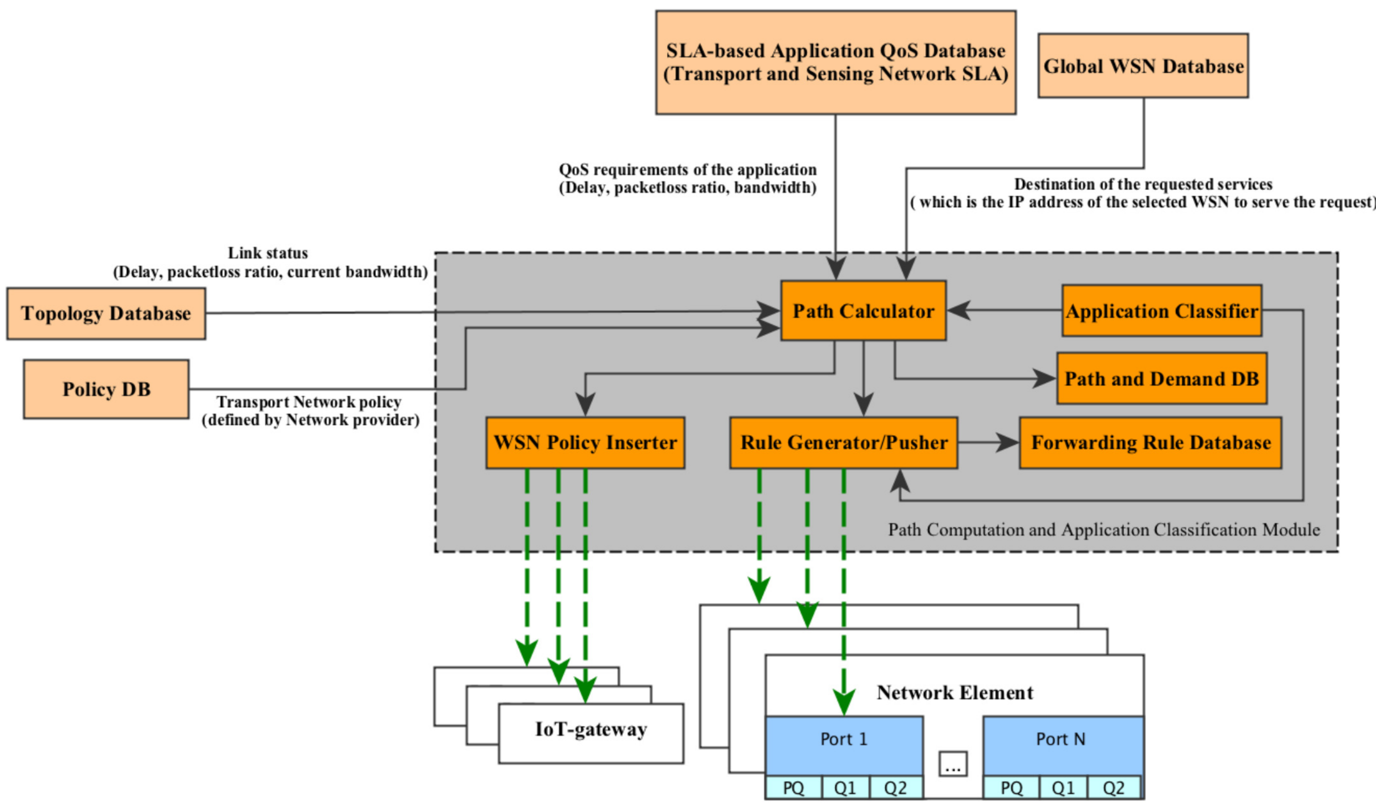
Policy Management Module

- ✓ Load balancing
- ✓ Bandwidth reservation
- ✓ SLA-violation
- ✓ Congestion



Path Computation and Application Classification Module

- ✓ End-to-end QoS support routing from source to the destination(IoT-gateway)
- ✓ WSN assignment and QoS information provider



Path Computation and Application Classification Module

Application Classifier :

Application Class	IoT application	QoS attributes	Priority	Type of queue	Cisco classification
Delay-Centric	Mission-critical (event-based application)	$D^k_{max} \leq D_{Threshold}$	1	PQ (Priority Queue)	EF (Expedited Forwarding)
Bandwidth-Centric (Multimedia application)	Real-time monitoring, query-driven application	$D^k_{max} \geq D_{Threshold}$ $BW^k_{min} \geq BW_{Threshold}$	2	Q1	AF (Assured Forwarding)
General	Non-Real time monitoring, analytic application	N/A	3	Q2	BE (Best Effort)

Queuing/Scheduling techniques :

- Queuing model
- Complete Buffer Sharing
- Preemptive Priority Scheduling

Path Computation and Application Classification Module

Path Calculator :

- ✓ WSNs determination for the requested services
- ✓ QoS support routing path calculation across the core transport network

Path and Demand Database :

- ✓ Database of currently active demand information and the associated paths

Rule Generator/Pusher :

- ✓ Flow rules generation based on the calculated paths
- ✓ Rule insertion in the Flow Table of network elements along the paths.

Path Computation and Application Classification Module

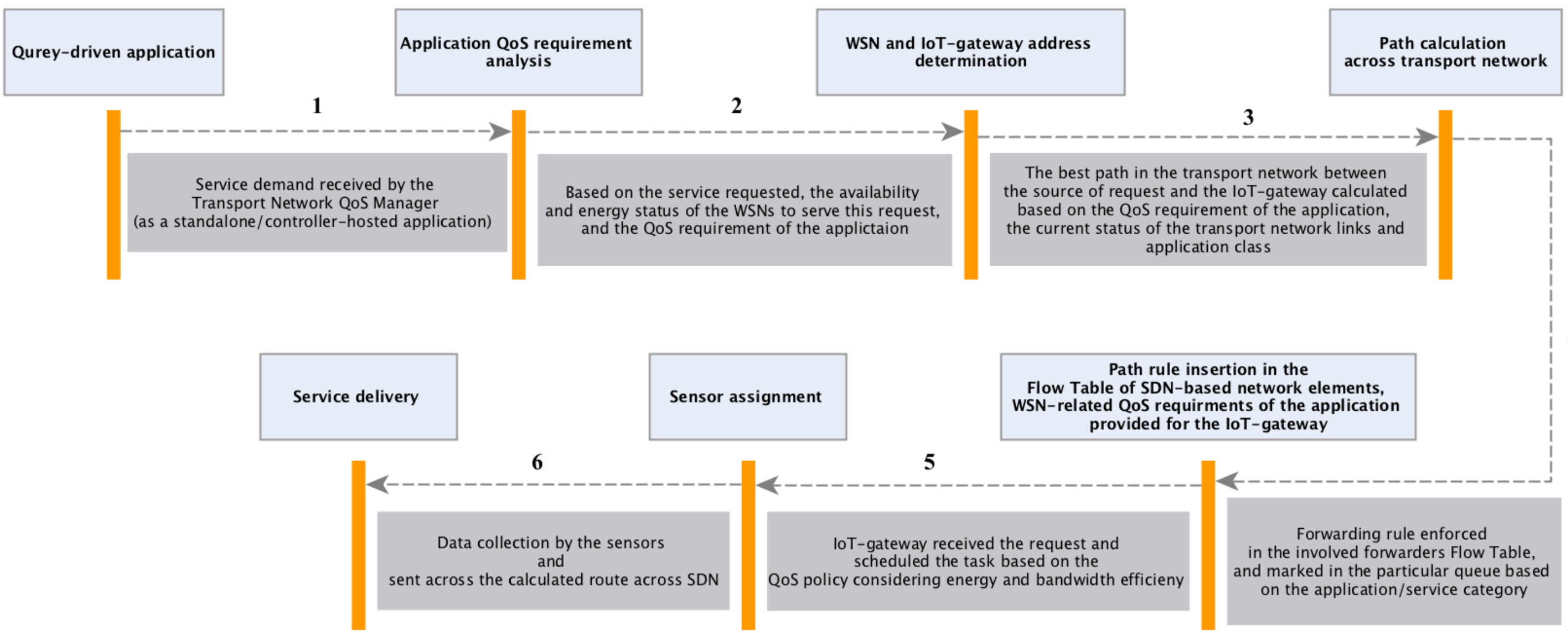
Forwarding Rule Database :

- ✓ Database of the configured flow rules (always updated)

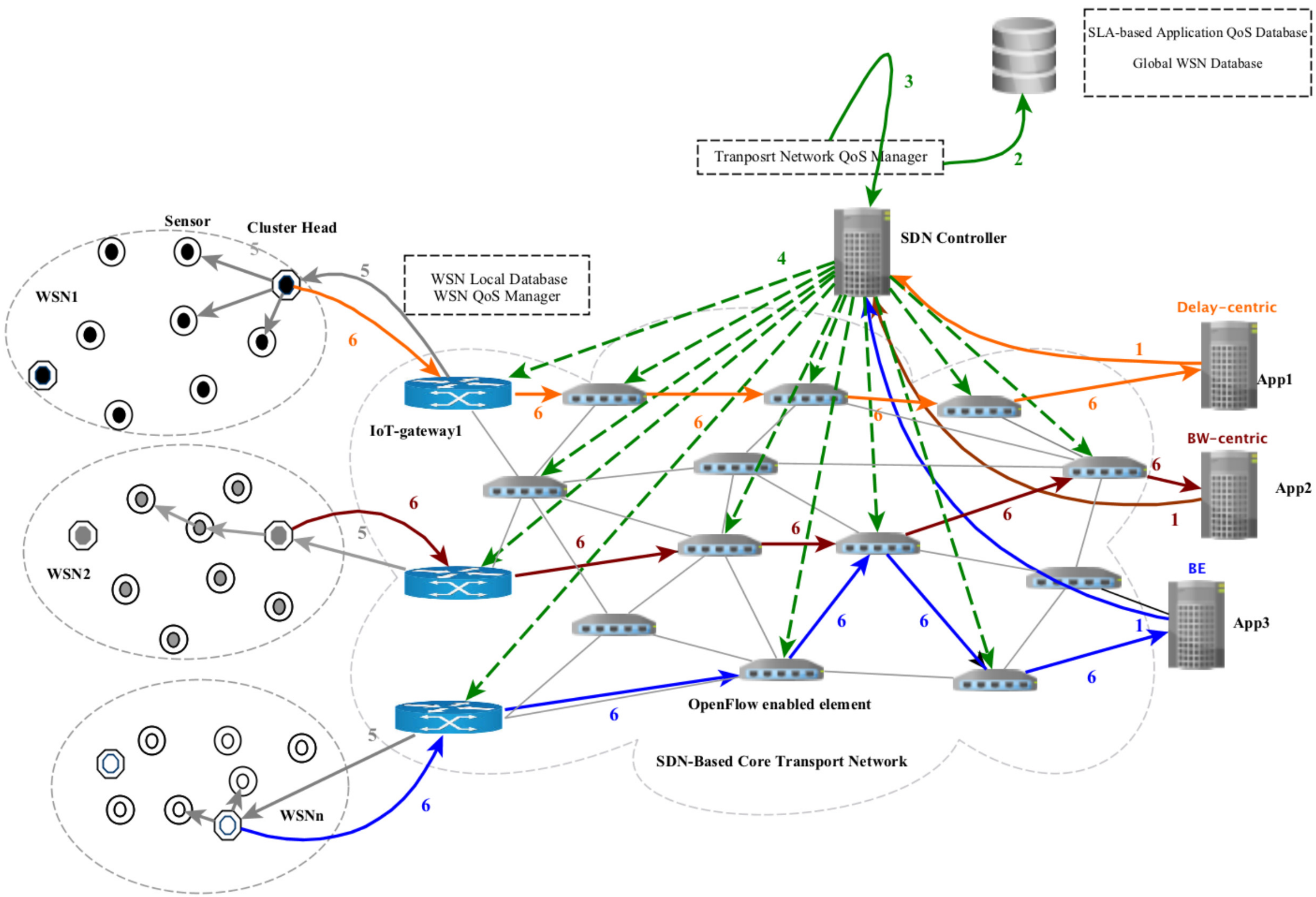
WSN policy pusher :

- ✓ Sensing-relevant application QoS requirements insertion in the determined IoT-gateway

Query-driven application scenario

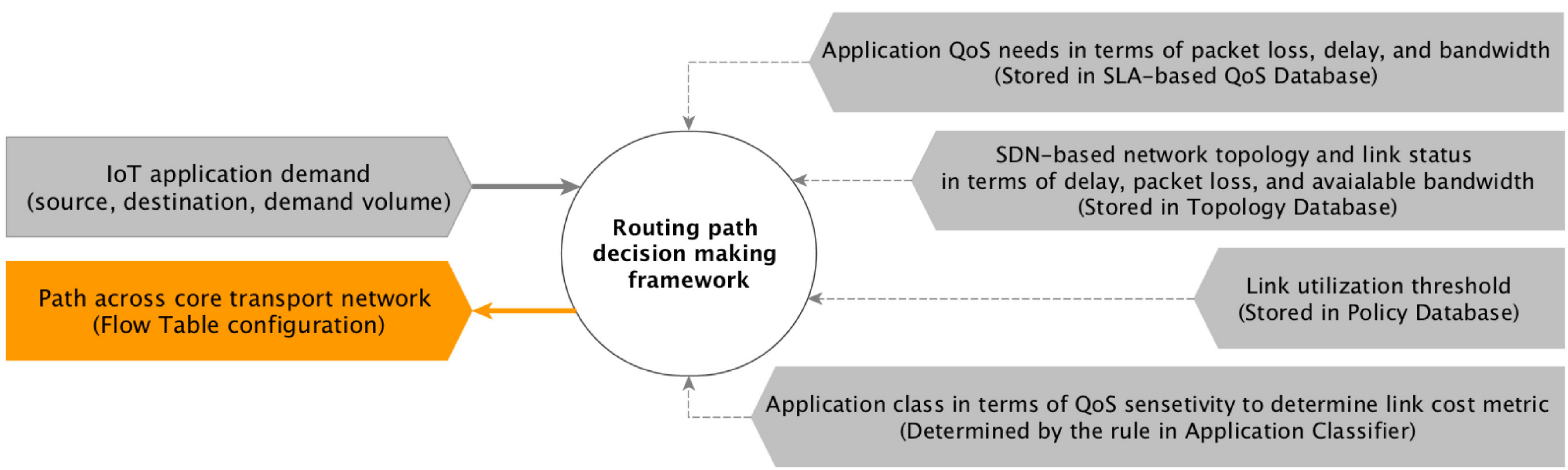


Query-driven application scenario



Making-decision framework

Status-aware and QoS-aware routing algorithm across core transport network :



QoS routing algorithm

Algorithm 1: Routing path algorithm to find the least-cost possible path across the core transport network

Procedure ;
Input : $G = (V, E)$ as the topology graph of the SDN network including nodes and bidirectional links : $V = \{1, 2, ..., v\}$ and $E = \{(i, j) : i, j \in V, i \neq j\}$
for k *in* K **do**
    Input : Source S^k, Destination T^k and Volume F^k
    Output : Paths from Source S^k to Destination T^k
end
for (i, j) *in* E **do**
    Read the link QoS parameters including b_{ij}, pl_{ij}, *and* d_{ij};
    Calculate the current link utilization rate;
    Read the link utilization limit $u_{Threshold}$;
    if *Link utilization rate* $>= u_{Threshold}$ **then**
        It excludes this link from the logical network topology used to calculate the path ;
    end
end
for k *in* K **do**
    Read SLA-DB to have Max acceptable delay D^k_{SLA}, *Max acceptable packet loss* PL^k_{SLA}, *and Min required bandwidth* B^k_{SLA};
    Set the link cost metrics depending on the application class;
end
for k *in* K **do**
    Make decision about the path based on the mathematical model;
end

Parameters

- Graph $G = (V, E)$ as the SDN-based Transport Network Topology :
  $V = \{1, ..., v\}$ as the list of network elements,
  $E = \{(i, j) : i, j \in V, i \neq j\}$ as list of bidirectional links between network elements.

- Network link status and network policy parameters :

$B_{ij} > 0$	Maximum capacity on the link(i, j), $[Mbps]$
$b_{ij} \geq 0$	Available capacity on the link(i, j), $[Mbps]$
$d_{ij} \geq 0$	Delay on the link(i, j), $[Second]$
$pl_{ij} \geq 0$	Packet loss ratio on the link(i, j), $[Percentage]$
$u_{Threshold} > 0$	Link utilization limit on the link, $[Percentage]$

Parameters

- Service demand parameters :

$S^k \in V$	Source of demand k
$T^k \in V$	Destination of demand k
$F^k \geq 0$	Total demand volume k
$D^k_{SLA} \geq 0$	Maximum acceptable delay for demand k, agreed in SLA
$PL^k_{SLA} \geq 0$	Maximum acceptable packet loss ratio for demand k, agreed in SLA
$B^k_{SLA} \geq 0$	Minimum bandwidth required for demand k, agreed in SLA
P^k	Selected path across the network for demand k

Objective function

$$\text{Minimize} \quad \sum_{(i,j)\in E} \sum_{k\in K} C_{ij} X_{ij}^k$$

X_{ij}^k :Amount of volume corresponding to the demand k on the link(i,j), $[Mbps]$

$$0 \leq X_{ij}^k \leq b_{ij}$$

C_{ij} :Cost of the link(i,j)

$$C_{ij} = \alpha * b_{ij} + \beta * pl_{ij} + \gamma * d_{ij} \quad , \forall (i,j) \in E$$

$\alpha, \beta,$ and γ : Scaling factor.

$$\alpha + \beta + \gamma = 1 \qquad , 0 \leq \alpha, \beta, \gamma \leq 1$$

Constraint function

- Link Bandwidth Constraint :

$$\sum_{k\in K} X_{ij}^k \leq b_{ij},\ \forall(i,j) \in E$$

- Link Utilization Constraint :

$$\sum_{k\in K} X_{ij}^k \leq b_{ij} - (1 - u_{Threshold}) * B_{ij},\ \forall(i,j) \in E$$

- Flow Conservation Law :

$$\sum_{(i,j)\in E} X_{ij}^k - \sum_{(j,i)\in E} X_{ji}^k = \begin{cases} F^k, & i = S^k \\ -F^k, & i = T^k \\ 0, & i \neq S^k and i \neq S^k T^k. \end{cases}$$

Constraint

- Delay Constraint :

$$d_p^k \leq D_{SLA}^k$$

$$d_p^k = \sum_{(i,j) \in E, p^k} d_{ij}$$

$$\sum_{(i,j) \in E, P^k} d_{ij} \leq D_{SLA}^k, \ \forall k \in K$$

- Packet Loss Constraint :

$$pl_p^k \leq PL_{SLA}^k$$

$$pl_p^k = \sum_{(i,j) \in E, p^k} pl_{ij}$$

$$\sum_{(i,j) \in E, P^k} pl_{ij} \leq PL_{SLA}^k, \ \forall k \in K$$

Link cost formula normalization

$$C_{ij} = \alpha * b_{ij} + \beta * pl_{ij} + \gamma * d_{ij} \quad ,\forall (i,j) \in E$$

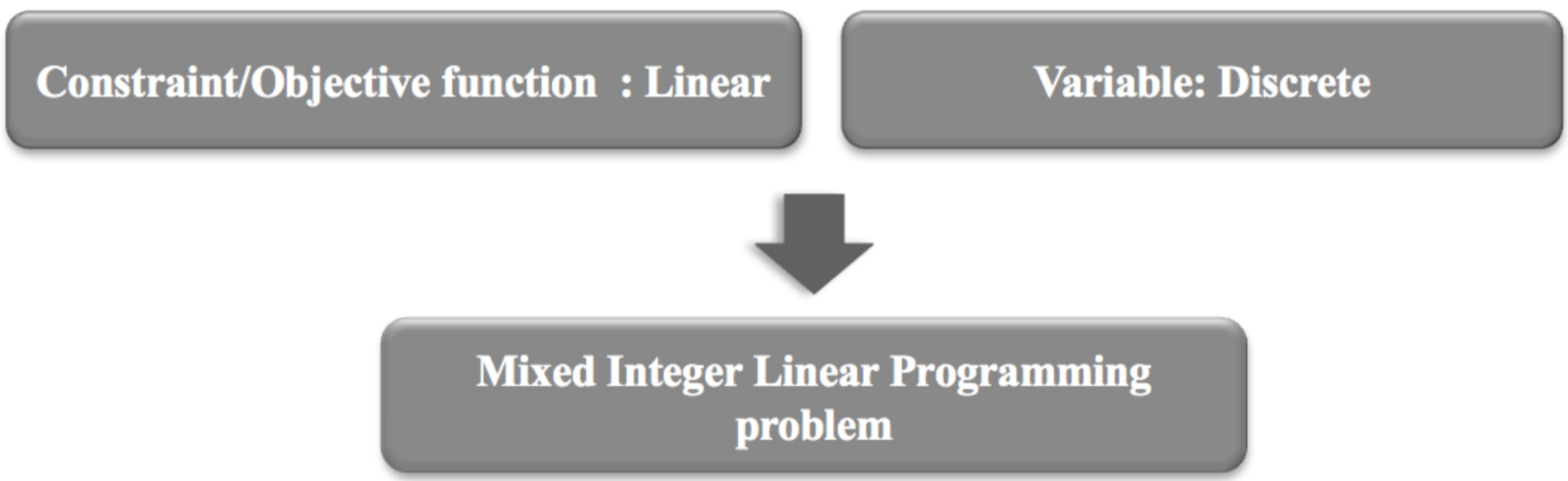
Feature scaling method

$$b'_{ij} = \frac{b_{max} - b_{ij}}{b_{max} - b_{min}} \quad , b_{min} \le b_{ij} \le b_{max} \quad , b_{max} \ne b_{min}$$

$$pl'_{ij} = \frac{pl_{ij} - pl_{min}}{pl_{max} - pl_{min}} \quad , pl_{min} \le pl_{ij} \le pl_{max} \quad , pl_{max} \ne pl_{min}$$

$$d'_{ij} = \frac{d_{ij} - d_{min}}{d_{max} - d_{min}} \quad , d_{min} \le d_{ij} \le d_{max} \quad , d_{max} \ne d_{min}$$

Model implementation



Tools

AMPL: A Mathematical Programming Languages

CPLEX: Integer linear programming solver

Assumption

- QoS parameters bound for normalization purpose :

	Minimum	Maximum
Link packet loss ratio range	0%	5%
Link delay range	0 s	0.0001 s
Link bandwidth range	0 bps	1000Mbps

- Link utilization rate limit= 75%
- Link cost metrics :

Application class	Link cost metric
Delay-centric	Delay, Packet Loss Ratio
BW-centric	BW, Packet Loss Ratio
BE	Delay, Packet Loss ratio, BW (or BE algorithm)

Experiment method

Demand Class

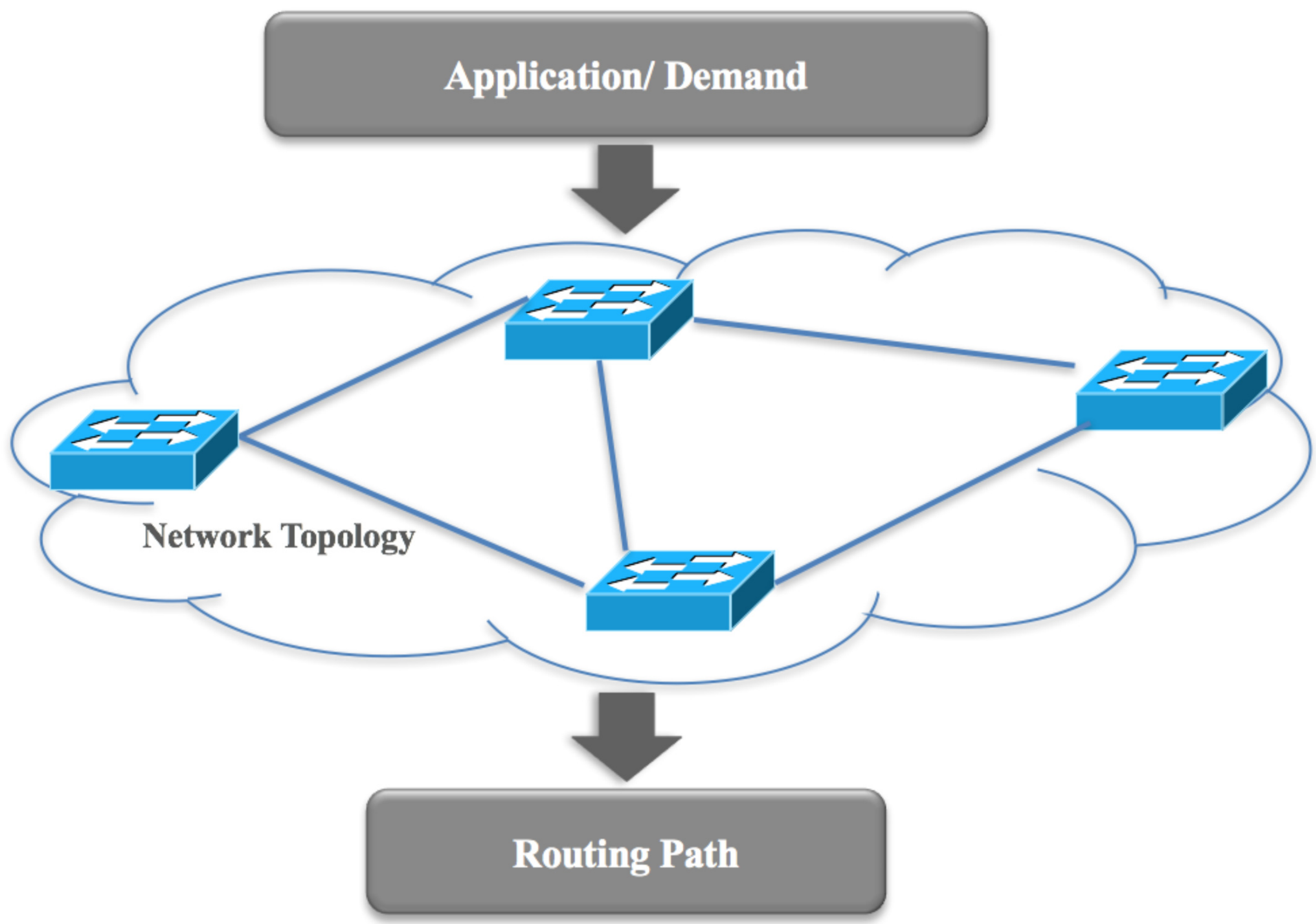
Delay-centric
BW-centric

Routing Model

Our proposed model
OSPF

Path Analysis

Delay
Packet Loss ratio
Link Utilization Rate

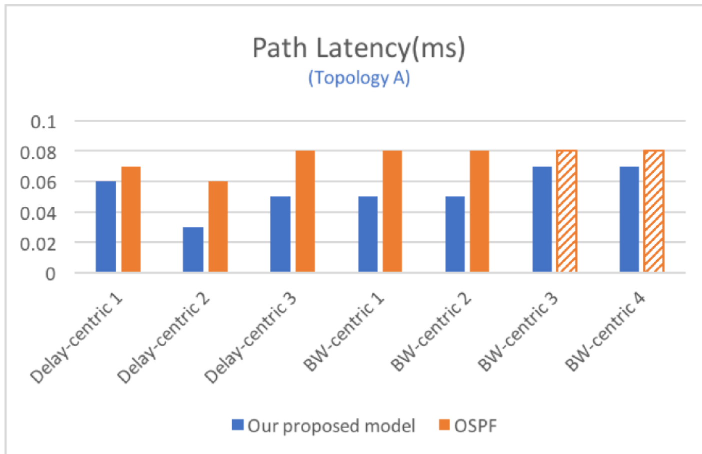


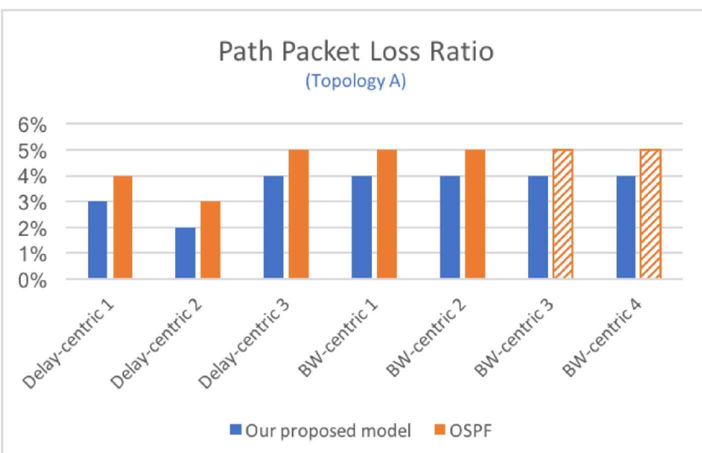
Single-demand and Multi-demand scenario

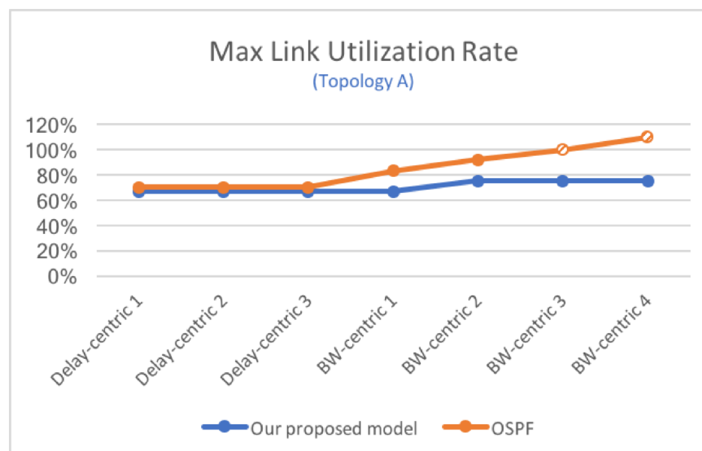
	Single-demand	
Topology A	Delay-centric	3
	BW-centric	4
Topology B	Delay-centric	3
	BW-centric	4
Topology C	Delay-centric	4
	BW-centric	4

	Multi-demand	
Topology A	Test 1	1 Delay-centric, 1 BW-centric
	Test 2	2 Delay-centric, 1 BW-centric
	Test 3	1 Delay-centric, 2 BW-centric
Topology B	Test 1	2 Delay-centric, 1 BW-centric
	Test 2	2 Delay-centric, 2 BW-centric
	Test 3	2 Delay-centric, 3 BW-centric
Topology D	Test 1	2 Delay-centric, 1 BW-centric
	Test 2	3 Delay-centric, 2 BW-centric
	Test 3	3 Delay-centric, 3 BW-centric

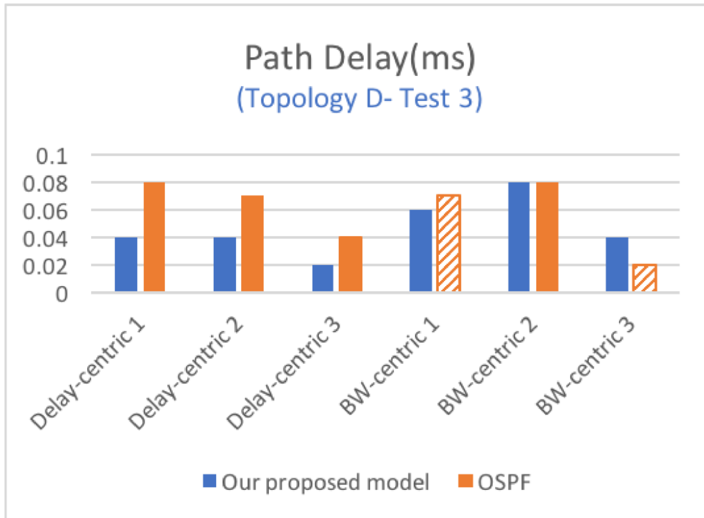
Example1 : Single demand- Topology A

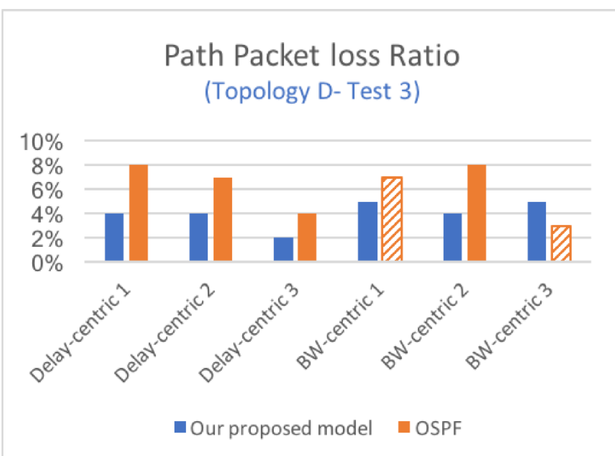


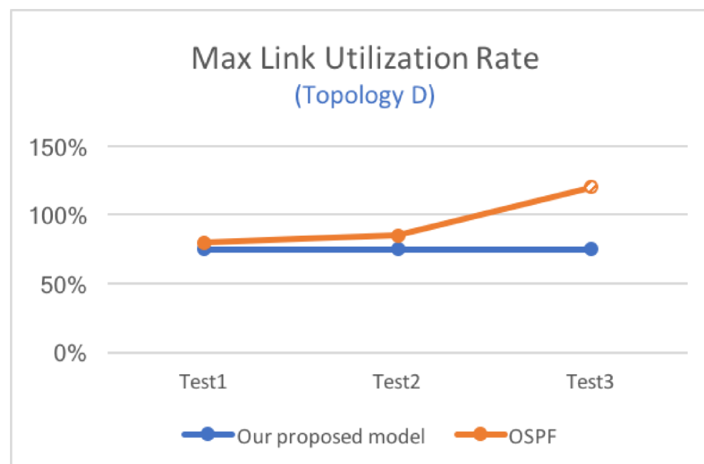




Example2 : Multi demand- Topology D- Test 3

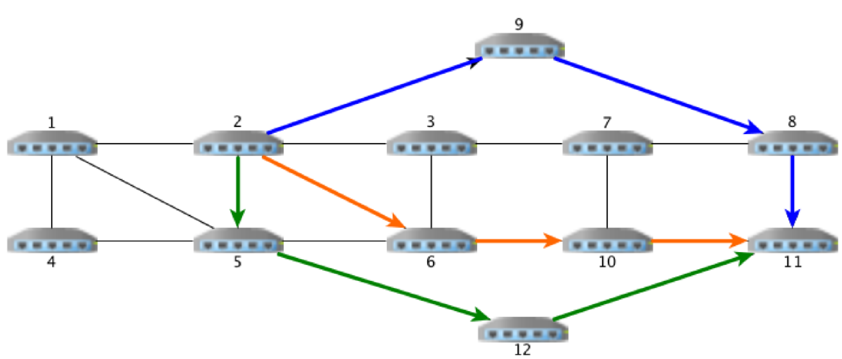


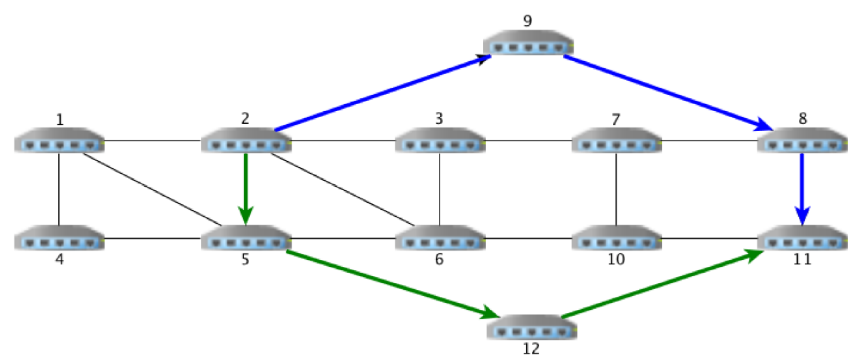




Example3 : Dynamic value of link utilization limit and multi-path approach

Topology C,Demand="BW-centric 4", Source= "node 2", Destination="node 11" :





$U_{Threshold} = 75\%$	$U_{Threshold} = 90\%$
$2 \rightarrow 9 \rightarrow 8 \rightarrow 11$	$2 \rightarrow 9 \rightarrow 8 \rightarrow 11$
$2 \rightarrow 6 \rightarrow 10 \rightarrow 11$	$2 \rightarrow 5 \rightarrow 12 \rightarrow 11$
$2 \rightarrow 5 \rightarrow 12 \rightarrow 11$	

Result Analysis

Latency/Loss Rate

- **Minimized delay and error probability for delay-application, even for BW-centric application**
- **Maximize the application efficiency**
- **Improve customer satisfaction**
- **SLA-respected QoS needs**

Link Utilization Rate

- **Less maximum link utilization**
- **Link load balancing**
- **Minimized congestion probability**
- **Increase network availability**
- **Improve customer satisfaction**

Work summary

- ✓ Common QoS support framework to provide QoS-aware cross layer resource allocation.
- ✓ QoS-aware resource allocation model in core transport network/Internet
- ✓ Analyze and evaluate the performance of the proposed QoS model

Architecture and model advantage

Architecture :

- ✓ The infrastructure resource abstraction from the application layer (SDN-based middleware)
- ✓ Implementation of the generic support services for the IoT system
- ✓ Efficient centralized network management(cost and physical resource)
- ✓ Programmable and innovative nature
- ✓ Flexible, scalable, and fast adaptive based on business goal changes

Architecture and model advantage

QoS model :

- ✓ SLA-aware per-flow QoS policy
- ✓ Network status-aware
- ✓ Dynamic link metrics based on application classes
- ✓ Congestion preventative by excluding the congested links from the path calculation process
- ✓ Congestion avoidance by balancing the link load
- ✓ Dynamic policy support

Contribution

- ✓ Overcome the challenge of the dynamic definition of the SLA for the IoT application and its enforcement in the resource allocation process
- ✓ Design a flexible and programmable middleware to provide any generic support services for the IoT system.
- ✓ Provide dynamic Traffic Engineering polices and dynamic routing to guarantee requirements of various services

Limitation and future work

Limitation :

- Statistical collection in the large network
- SDN-controller as a single point of failure
- Multi-domain network

Future work :

- Learning-based control policy
- Model performance enhancement : Integration of the *k shortest path routing algorithm* in our QoS model

Q&A

Thank you.