# Delivering Application-Layer Traffic Optimization Services based on Public Routing Data at IXPs

Danny Alex Lachos Perez

Orientador: Prof. Dr. Christian Rodolfo Esteve Rothenberg

INTRIG INFORMATION & NETWORKING TECHNOLOGIES RESEARCH & INNOVATION GROUP

FEEC Faculdade de Engenharia Elétrica e de Computação

UNICAMP

# Agenda

1. Introduction
2. Background
3. Design of ALTO-as-a-Service (AaaS)
4. AaaS Prototype
5. Experimental Evaluation
6. Conclusions / Future Work

# Distributed Applications

File sharing applications, Content Delivery Networks (CDNs), real-time communication, among others, use a significant amount of network resource to connect nodes across the Internet and transfer a large amount of data.

# Distributed Applications

- ✓ Topological information used only has a localized view provided by the ISPs.
- ✓ Without knowledge of the underlying network topology.
- ✓ Selection of resources are provided randomly.
- ✓ Impacting both applications and networking infrastructure.

# Introduction

MOTIVATION

Application-Layer Traffic Optimization (ALTO)

ISP => Focus of ALTO implementations

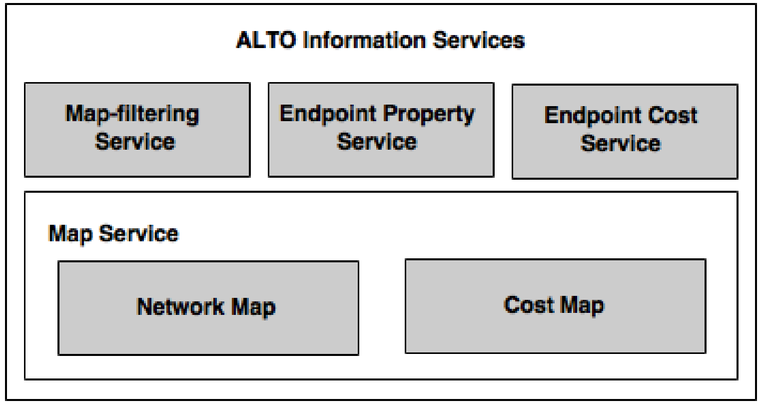
Internet eXchange Point (IXP)

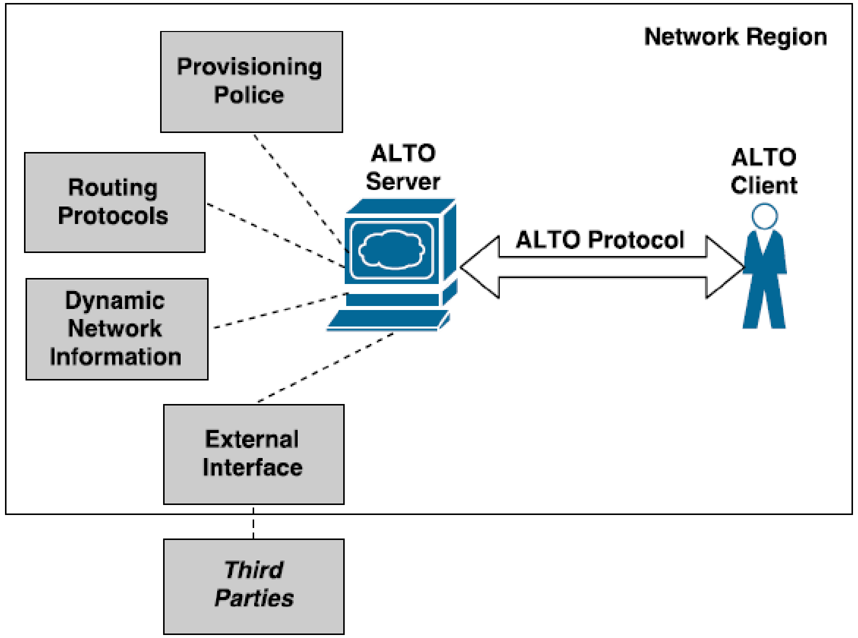
Third-parties => ALTO Information <=ALTO Protocol

# Agenda

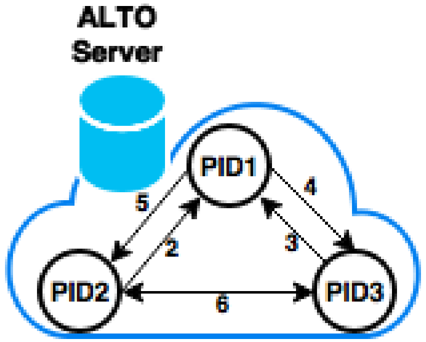
1. Introduction
2. **Background**
3. Design of ALTO-as-a-Service (AaaS)
4. AaaS Prototype
5. Experimental Evaluation
6. Conclusions / Future Work

## Introduction :
# Application-Layer Traffic Optimization (ALTO)








| | |
|---|---|
| *PID1* | 198.51.100.0/24<br>192.0.2.0/24 |
| *PID2* | 203.0.113.0/24 |
| *PID3* | 2001:db8::/32 |

NETWORK MAP

| *Cost* | *PID1* | *PID2* | *PID3* |
|---|---|---|---|
| *PID1* | 0 | 5 | 4 |
| *PID2* | 2 | 0 | 6 |
| *PID3* | 3 | 6 | 0 |

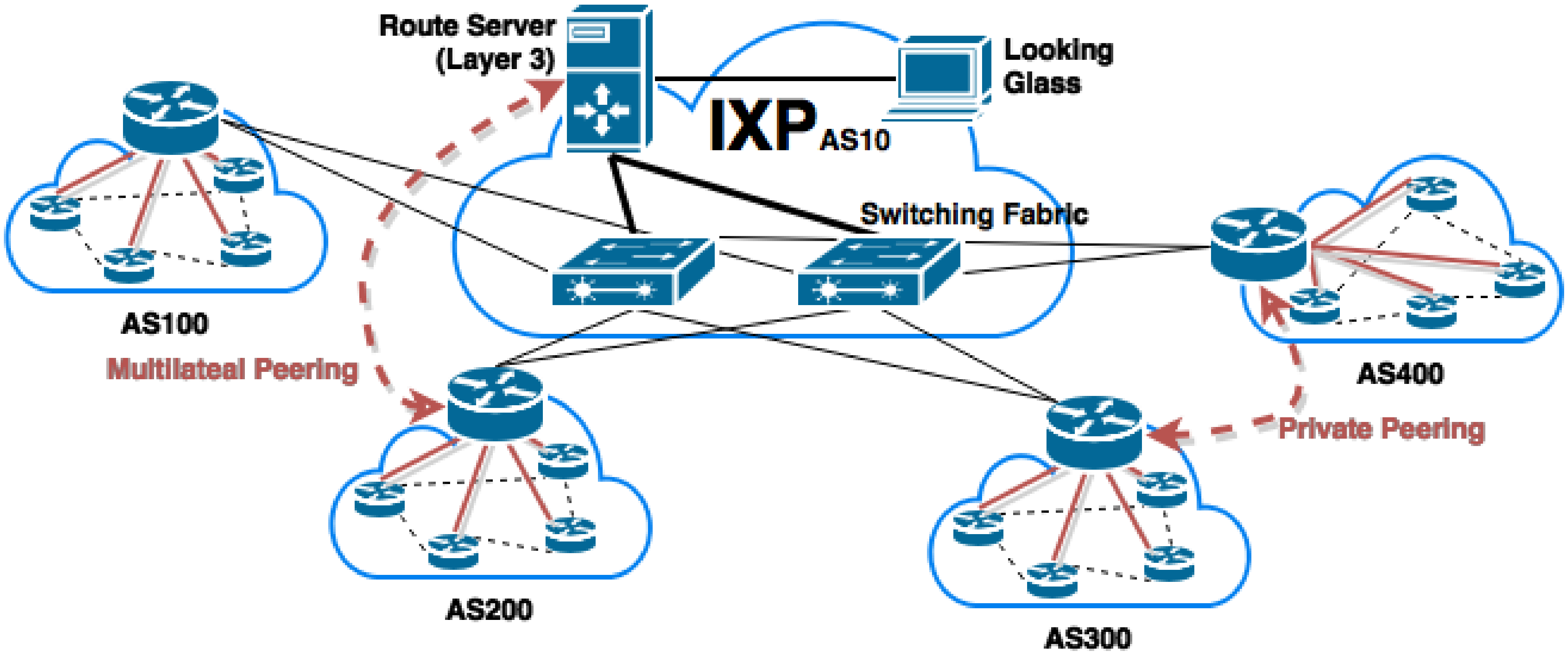
COST MAP


RESTful API
GET PUT POST DELETE

{JSON}

Introduction :

# Internet eXchange Point (IXP)



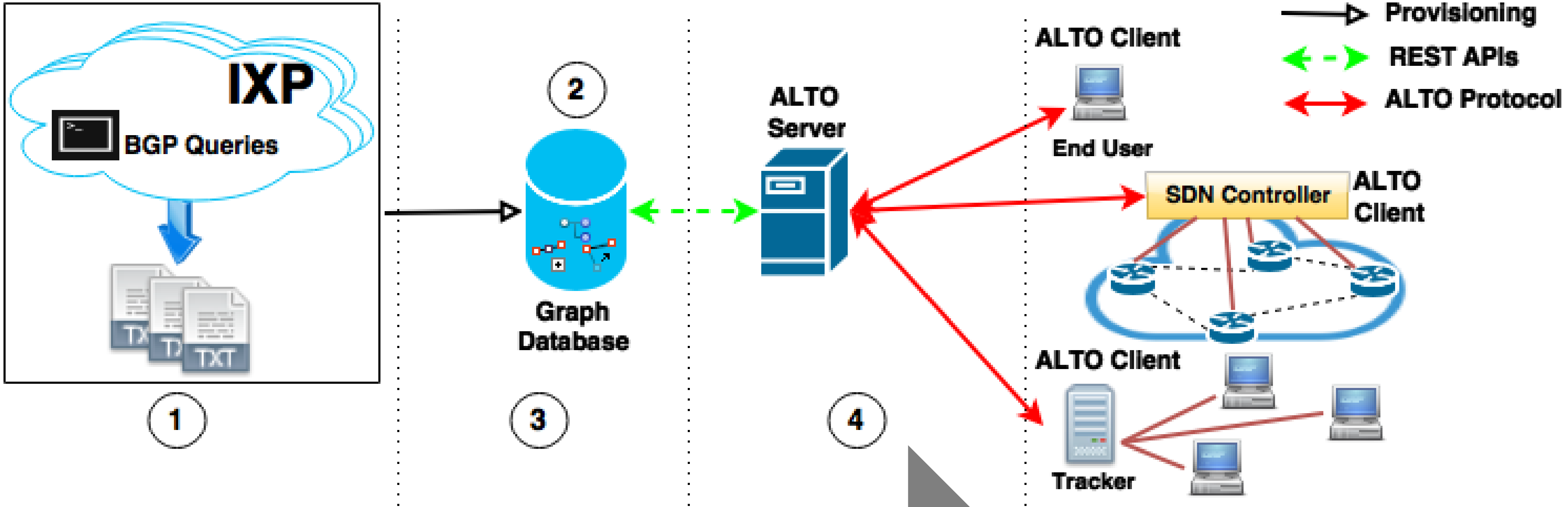

| | | | | |
|---|---|---|---|---|
| A shared network infrastructure | The Internet performance is optimized | Multilateral peering | Allow bilateral agreements | See IX.br public service approach (CGI) |

# Agenda

1. Introduction
2. Background
3. **Design of ALTO-as-a-Service (AaaS)**
4. AaaS Prototype
5. Experimental Evaluation
6. Conclusions / Future Work

## Design of ALTO-as-a-Service (AaaS) :

# AaaS Workflow

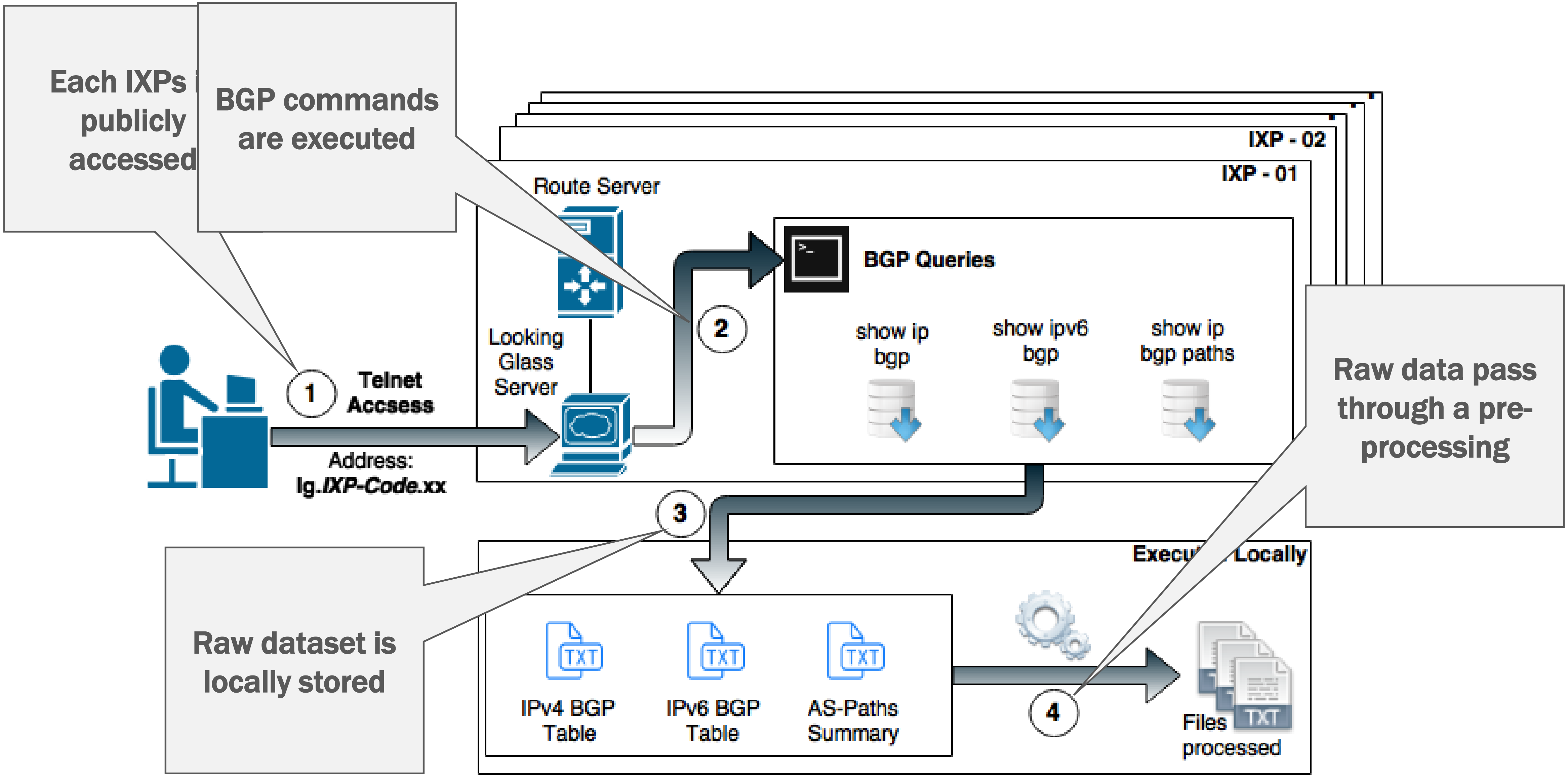



**1**

**Acquiring Input Data**

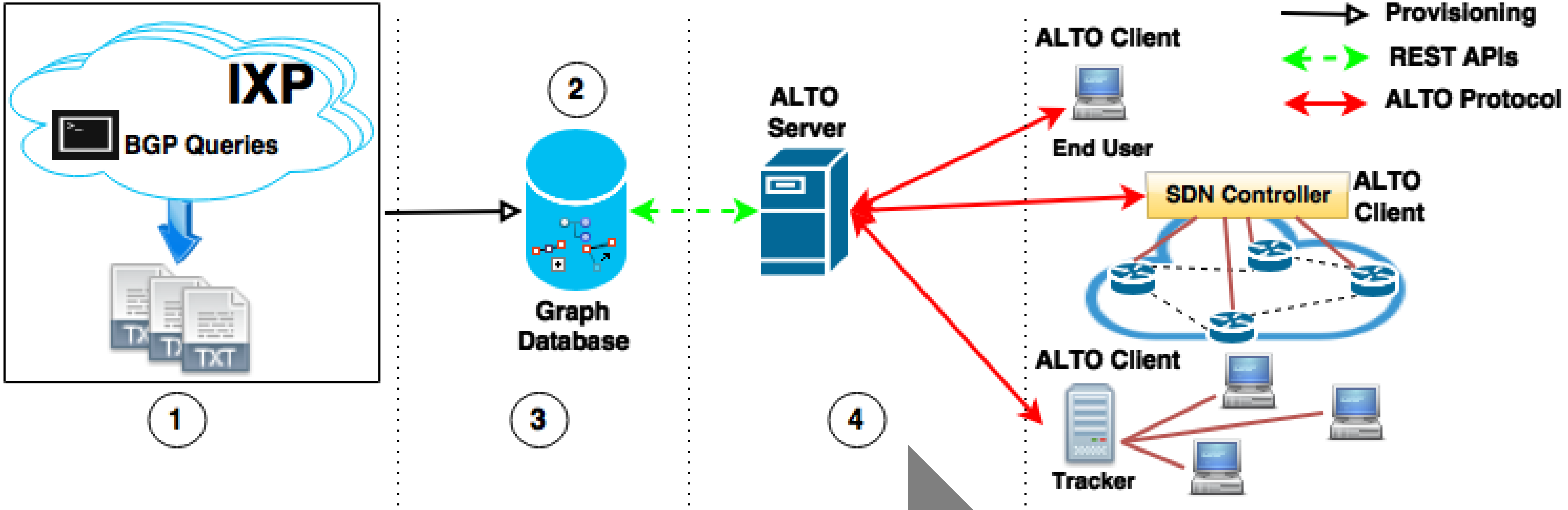
## Design of ALTO-as-a-Service (AaaS) :

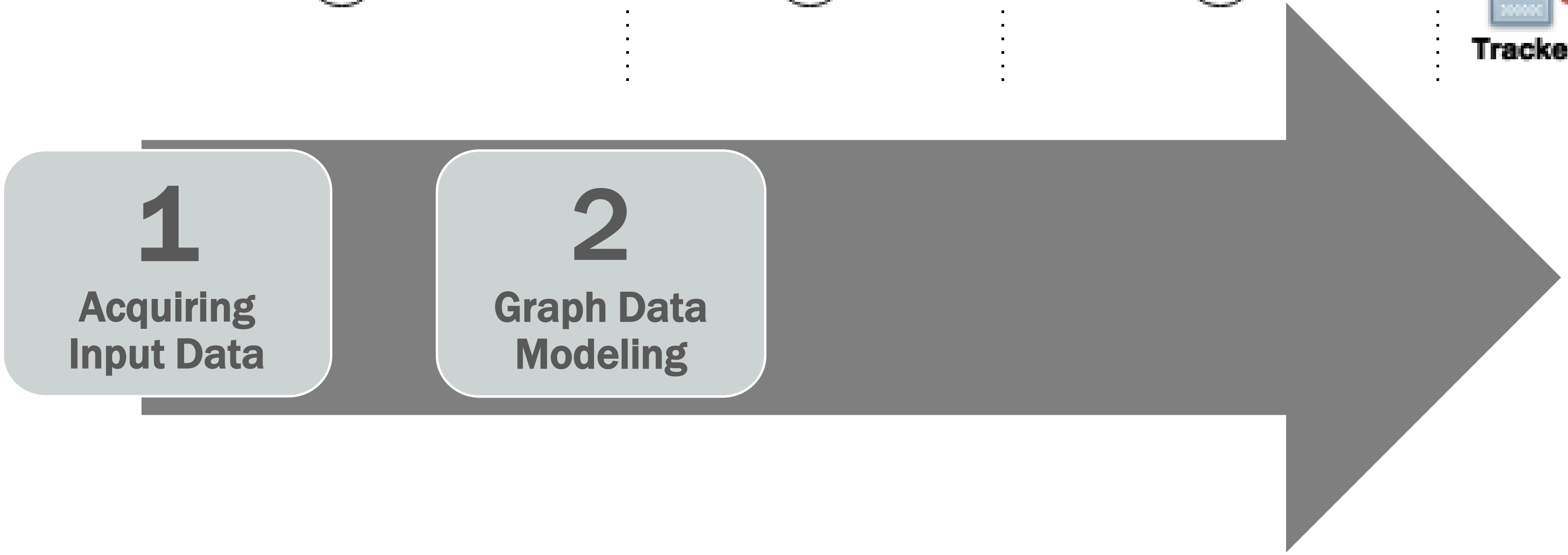
# Acquiring Input Data



## Design of ALTO-as-a-Service (AaaS) :

# AaaS Workflow



Design of ALTO-as-a-Service (AaaS) :

# Building Graph Data Models

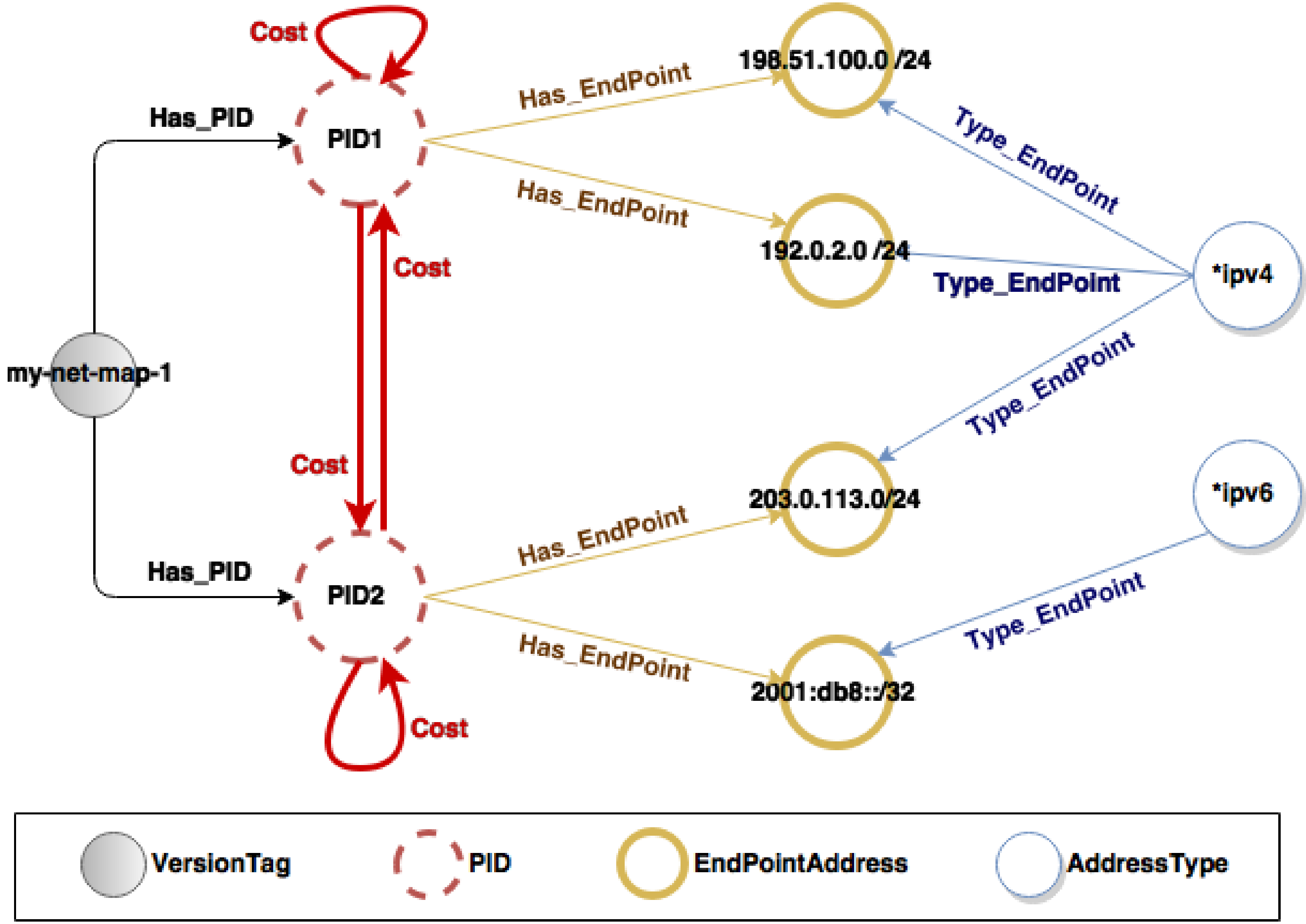
RFC 7285 IETF

## Nodes:

1. Provider-Defined Identifier
2. Endpoint Address
3. Address Type
4. Version Tag

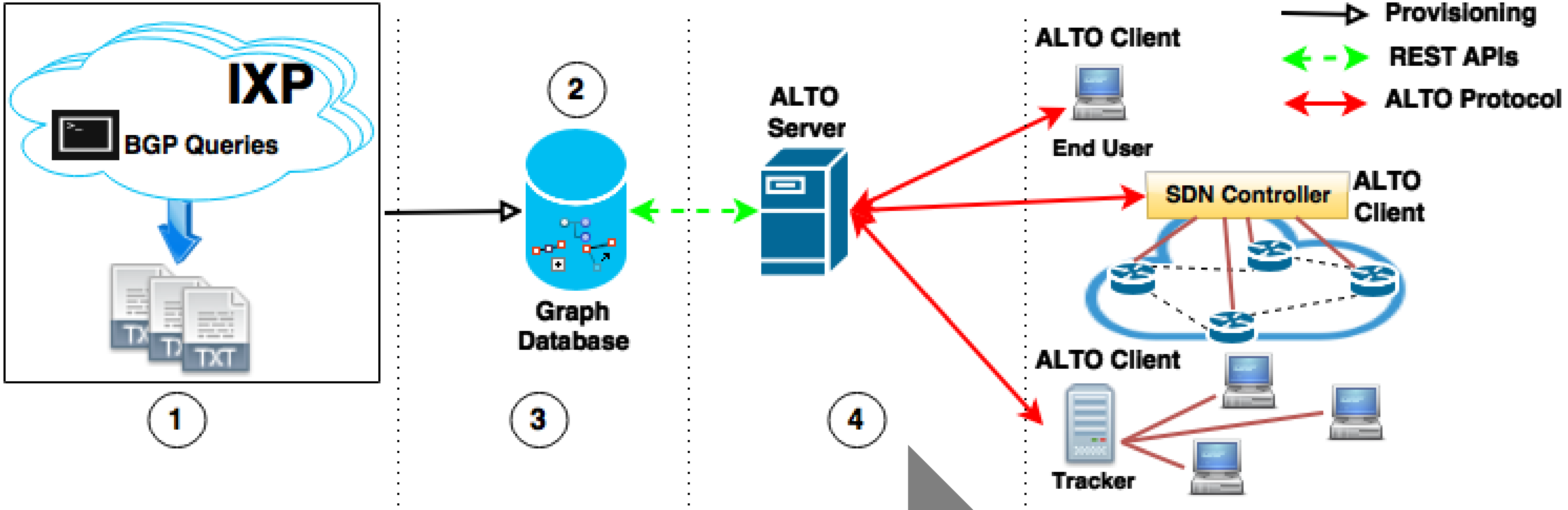
## Relationships:

1. Has_PID
2. Has_EndPoint
3. Type_EndPoint
4. Cost



## Design of ALTO-as-a-Service (AaaS) :

# AaaS Workflow




**1** Acquiring Input Data

**2** Graph Data Modeling

**3** Creating ALTO Information

**4** Delivering ALTO Services

# Agenda

1. Introduction
2. Background
3. Design of ALTO-as-a-Service (AaaS)
4. **AaaS Prototype**
5. Experimental Evaluation
6. Conclusions / Future Work

AaaS Prototype :

# Input: IX.br BGP Data Set

**25** IXPs are part of PTTMetro project[1], created to promote the necessary infrastructure for direct interconnection between networks of Brazilian metropolitan regions.

**1** The largest IXP ecosystem in Latin America and one of the world's top ten.

**1374** Members[2].

1184 ASes registered.

**+1.0** Tbps average throughput [3]

[1]http://ix.br

[2]http://ix.br/particip, Accessed: September, 2015

[3]http://ix.br/trafego, Accessed: May, 2015

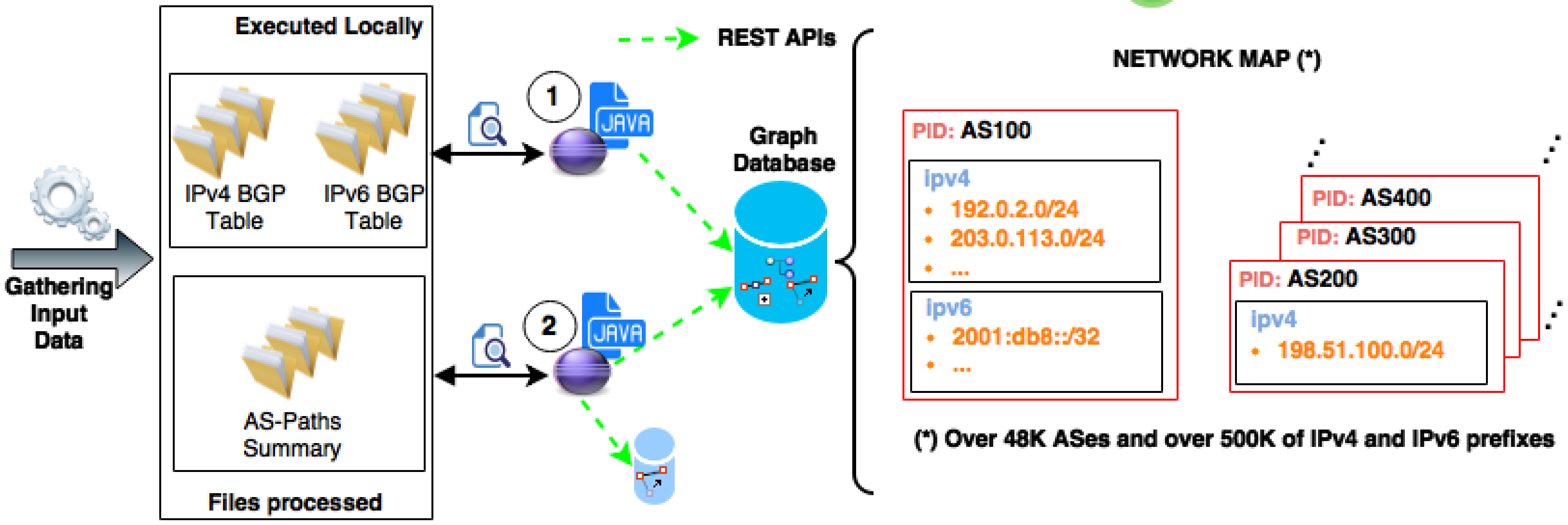
Table 4.1: Public IXPs Operating in Brazil

| # | City | State | Looking Glass | Gbps | Members | |
|---|---|---|---|---|---|---|
| 01 | Belem | PA | lg.bel.ptt.br | 0.44 | 14 | |
| 02 | Belo Horizonte | MG | lg.mg.ptt.br | 2.07 | 33 | |
| 03 | Brasilia | DF | lg.df.ptt.br | 2.98 | 30 | |
| 04 | Campina Grande | PB | lg.cpv.ptt.br | 0.69 | 10 | |
| 05 | Campinas | SP | lg.cas.ptt.br | 3.57 | 35 | (*) |
| 06 | Cuiaba | MT | lg.cgb.ptt.br | 0.00 | 9 | (*) |
| 07 | Caxias do Sul | RS | lg.cxj.ptt.br | 0.08 | 5 | (*) |
| 08 | Curitiba | PR | lg.pr.ptt.br | 16.10 | 68 | (1) |
| 19 | Florianopolis | SC | lg.sc.ptt.br | 1.28 | 34 | (*) |
| 10 | Fortaleza | CE | lg.ce.ptt.br | 2.72 | 29 | |
| 11 | Goiania | GO | lg.gyn.ptt.br | 1.06 | 24 | |
| 12 | Lajeado | RS | lg.laj.ptt.br | 0.01 | 8 | (*) |
| 13 | Londrina | PR | lg.lda.ptt.br | 1.62 | 32 | |
| 14 | Manaus | AM | lg.mao.ptt.br | 0.02 | 8 | (*) |
| 15 | Maringa | PR | lg.mgf.ptt.br | 0.28 | 21 | (*) |
| 16 | Natal | RN | lg.nat.ptt.br | 0.26 | 13 | (*) |
| 17 | Porto Alegre | RS | lg.rs.ptt.br | 20.85 | 117 | |
| 18 | Recife | PE | lg.pe.ptt.br | 0.69 | 16 | |
| 19 | Rio de Janeiro | RJ | lg.rj.ptt.br | 39.22 | 68 | |
| 20 | Salvador | BA | lg.ba.ptt.br | 1.47 | 47 | (*) |
| 21 | Sao Carlos | SP | lg.sca.ptt.br | 0.00 | 3 | (*) |
| 22 | Sao Jose dos Campos | SP | lg.sjc.ptt.br | 0.47 | 13 | |
| 23 | Sao Jose do Rio Preto | SP | lg.sjp.ptt.br | 0.62 | 11 | (*) |
| 24 | Sao Paulo | SP | lg.sp.ptt.br | 429.45 | 667 | (1) |
| 25 | Vitoria | ES | lg.vix.ptt.br | 0.80 | 22 | |

(1) There are filters in LG compromising the BGP table.

(*) Data provided by NIC.br, since publicly access was denied.

## AaaS Prototype :
# ALTO Server Backend: Neo4j




COST MAP (+)

| AS hops | AS100 | AS200 | AS300 | ... |
|---|---|---|---|---|
| AS100 | 0 | 5 | 4 | ... |
| AS200 | 2 | 0 | 6 | ... |
| AS300 | 3 | 3 | 0 | ... |
| ... | ... | ... | ... | ... |

(+) Over 2.3 billions of pah costs

### 1. Network Map

✓ The ASNs serve as our grouping, with each ASN corr

✓ Each prefix (be it IPv4 and IPv6) will be associate w

### 2. Cost Map

✓ Path costs = The topological distance, expressed in the n of traversed AS.

✓ Two Cost Maps: Absolute distance and Relative distance.

Hops between ASes present in the same IXPs are zeroed to favor intra-IXP traffic.

AaaS Prototype :

# ALTO Server Front-End: OpenDaylight


- ✓ ALTO in ODL is focused on implementing basic ALTO as services RESTful web services (Northbound APIs) for ALTO client/server communications.
- ✓ ALTO Northbound APIs generate ALTO services from data stored in the MD-SAL data store (an ODL core component)
  - It was necessary to modify the Northbound APIs to generate ALTO services from the data stored in the Neo4j GDB (instead of the MD-SAL topology).

# Agenda

1. Introduction
2. Background
3. Design of ALTO-as-a-Service (AaaS)
4. AaaS Prototype
5. **Experimental Evaluation**
6. Conclusions / Future Work

Experimental Evaluation :

# Functional Evaluation

RFC 7285 IETF

Postman - REST Client

{JSON}

```
URI : "http://intrig.dca.fee.unicamp.br/controller/nb/v2/alto/
        filtered/networkmap/my-default-network-map"
HTTP Method : "POST"
Content-Type : "application/alto-networkmapfilter+json"
Input Parameters :
{
        "pids" : ["AS100","AS200"]
}
```

```
HTTP Response :
{
   "meta": {
      "vtag": { "resource-id": "my-default-network-map",
       "tag": "da65eca2eb7a10ce" }
   },
   "network-map": {
     "AS100": { "ipv4": [ "192.0.2.0/24","203.0.113.0/24" ],
                         "ipv6": [ "2001:db8::/32" ],
     "AS200": { "ipv4": [ "198.51.100.0/24" ] }
      }
   }
}
```

(a) Filtered Network Map

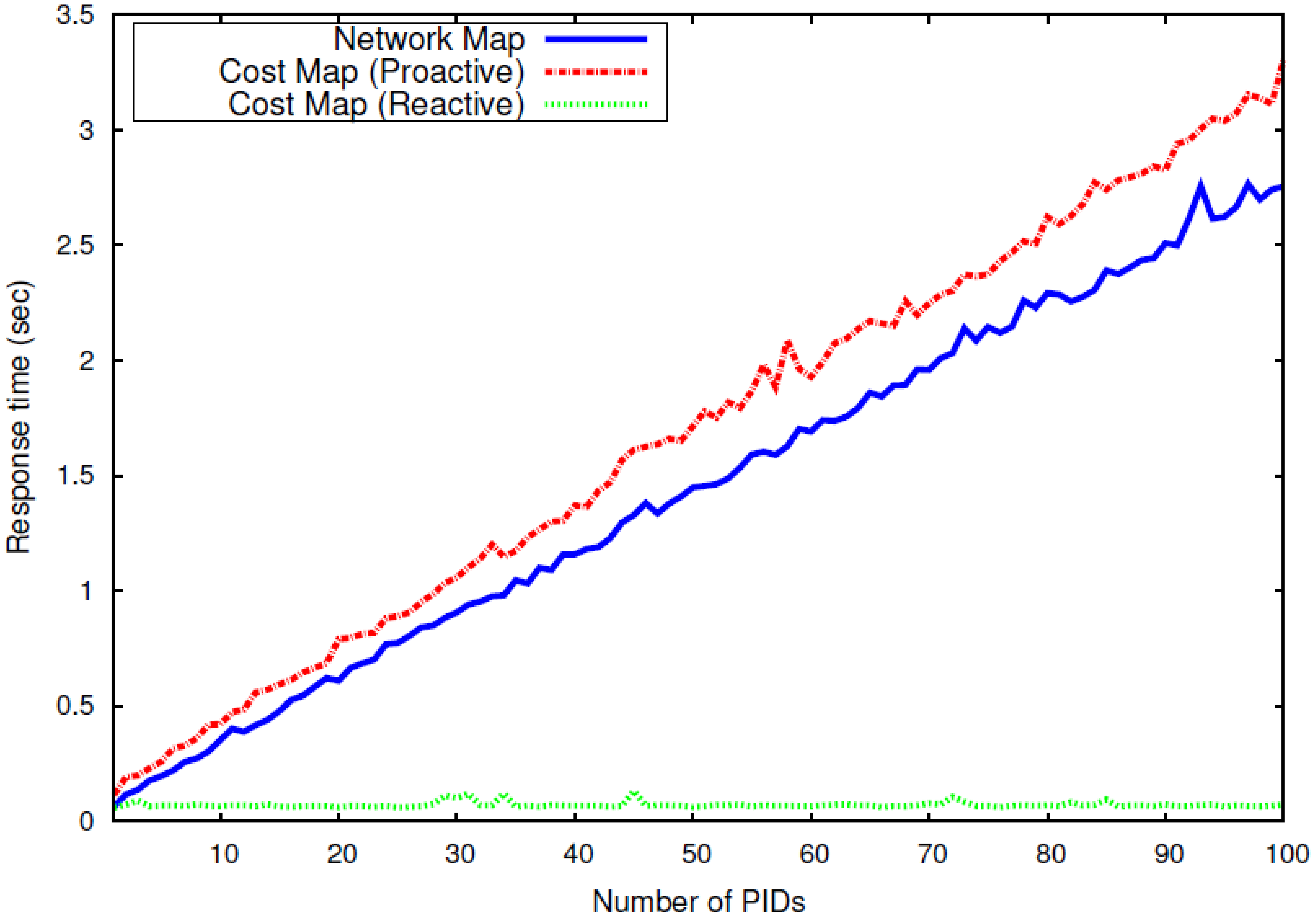
```
URI : "http://intrig.dca.fee.unicamp.br/controller/nb/v2/alto/
        filtered/costmap/my-default-network-map"
HTTP Method : "POST"
Content-Type : "application/alto-costmapfilter+json"
Input Parameters :
{   "cost-type" :{"cost-mode": "Numerical",
    "cost-metric": "HopsNumber" },
    "pids" : { "srcs" : [ "AS100" ],
    "dsts" : [ "AS100","AS200","AS300" ] }
}
```

```
HTTP Response :
{
  "meta": {
    "dependent-vtags": [
                  { "resource-id": "my-default-network-map",
                            "tag": "da65eca2eb7a10ce" } ],
    "cost-type": { "cost-mode": "Numerical",
                   "cost-metric": "HopsNumber" }},
    "cost-map": { "AS100":
                  { "AS100": 0, "AS300": 4, "AS200": 5 } }
}
```

(b) Filtered Cost Map (abs. distance)

## Experimental Evaluation :

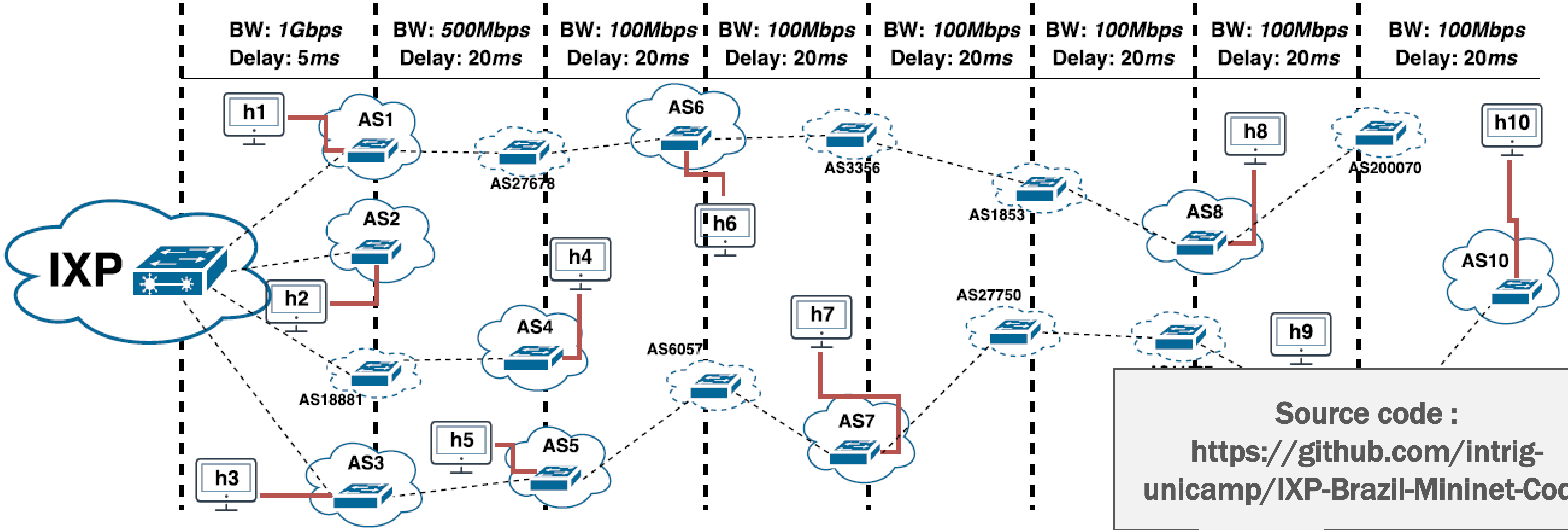
# System Performance Profiling




Response processing time for the Network and Cost Map (Absolute Distance) services

## Experimental Evaluation :

# Use Case Scenario (*strawman*)




Source code :
https://github.com/intrig-unicamp/IXP-Brazil-Mininet-Code

- ✓ 19 ASes, each represented by a switch abstraction in the Mininet emulator.
- ✓ A sample AS-Path summary file was used to create the AS-level connectivity.
- ✓ The large AS switch represents the IXP.
- ✓ 10 communicating peers are represented as Mininet hosts attached to the switches.
- ✓ Links between ASes were set with larger bandwidth and lower delay when closer to the IXP.

## Experimental Evaluation :

# Use Case Scenario

### ALTO information

✓ Two Cost Maps: based on the topological distance (absolute and relative) expressed as the number of hops between ASes.

**COST MAP RANKING 1: Absolute distance (HopsNumber)**

"cost-map": {
**AS1** {**AS2: 2**, AS3: 2, AS6: 2, AS4: 3, AS5: 3, AS7: 5, AS8: 5, AS9: 8, AS10: 9 },
**AS2** {**AS1: 2**, AS3: 2, AS4: 3, AS5: 3, AS6: 4, AS7: 5, AS8: 7, AS9: 8, AS10: 9 },
**AS3** {**AS5: 1**, AS1: 2, AS2: 2, AS4: 3, AS7: 3, AS6: 4, AS9: 6, AS10: 7, AS8: 7 }}

**COST MAP RANKING 2: Relative distance (HopsNumberPTT)**

"cost-map": {
**AS1** {**AS2: 0**, AS3: 0, AS4: 1, AS5: 1, AS6: 2, AS7: 3, AS8: 5, AS9: 6, AS10: 7 },
**AS2** {**AS1: 0**, AS3: 0, AS4: 1, AS5: 1, AS6: 2, AS7: 3, AS8: 5, AS9: 6, AS10: 7 },
**AS3** {**AS1: 0**, AS2: 0, AS4: 1, AS5: 1, AS6: 2, AS7: 3, AS8: 5, AS9: 6, AS10: 7 }}

### ALTO Clients

✓ Hosts that belong to ASes present at the IXP (ie., h1, h2, h3).

### Workload and Metrics

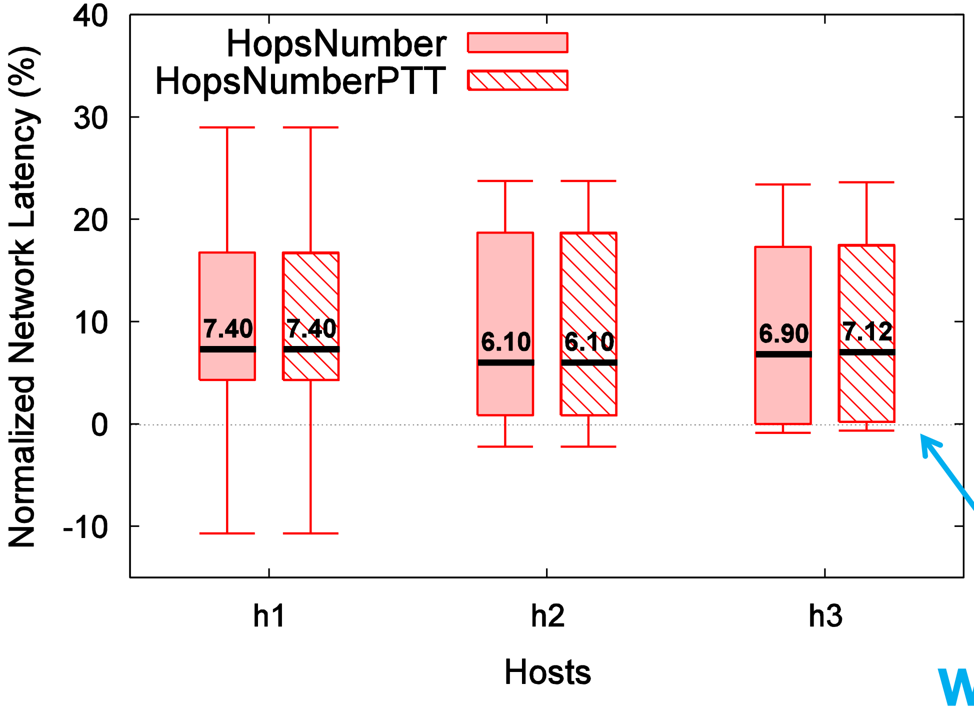
✓ For each ALTO client (h1, h2, h3), we run end-to-end round-trip time measurements and available bandwidth with the remaining nine hosts using ping and iperf tools, respectively.

✓ Two scenarios:

1. Without traffic
2. With a background traffic using the D-ITG traffic sending TCP traffic (512 byte packet size, 1,000 pps rate).

## Experimental Evaluation :

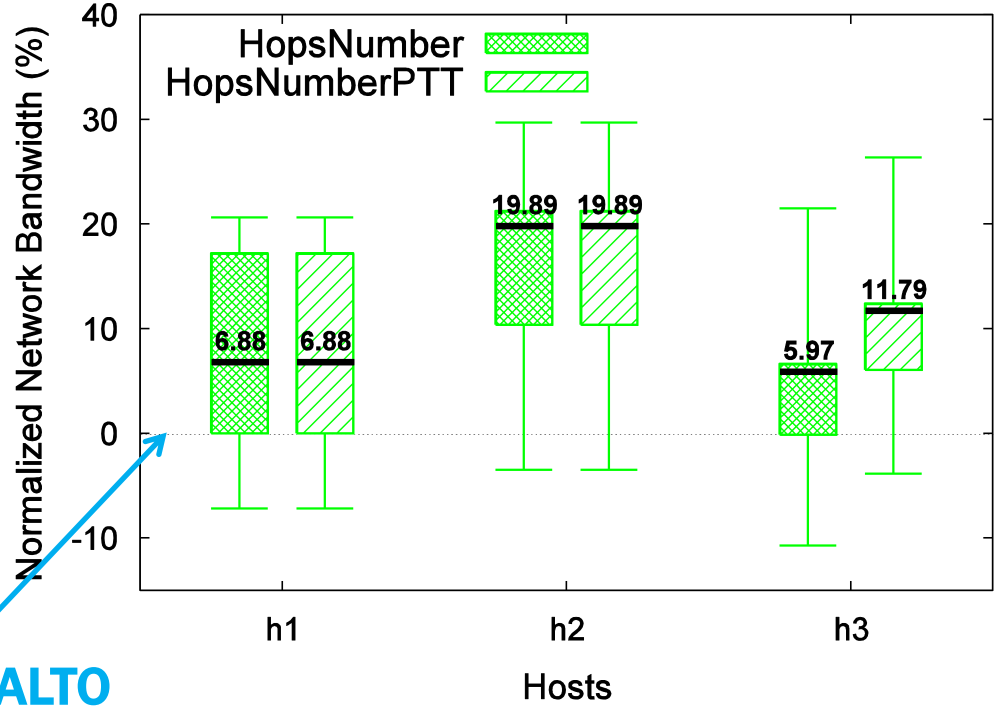
# Use Case Scenario




Latency

Without ALTO

(Random selection)




Bandwidth

**Results Analysis**

- ✓ improvements in latency and throughput of up to 29%
- ✓ In a few cases, peers selected through AaaS ended with less bandwidth or higher latency (between 1% and 11%) : Shortest paths not necessary the best performing ones.
- ✓ When h3 uses the IXP infrastructure to select a peer (h1), as HopsNumberPTT suggests, further throughput improvements (up to 26%) and lower latency (up to 24%) can be obtained.

# Agenda

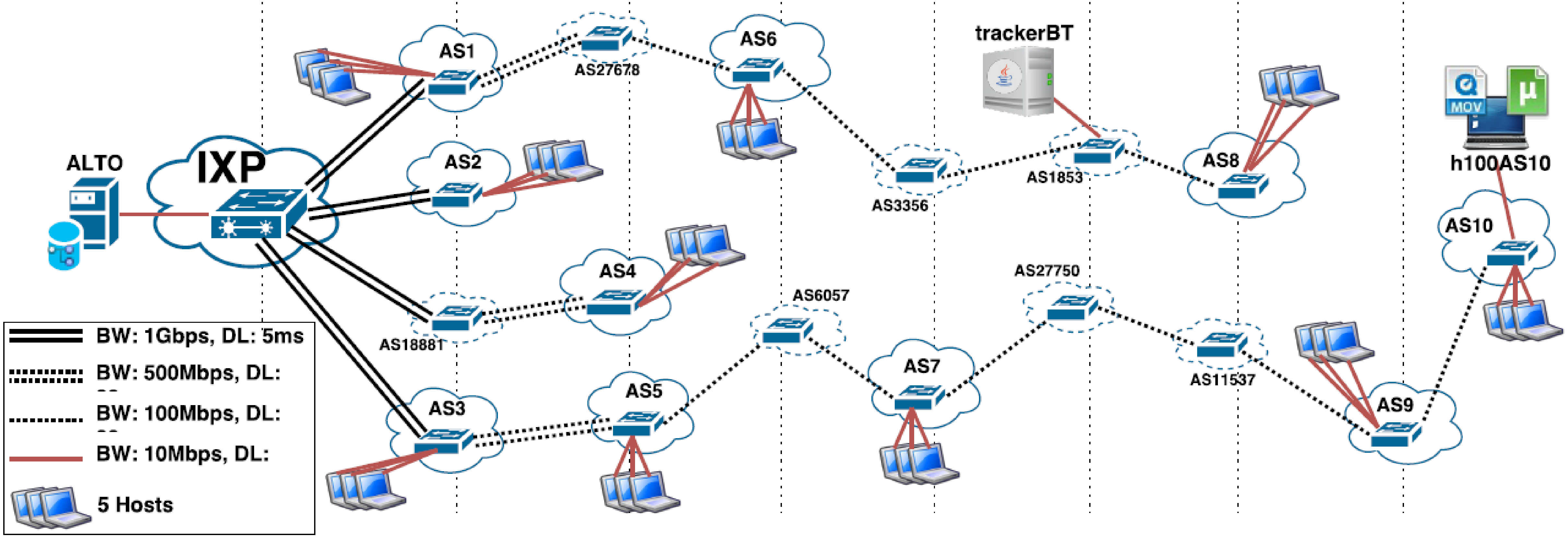
1. Introduction
2. Background
3. Design of ALTO-as-a-Service (AaaS)
4. AaaS Prototype
5. Experimental Evaluation
6. **Conclusions / Future Work**

# Future work

- ✓ The Cost Map rankings are based on relatively static AS-Path distance and do not consider more dynamic information such as actual bandwidth, latency, packet loss rate, etc.
  - Thus, dynamic updates of cost maps based on public Internet quality measurements (e.g., SIMET, RIPE TTM) are in our research agenda.
- ✓ Implement the remaining ALTO services (e.g., full Map-Service, EPS and ECS).
- ✓ Prototype performance optimization (e.g., Neo4j query tuning techniques, Linux fs configuration).

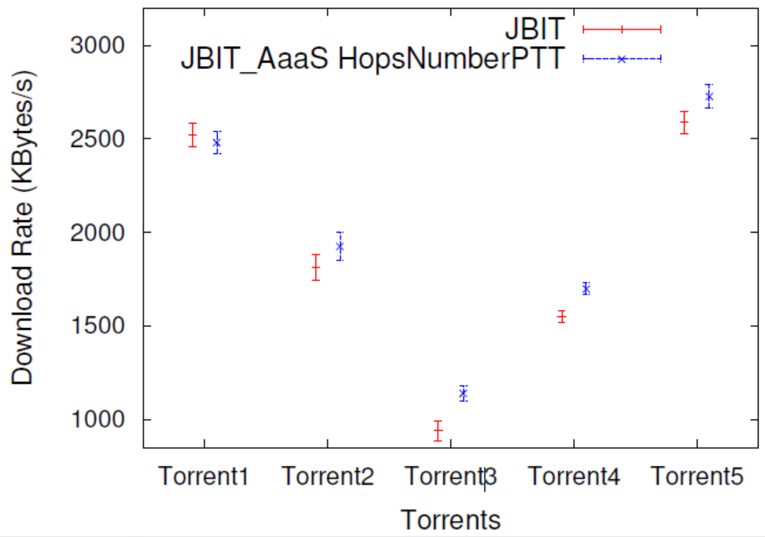
# Future work

✓ Further experimentation on more complex scenarios.



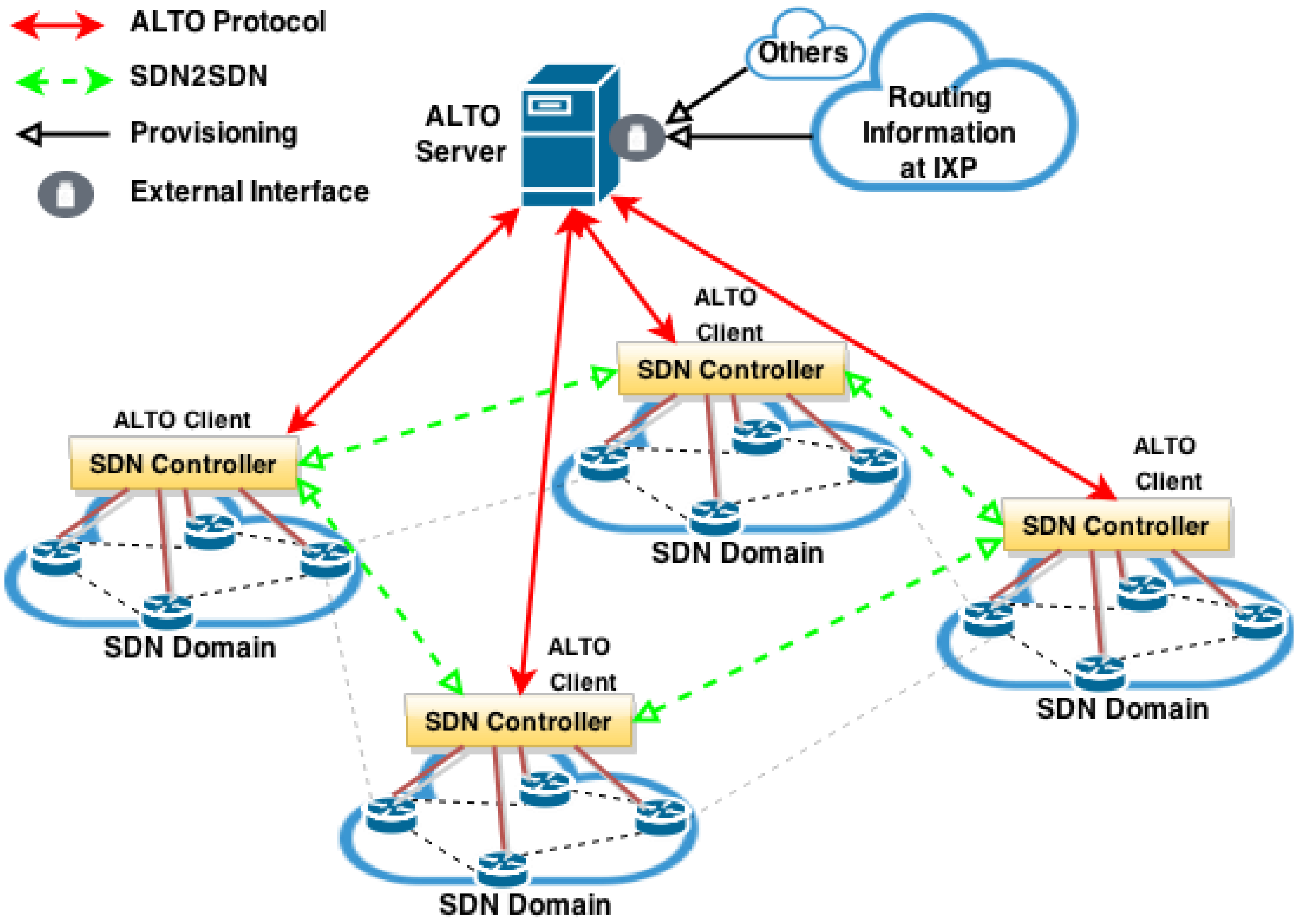

The IXP-based testing network model (**EXTENDED VERSION**)

# Ongoing work





## Real P2P scenarios




## ALTO – SDN Use Cases

# Conclusions

- ✓ This is the first work that explores the use of inter-domain routing data publicly available at IXPs to create abstract topology and cost maps following the recently standardized IETF ALTO protocol.
- ✓ Our proposed architecture encompasses the whole process of ALTO service delivery.
- ✓ Our proof-of-concept implementation is based on the popular Neo4j graph database and the OpenDaylight controller and validated the potential of applications to leverage the network awareness provided by ALTO servers.
- ✓ At the same time, ISPs and ALTO service providers (in our case IX.br operators) can benefit from increased and localized IXP traffic exchange.

# Thanks! Obrigado! (More) Questions?

dlachosp@dca.fee.unicamp.br

http://lattes.cnpq.br/5466177320244302

INTRIG
INFORMATION & NETWORKING TECHNOLOGIES RESEARCH & INNOVATION GROUP

FEEC
Faculdade de Engenharia Elétrica e de Computação
UNICAMP

# BACKUP

# Objectives

**MAIN OBJECTIVE:** Create and provide Application-Layer Traffic Optimization (ALTO) services based on public information, more specifically BGP routing information publicly available at Internet eXchange Points (IXPs).

| | |
|---|---|
| Issue: Random selection of available peers | **Providing a mechanism for giving peers information which enables choosing closest neighbors rather randomly.** |
| Issue: Cannot directly use traditional traffic optimization techniques | **Applications could have a large impact on the overall traffic generated using services with knowledge of network topology.** |
| Issue: Limited knowledge of the underlying network topology | **Working at a layer above, with a better understanding of the underlying network would give applications a detailed guide to achieve the desired resources.** |

Design of ALTO-as-a-Service (AaaS) :

# Creating and Delivering ALTO Information

## Create the ALTO information and populate the graph DB

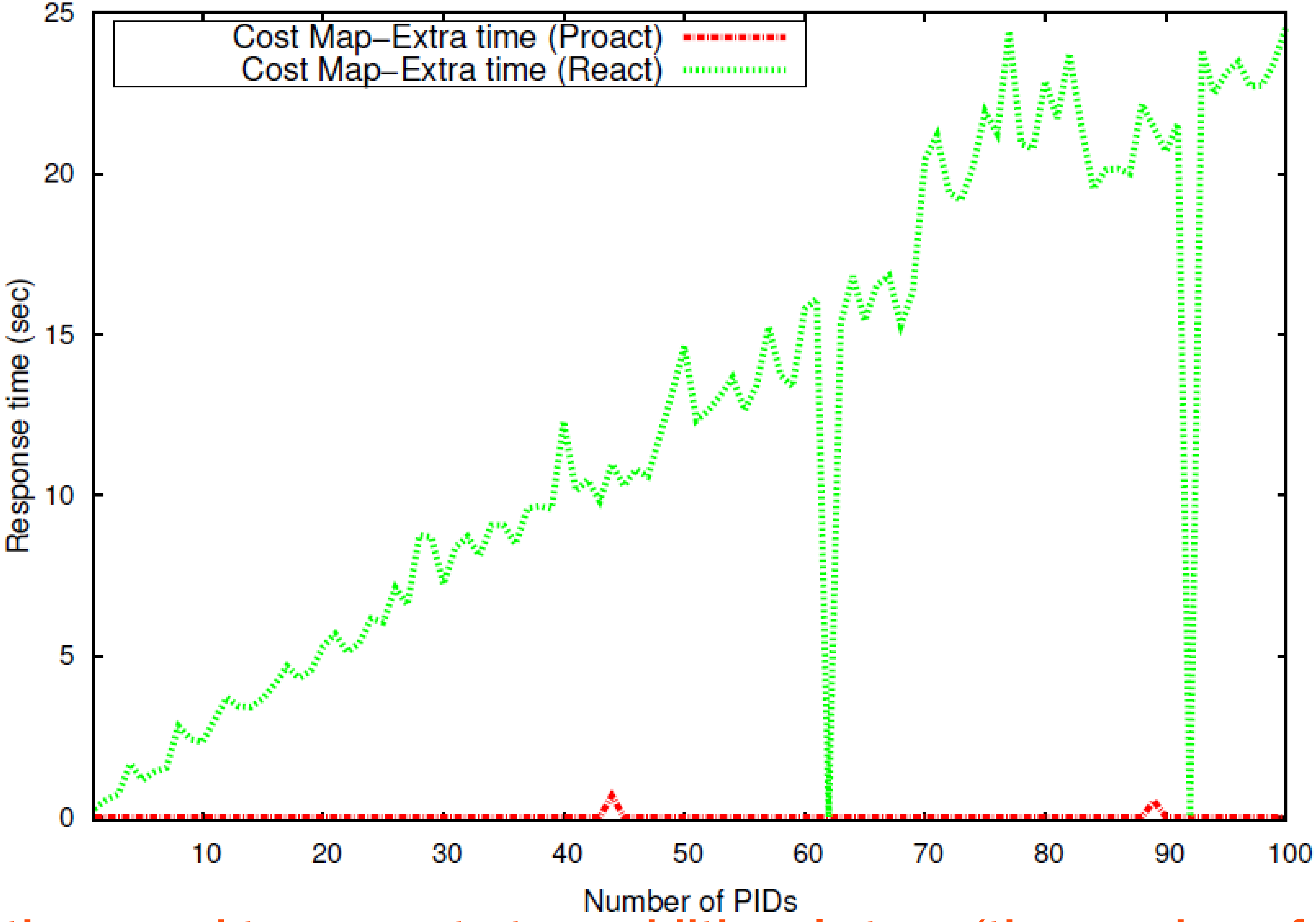
1. Grouping of prefixes into PID (Network Map) by ASes, IXPs, BGP communities, points of presence, just to cite a few examples.
2. Defining the preferences / costs between the groups PID (Cost Map) expressed on a path cost such as physical distance between IXPs, topological distances between ASes, etc.

## Delivering ALTO Services

The last step is deploying an ALTO web server implementing the client-server protocol delivering the REST/JSON APIs to ALTO clients as defined by RFC7285.

Experimental Evaluation :

# System Performance Profiling




Processing time used to compute two additional steps (the number of hops and insert it into database).

Experimental Evaluation :

# Use Case Scenario

| | h1 | h2 | h3 | h4 | h5 | h6 | h7 | h8 | h9 | h10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **h1** | — | 120.5 | 120.6 | 160.6 | 160.6 | 180.5 | 240.8 | 300.7 | 360.9 | 401.0 |
| | — | (±0.4) | (±0.3) | (±0.5) | (±0.5) | (±0.7) | (±0.5) | (±0.8) | (±0.8) | (±1.0) |
| **h2** | 120.5 | — | 120.4 | 160.6 | 160.6 | 200.7 | 240.9 | 320.9 | 360.9 | 401.1 |
| | (±0.3) | — | (±0.2) | (±0.4) | (±0.4) | (±0.3) | (±0.6) | (±0.7) | (±0.8) | (±0.6) |
| **h3** | 120.5 | 120.5 | — | 160.5 | 140.5 | 200.6 | 220.6 | 320.8 | 340.8 | 380.8 |
| | (±0.4) | (±0.5) | — | (±0.6) | (±0.4) | (±0.5) | (±0.4) | (±0.5) | (±0.4) | (±0.8) |

Processing Network Latency (ms) in a scenario with no traffic expressed as RTT AVG and RTT MDEV.