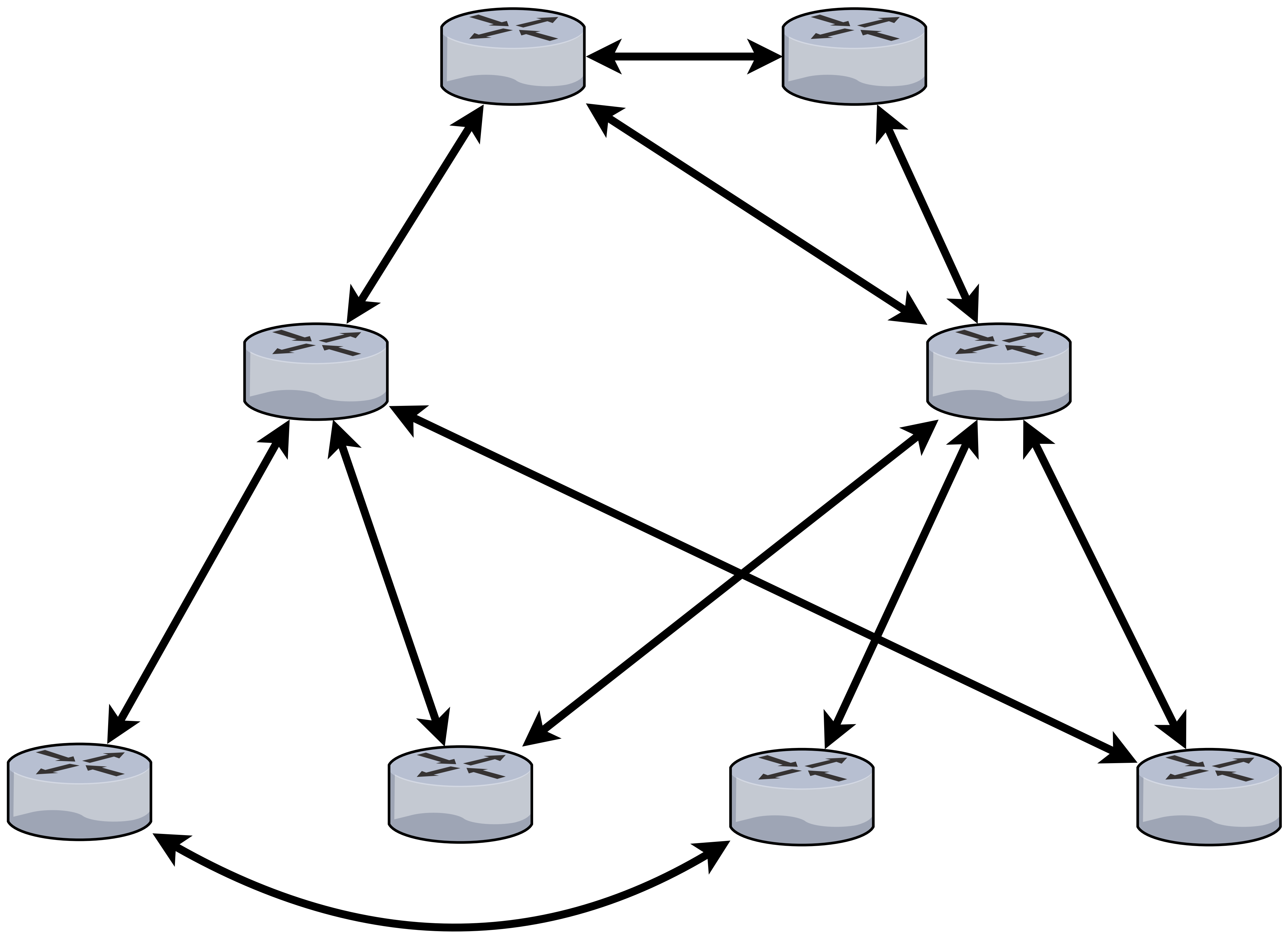
Tarpan: a router that supports evolvability

Harshal Sheth, Andrew Sun

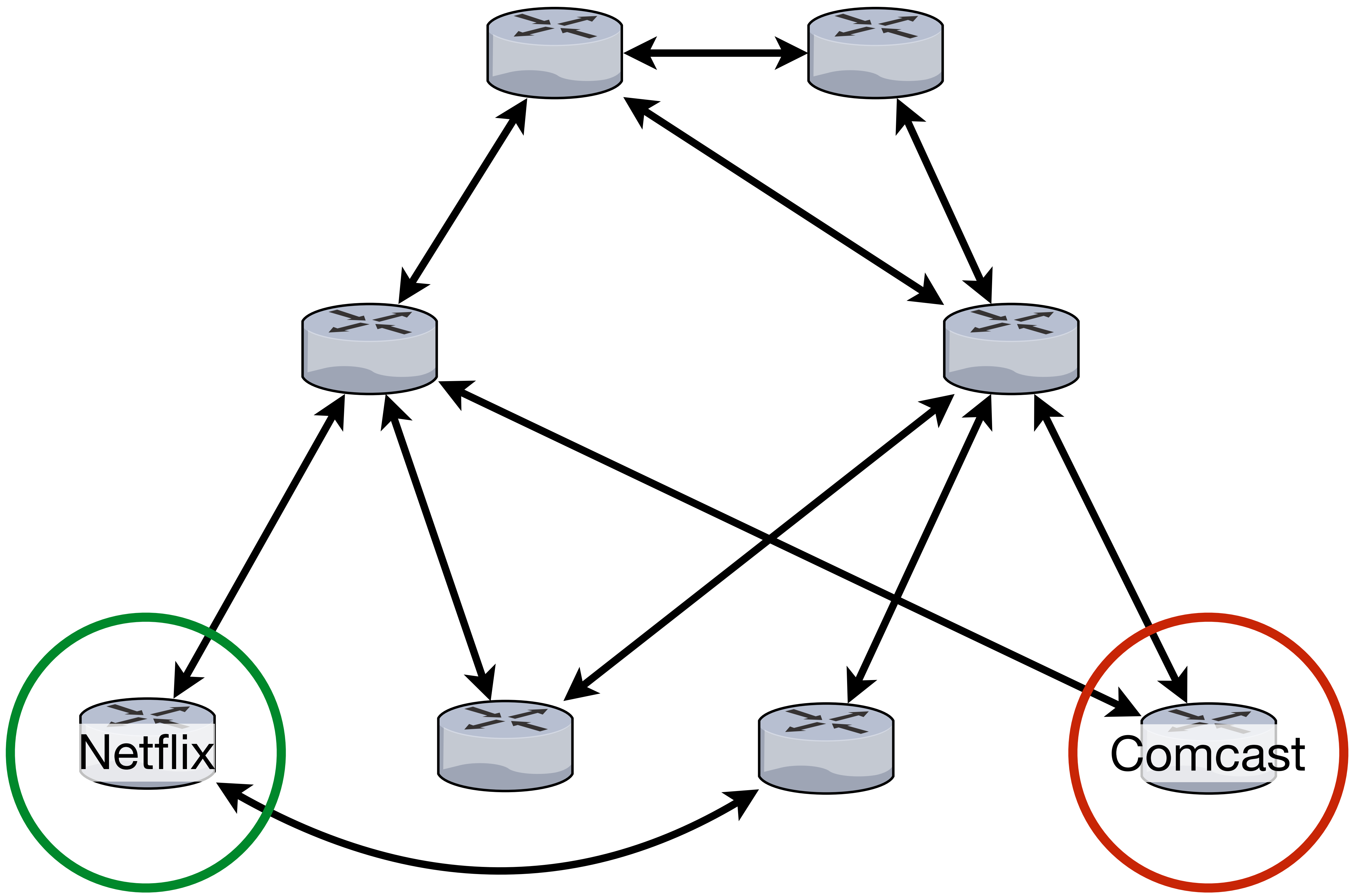
Inter-Domain Routing

- Creates paths between destination ISPs and source ones
  - e.g. Netflix to Comcast
- Paths can be used to deliver traffic (from sources to destinations)
- Provided today by BGP



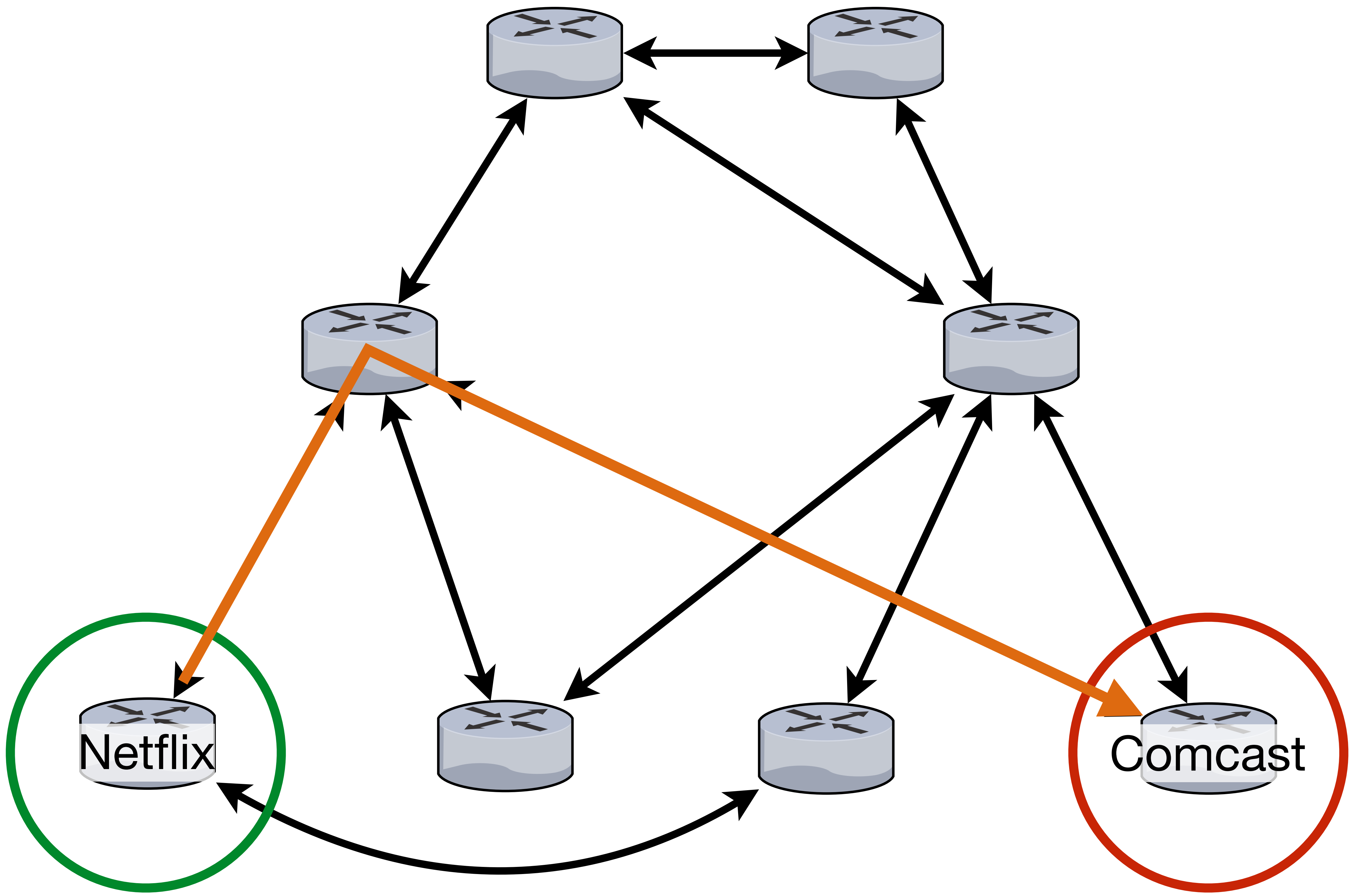
Inter-Domain Routing

- Creates paths between destination ISPs and source ones
  - e.g. Netflix to Comcast
- Paths can be used to deliver traffic (from sources to destinations)
- Provided today by BGP



Inter-Domain Routing

- Creates paths between destination ISPs and source ones
  - e.g. Netflix to Comcast
- Paths can be used to deliver traffic (from sources to destinations)
- Provided today by BGP



Problems with BGP and Proposed Solutions

Problems with BGP and Proposed Solutions

Problem	Proposed Solutions
Only one possible route	Pathlets: gives greater freedom in route selection by presenting multiple path fragments

Problems with BGP and Proposed Solutions

Problem	Proposed Solutions
Only one possible route	Pathlets: gives greater freedom in route selection by presenting multiple path fragments
Does not account for cost or link quality	Wiser: adds a *cost* value that could represent either cost or link quality

Problems with BGP and Proposed Solutions

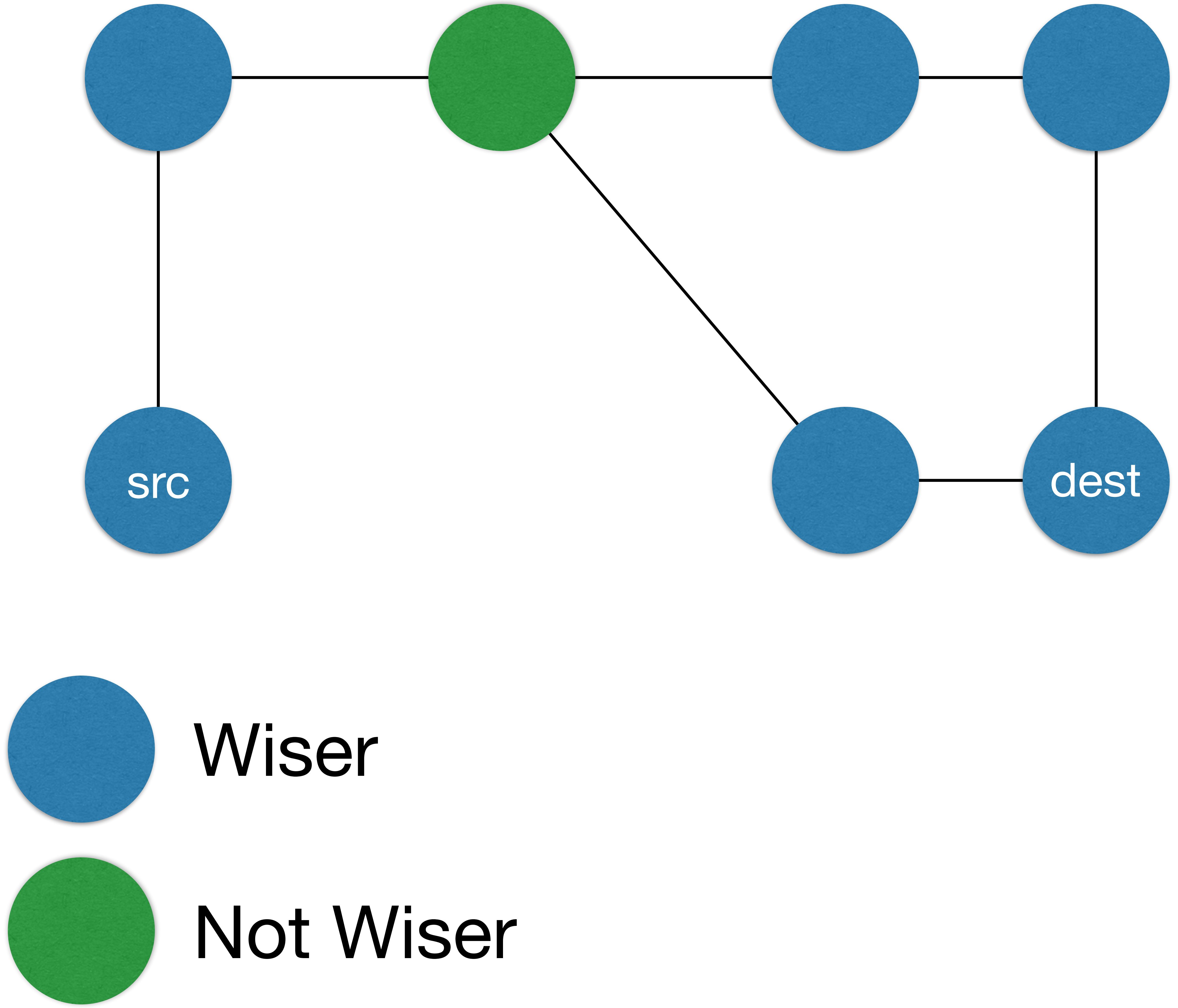
Problem	Proposed Solutions
Only one possible route	Pathlets: gives greater freedom in route selection by presenting multiple path fragments
Does not account for cost or link quality	Wiser: adds a *cost* value that could represent either cost or link quality
Trivial to hijack routes	S-BGP: adds authentication to BGP BGPsec: digital signatures in advertisements

Problems with BGP and Proposed Solutions

Problem	Proposed Solutions
Only one possible route	Pathlets: gives greater freedom in route selection by presenting multiple path fragments
Does not account for cost or link quality	Wiser: adds a *cost* value that could represent either cost or link quality
Trivial to hijack routes	S-BGP: adds authentication to BGP BGPsec: digital signatures in advertisements
BGP is rigid, assumes that all other routers are using BGP	Evolvable routing protocols

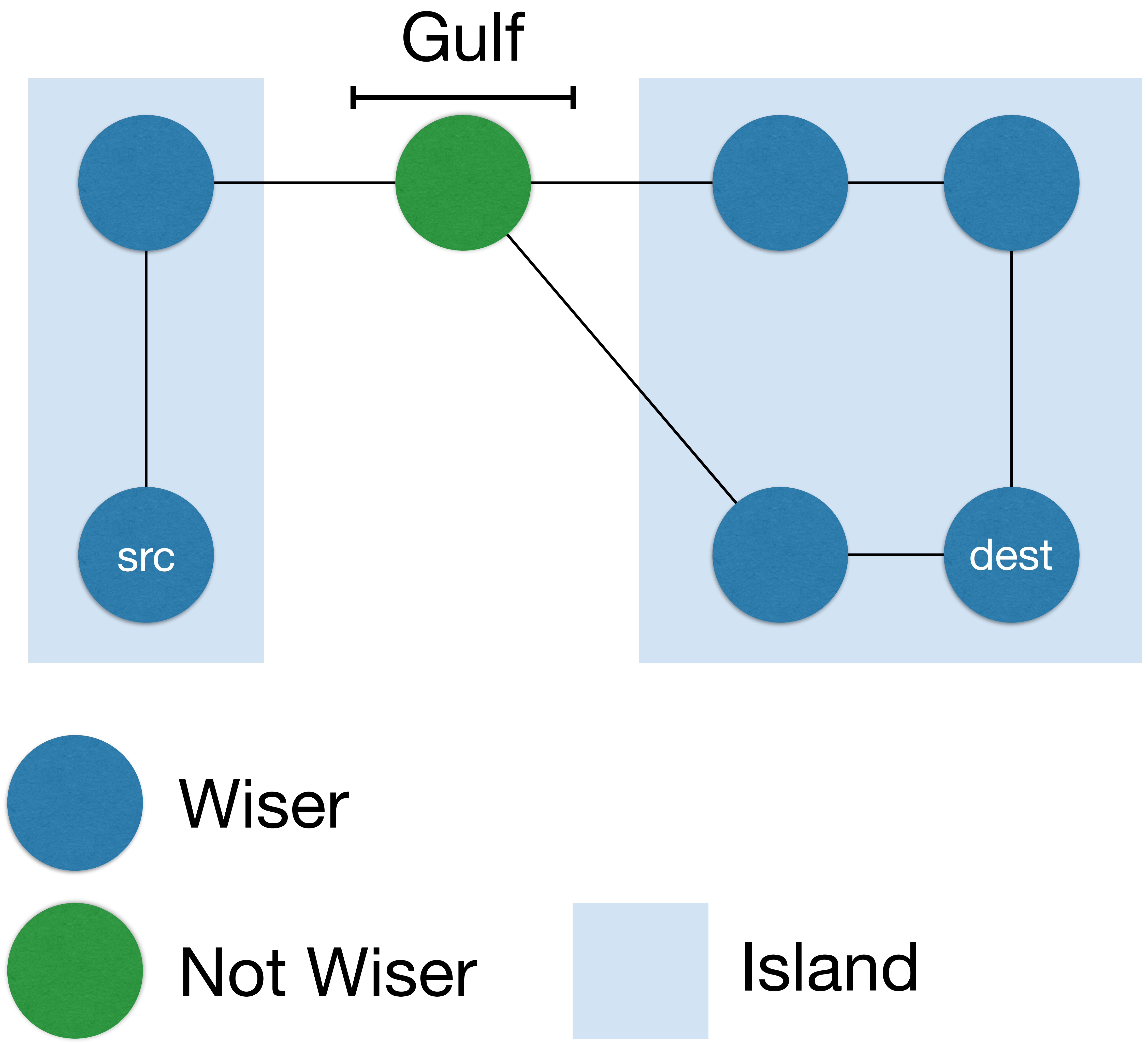
Example: Deploying Wiser

- Path costs
  - represents cost of sending on link and/or link saturation
- Cost normalization
  - prevents ISPs from inflating their path costs
- Bidirectional protocol



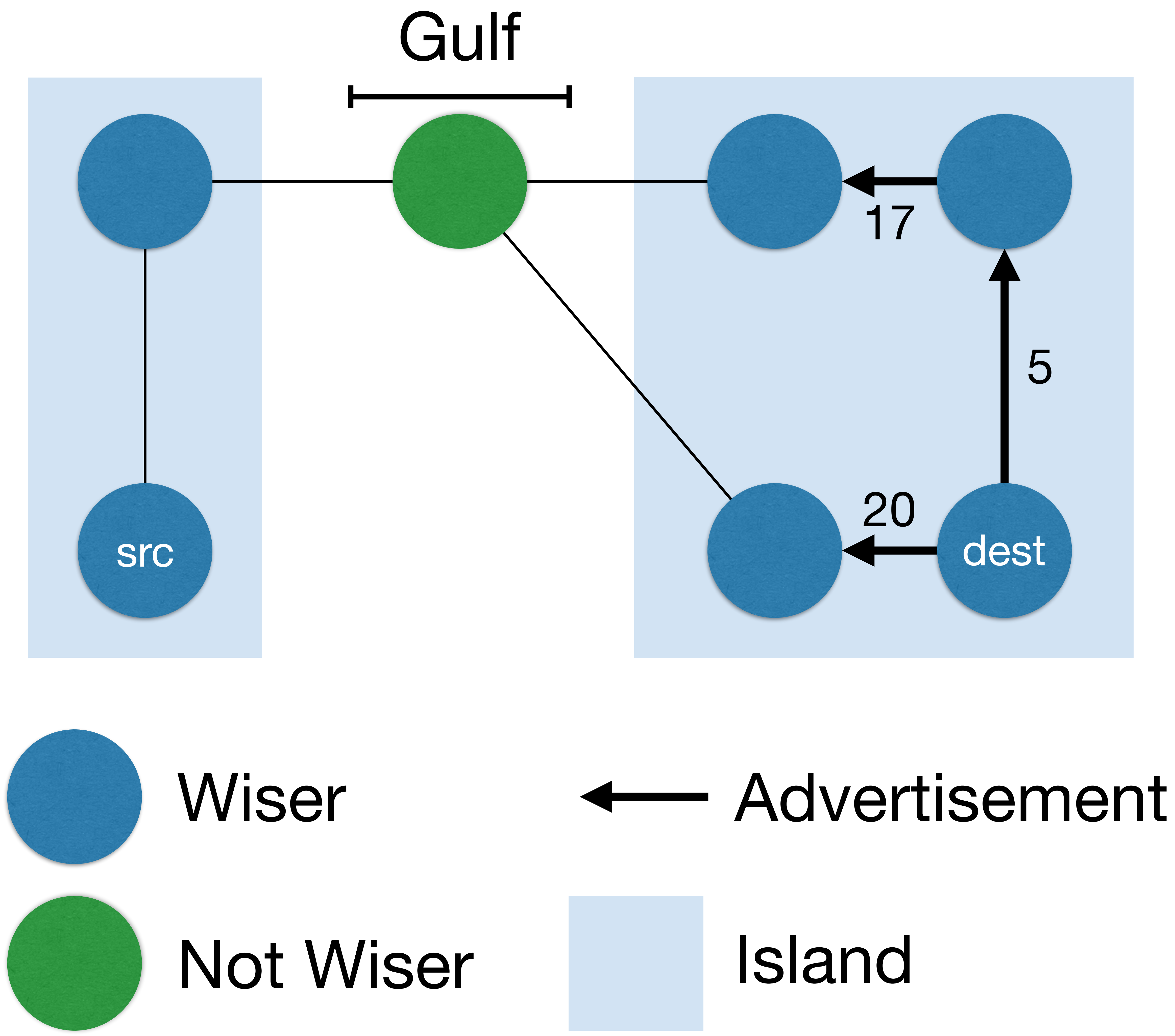
Example: Deploying Wiser

- Path costs
  - represents cost of sending on link and/or link saturation
- Cost normalization
  - prevents ISPs from inflating their path costs
- Bidirectional protocol



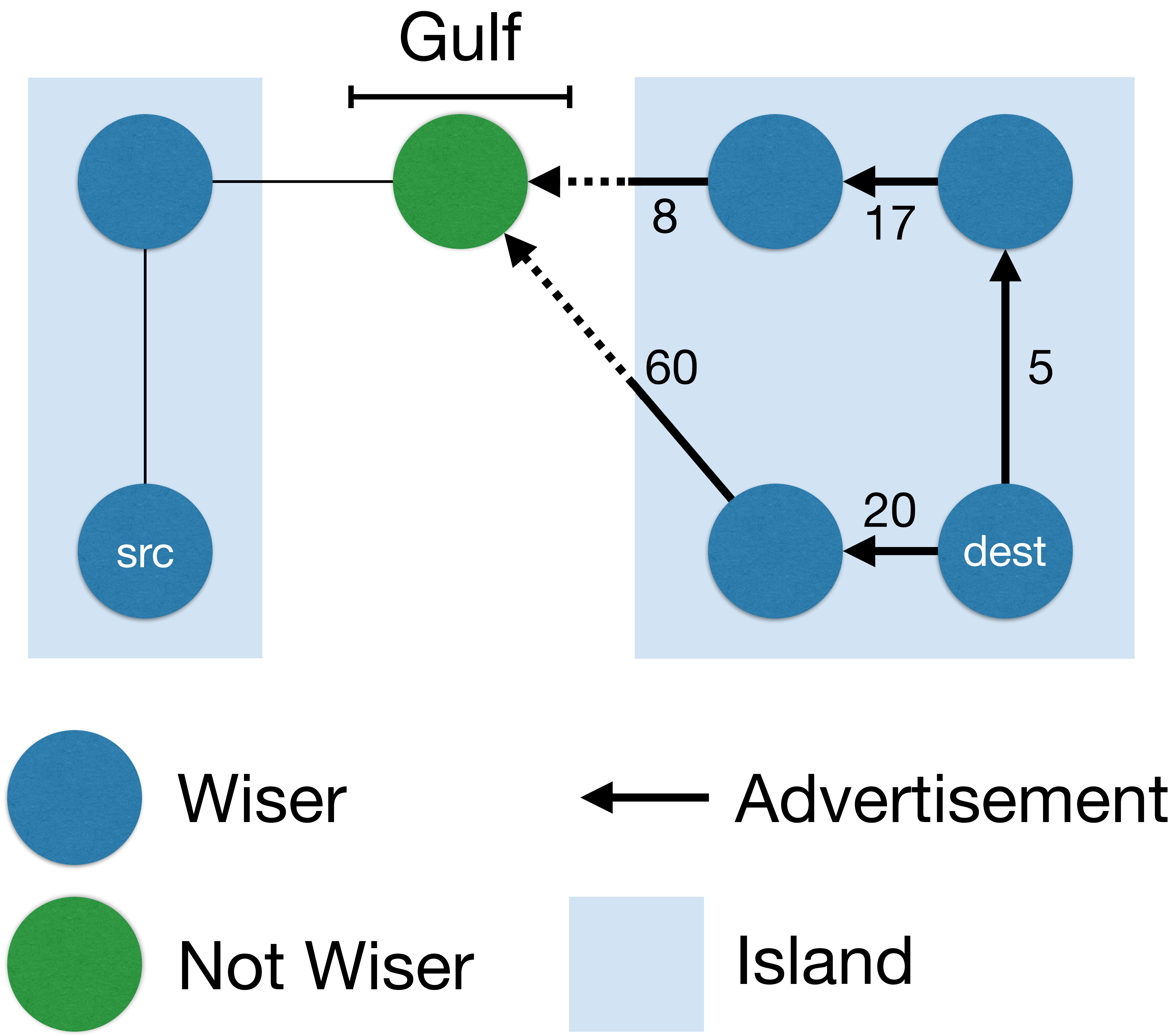
Example: Deploying Wiser

- Path costs
  - represents cost of sending on link and/or link saturation
- Cost normalization
  - prevents ISPs from inflating their path costs
- Bidirectional protocol



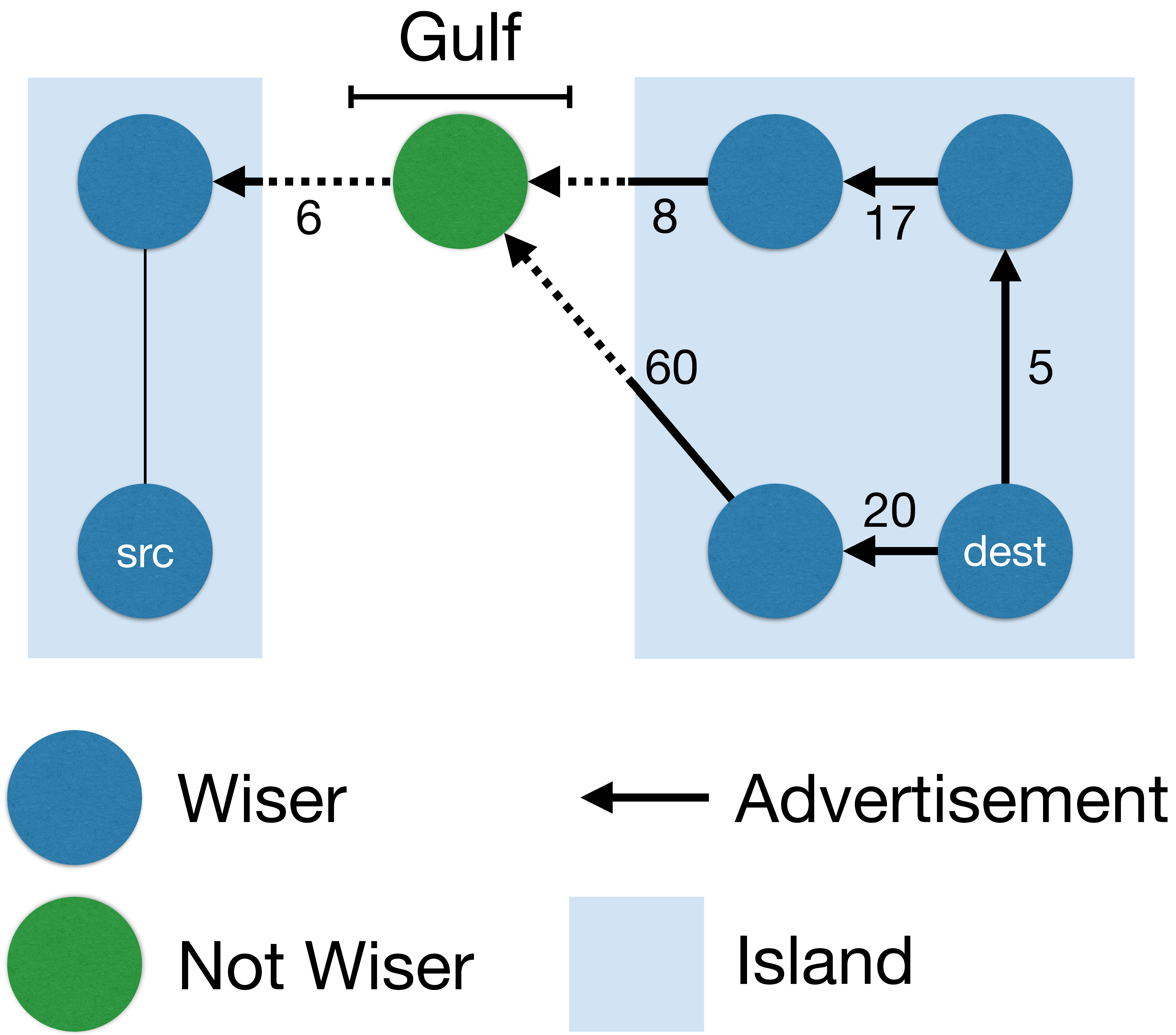
Example: Deploying Wiser

- Path costs
  - represents cost of sending on link and/or link saturation
- Cost normalization
  - prevents ISPs from inflating their path costs
- Bidirectional protocol



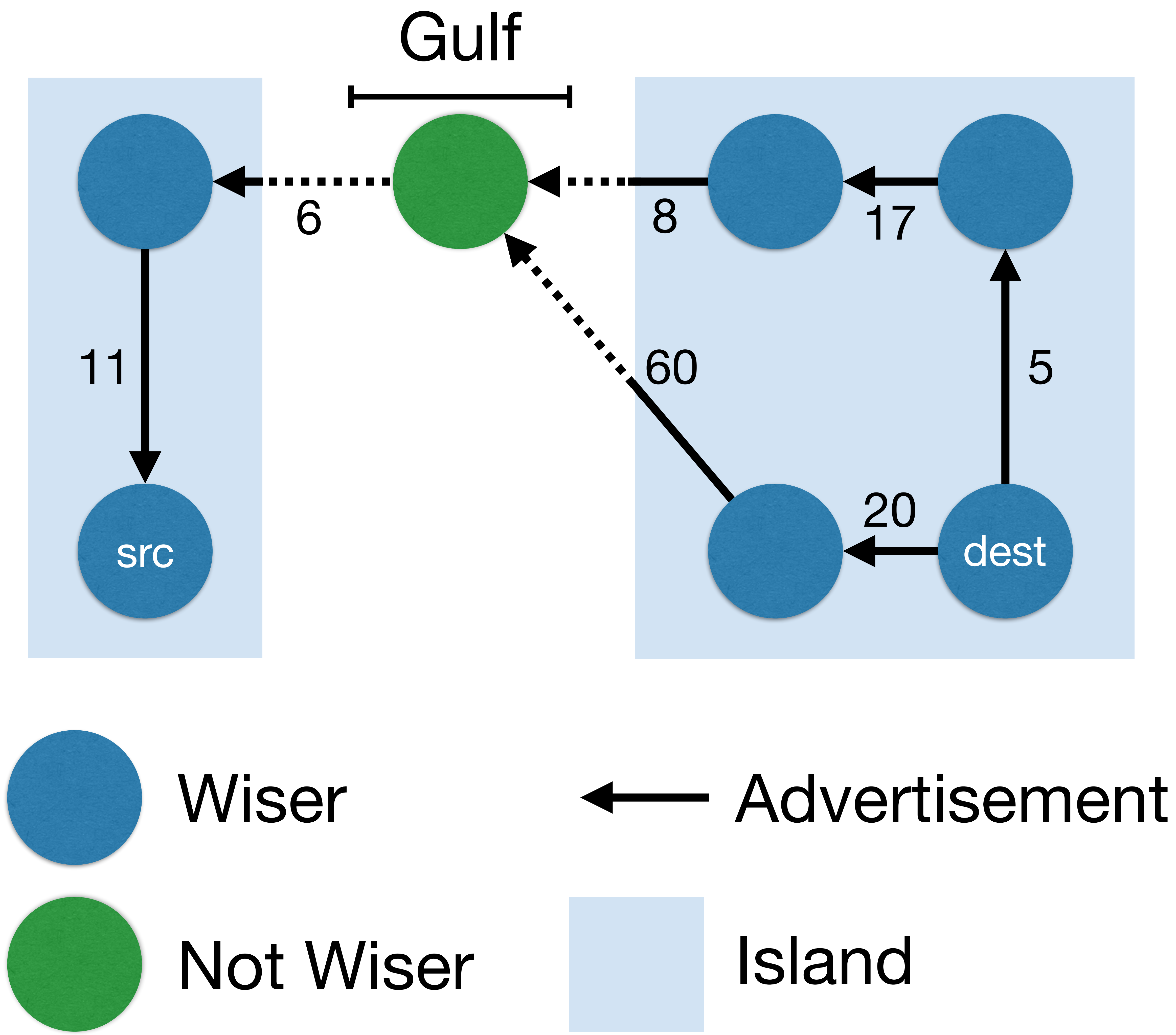
Example: Deploying Wiser

- Path costs
  - represents cost of sending on link and/or link saturation
- Cost normalization
  - prevents ISPs from inflating their path costs
- Bidirectional protocol



Example: Deploying Wiser

- Path costs
  - represents cost of sending on link and/or link saturation
- Cost normalization
  - prevents ISPs from inflating their path costs
- Bidirectional protocol



Previous Work

Previous Work

- **Identified features needed to support evolvability**
  - Pass-through support
  - Multi-protocol support

Previous Work

- **Identified features needed to support evolvability**
  - Pass-through support
  - Multi-protocol support
- **Initial implementation sends multi-protocol data out-of-band**
  - i.e., Wiser routers must use external service to communicate info
  - Allows routers to be simpler, but hides info from ISPs

Previous Work

- **Identified features needed to support evolvability**
  - Pass-through support
  - Multi-protocol support
- **Initial implementation sends multi-protocol data out-of-band**
  - i.e., Wiser routers must use external service to communicate info
  - Allows routers to be simpler, but hides info from ISPs

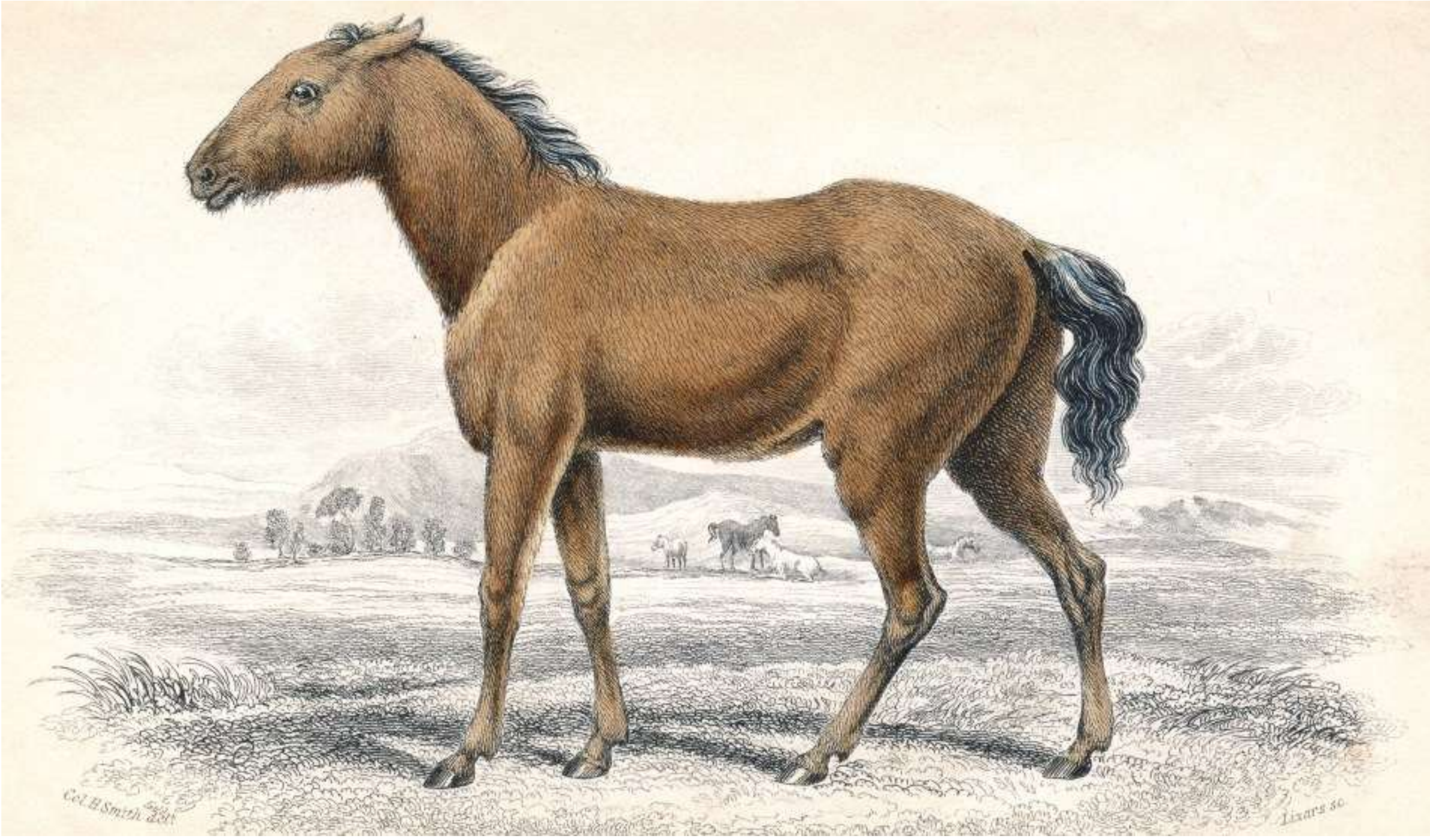
Our work: testing BGP evolvability in-band

Outline

- BGP and its Shortfalls
- Previous Work
- **Tarpan**
  - Design
  - Implementation
- Evaluation
- Conclusion and Future Work

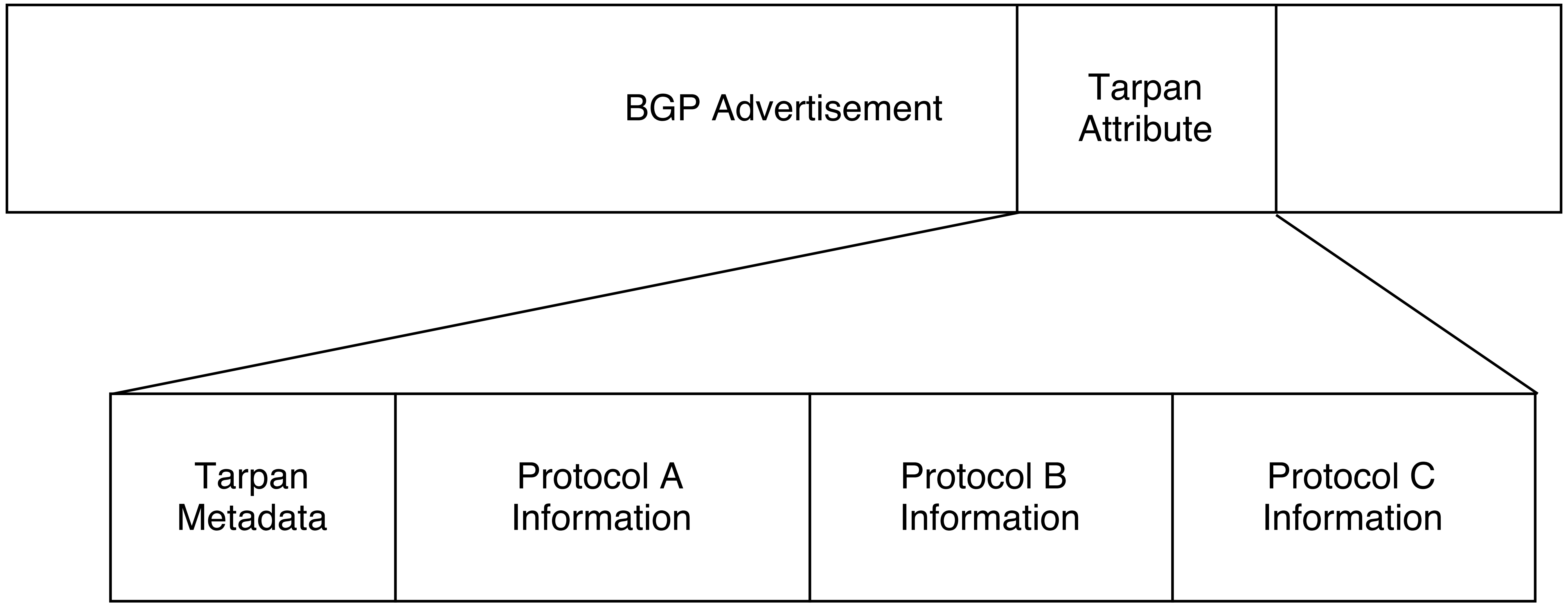
Tarpan

- Favors in-band communication
- Encodes multiple protocols
- Passes through unknown protocols
- Can cross gulfs
- Inter-operates with non-Tarpan routers



Data Structure

- Includes information from multiple protocols
- Tarpan operates as an extension to BGP

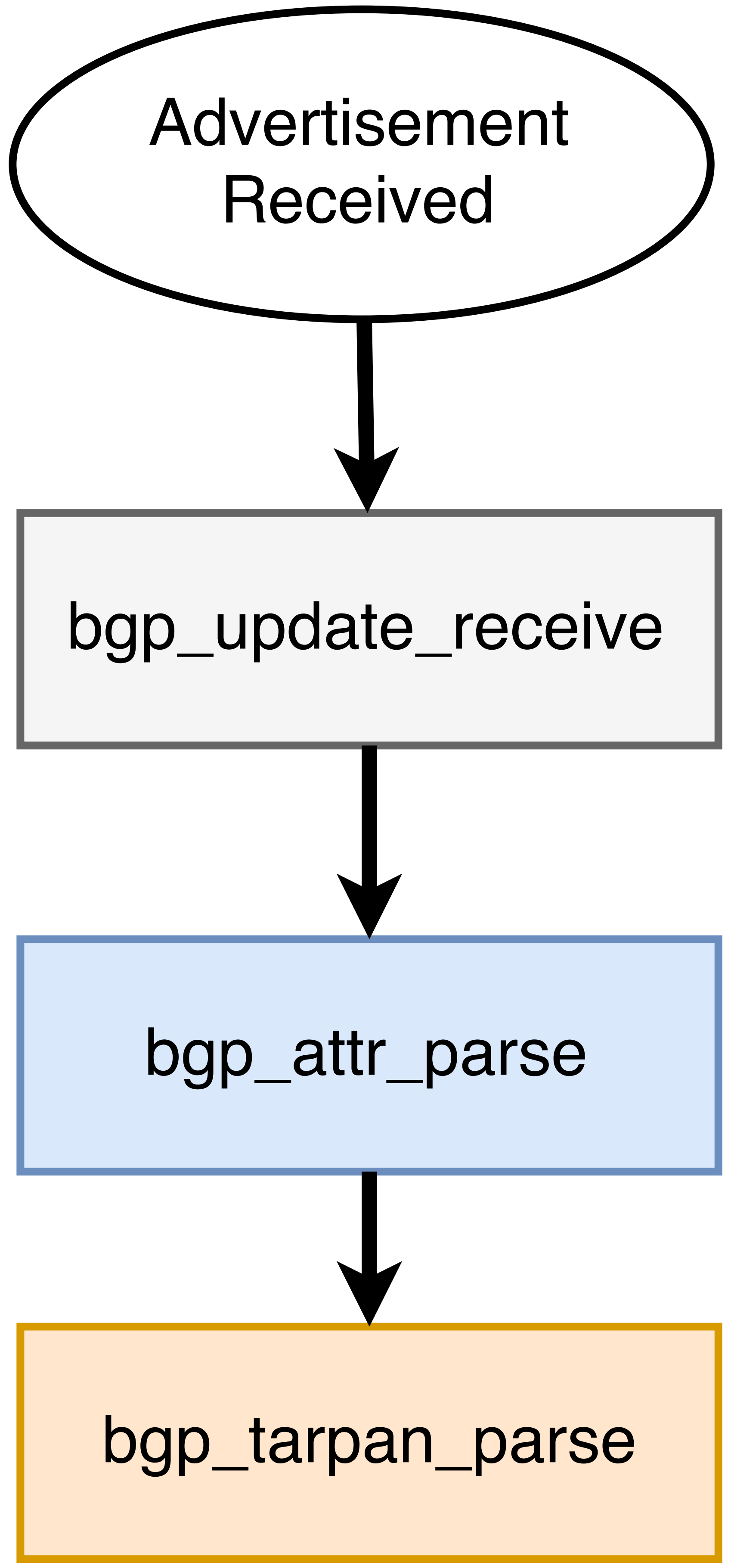


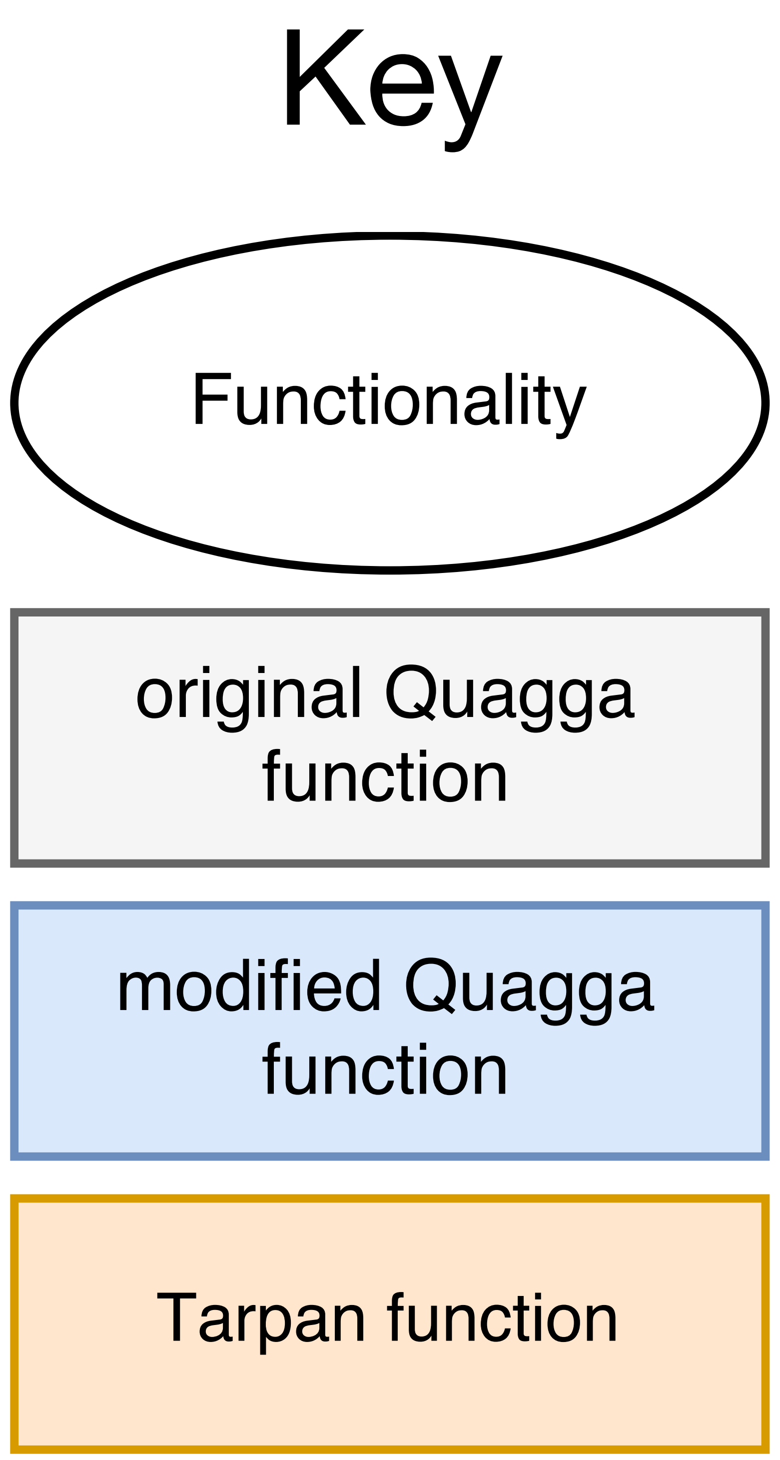
Implementation

- Implemented within Quagga, an open-source network routing suite
  - Quagga itself was a fork of Zebra
- Tarpan API for simply protocol addition
- Protocol Buffers for efficient data transfer
- Interposes on BGP route selection mechanisms
- About 2000 lines of code added or changed

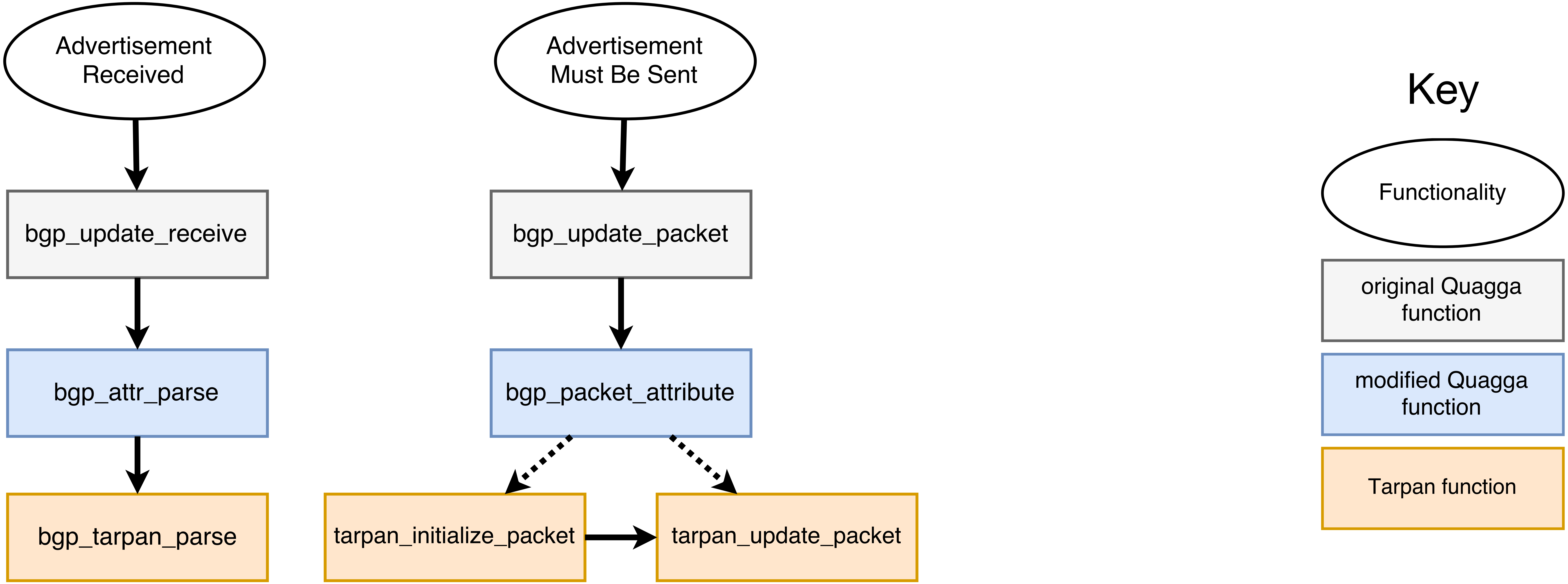
Modifications within Quagga

Modifications within Quagga

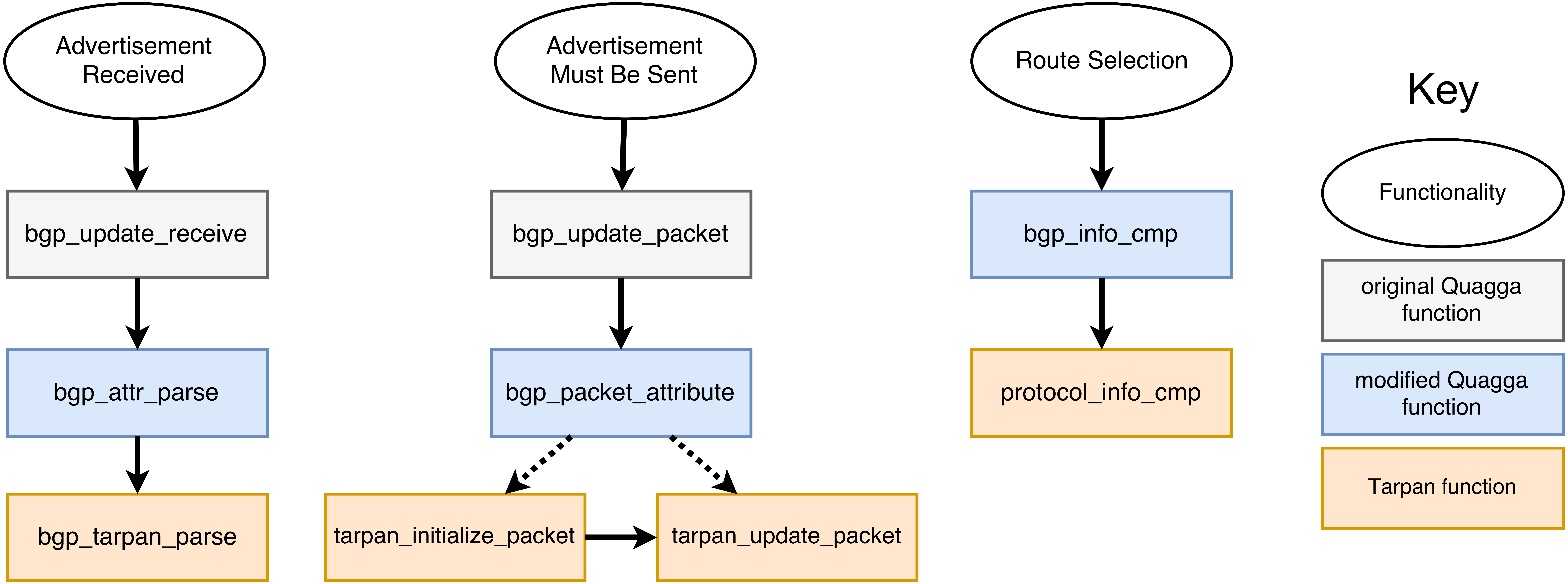




Modifications within Quagga



Modifications within Quagga



Challenges

- Interning
  - Custom memory management scheme
  - Breaks attributes into pieces for memory de-duplication
  - Interns most internal data structures
- Integrating with Quagga’s interning system was major source of frustration

Outline

- BGP and its Shortfalls
- Previous Work
- Tarpan
- **Evaluation**
  - Wiser Testing
  - Throughput Measurement
  - Large Payload Behavior
- Conclusion and Future Work

Experimental Setup

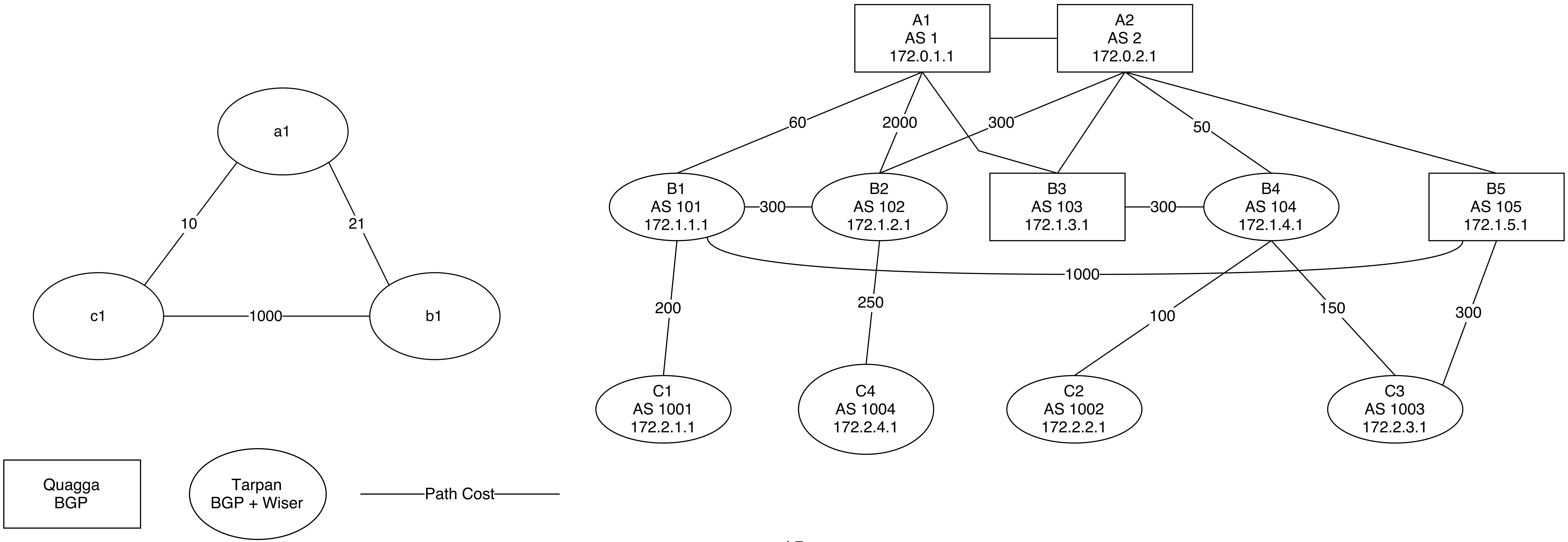
- Topologies emulated with miniNExT
- Ubuntu virtual machine on Massachusetts Open Cloud
  - 16 vCPUs
  - 64 GB RAM

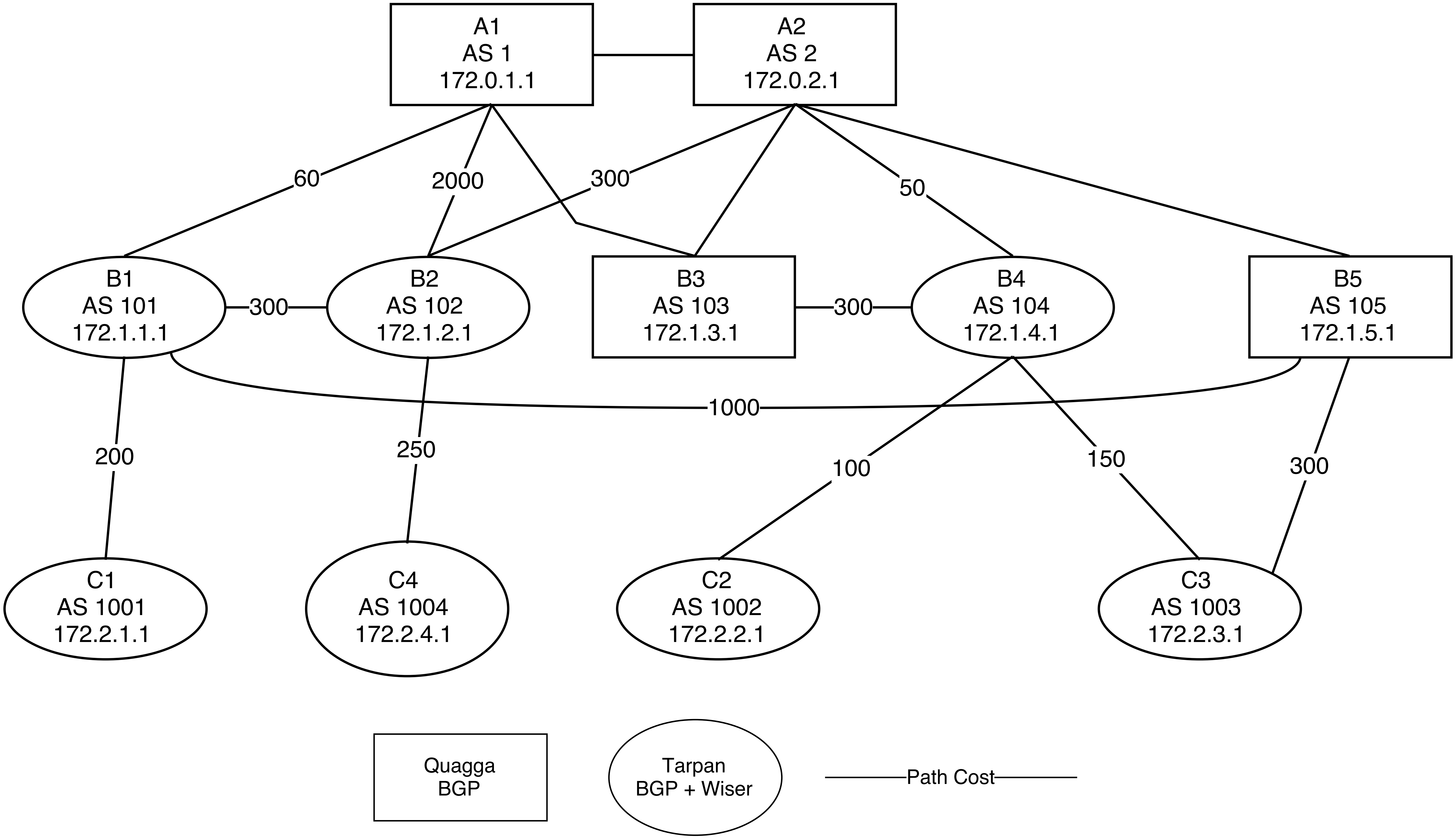
Wiser Experiment Overview

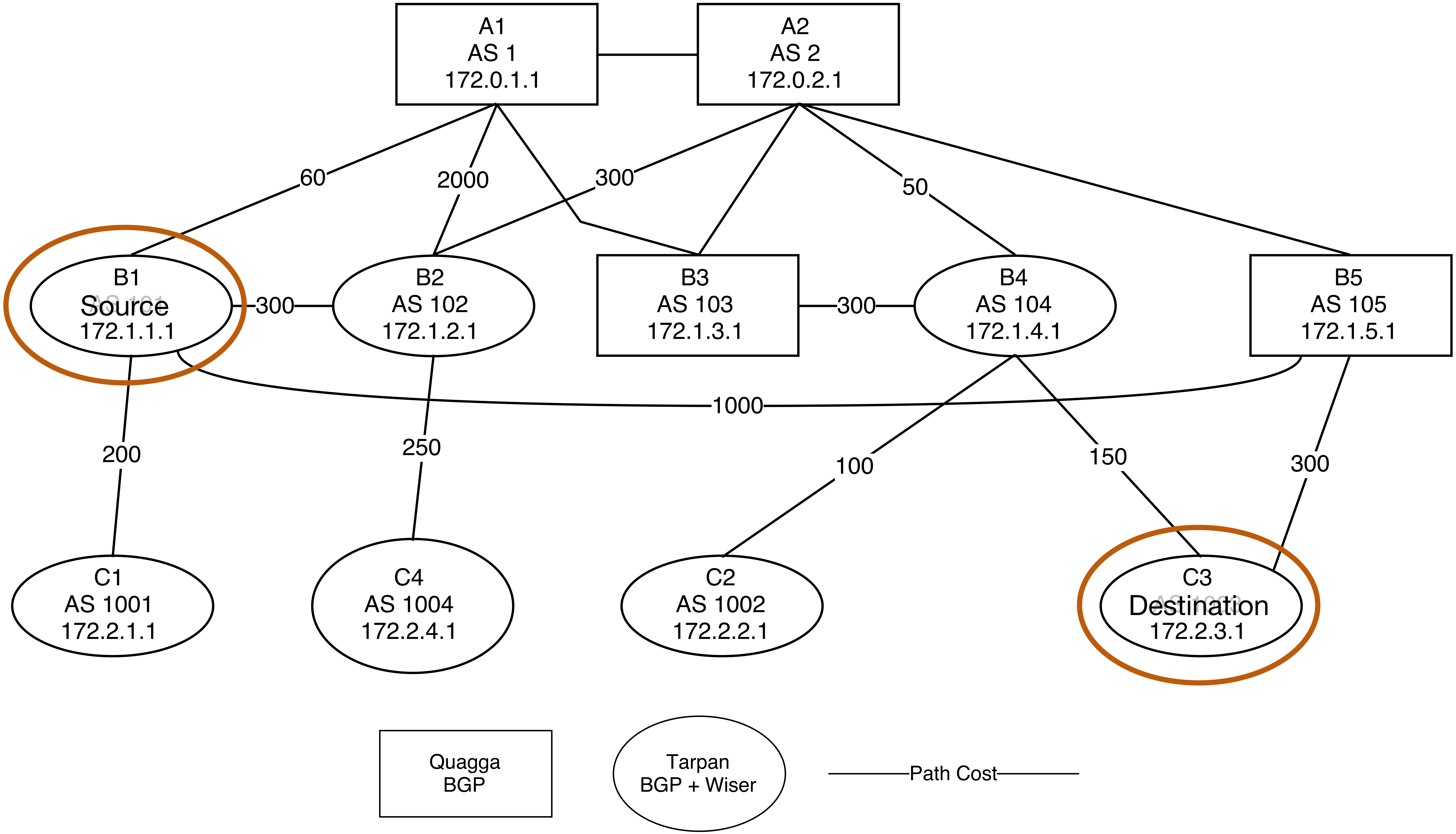
- Proof of concept to demonstrate that Tarpan functions as intended
- Tarpan was extended to use out-of-band communication to support Wiser’s cost normalization using bidirectional communication
- Implemented within Tarpan

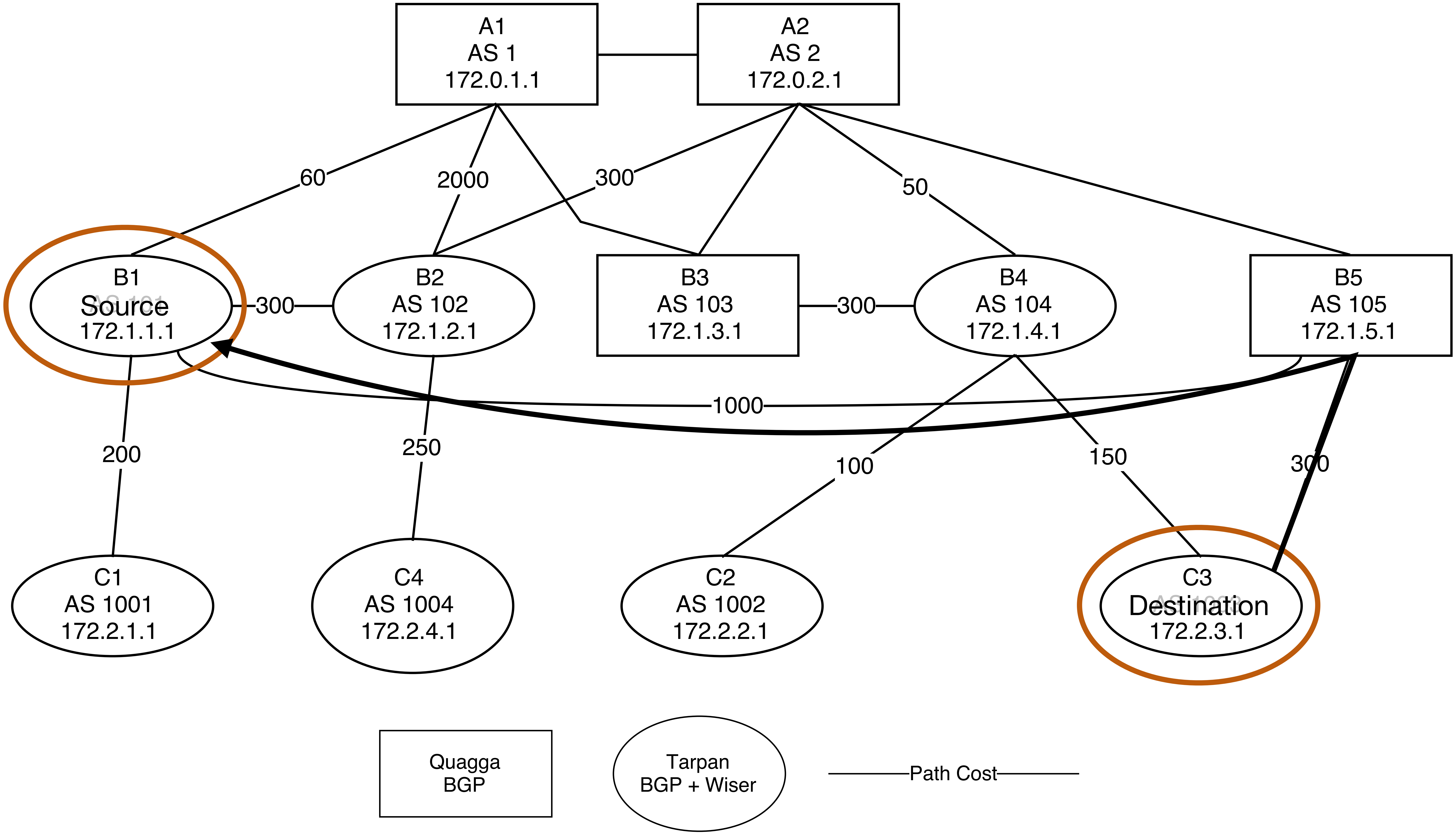
Wiser Testing

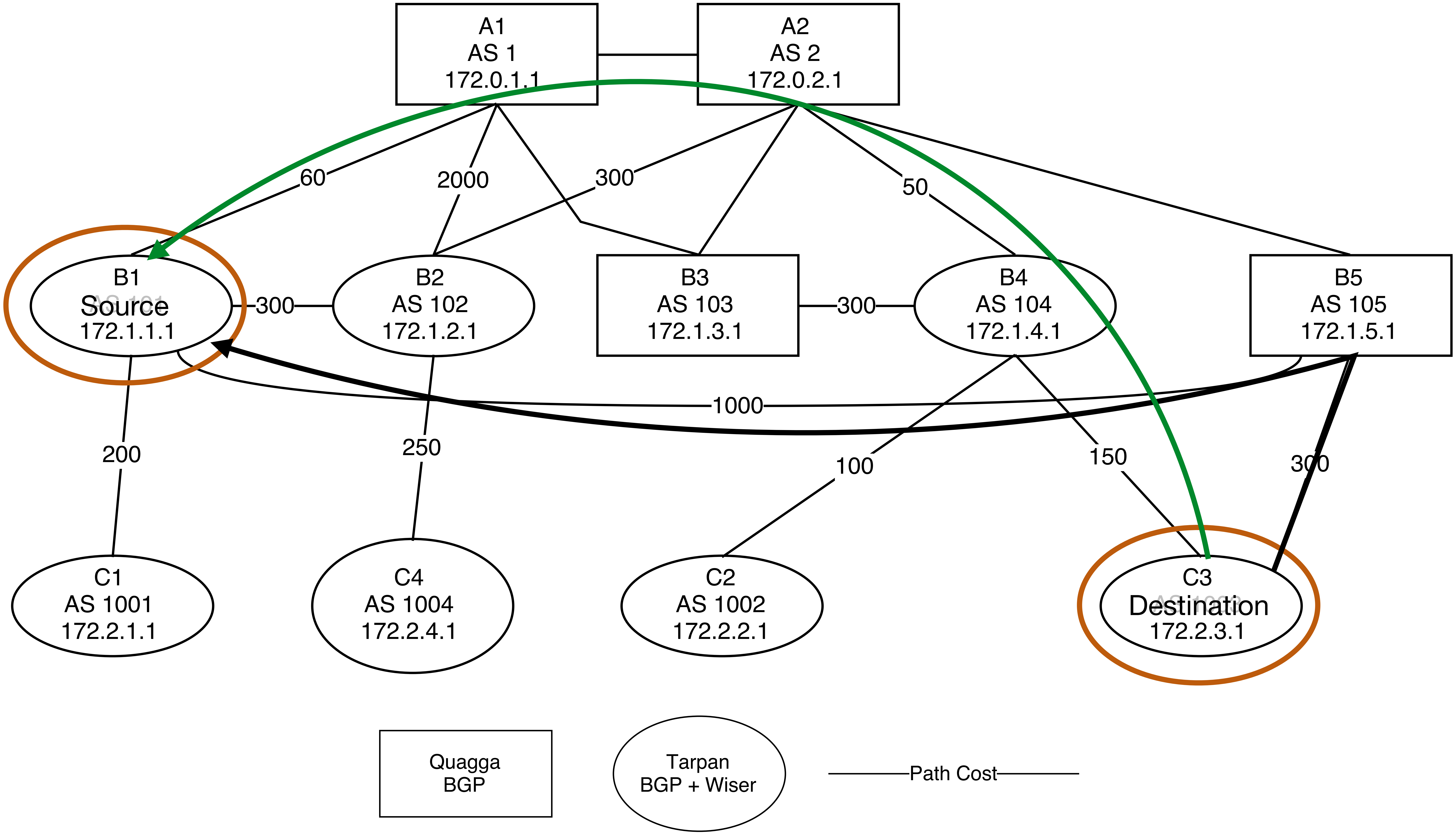
- Ensuring proper Wiser functionality
- Manual verification of route selections



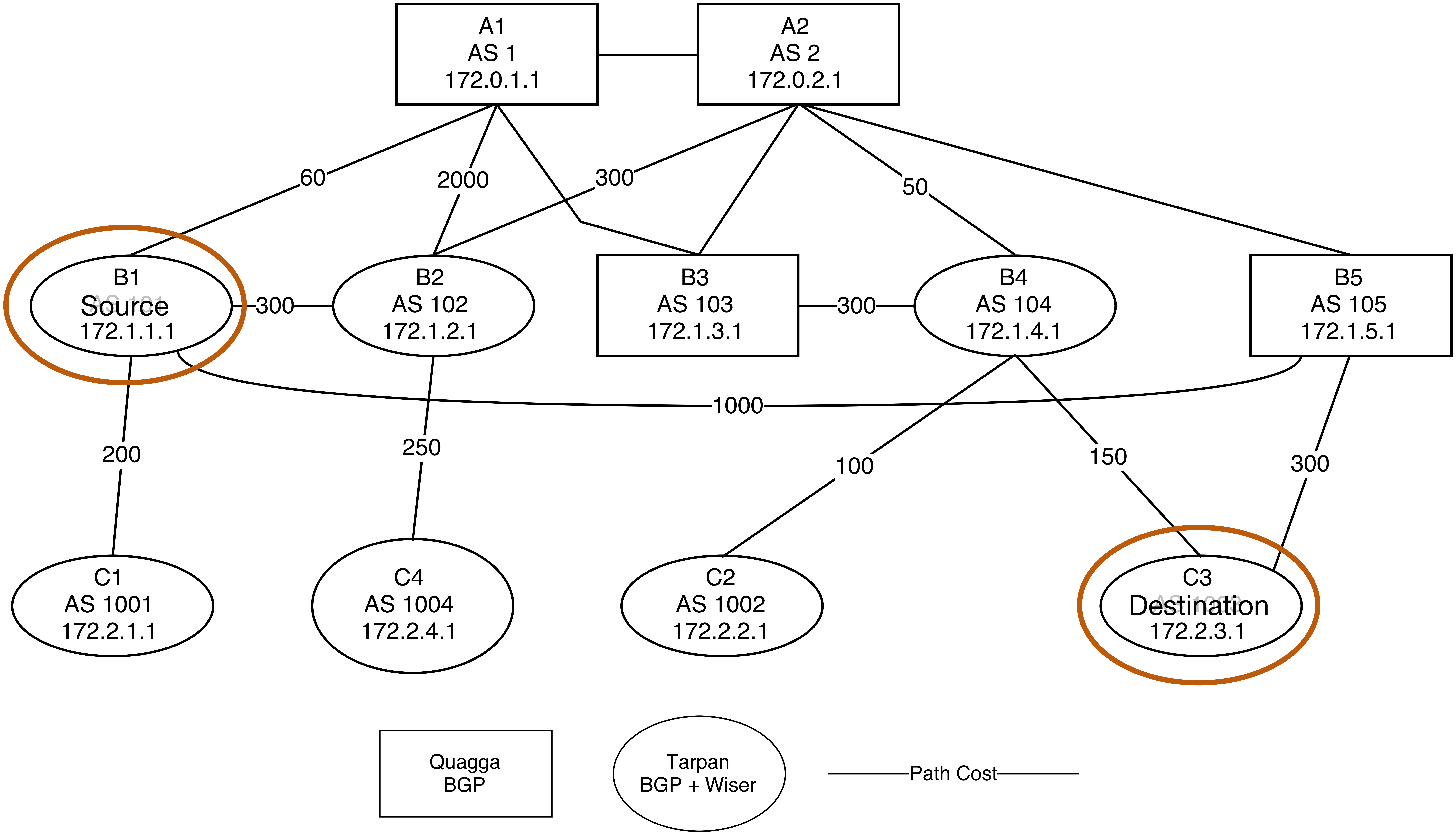

A1
AS 1
172.0.1.1
A2
AS 2
172.0.2.1
60
2000
300
50
B1
AS 101
172.1.1.1
300
B2
AS 102
172.1.2.1
B3
AS 103
172.1.3.1
300
B4
AS 104
172.1.4.1
B5
AS 105
172.1.5.1
1000
200
250
100
150
300
C1
AS 1001
172.2.1.1
C4
AS 1004
172.2.4.1
C2
AS 1002
172.2.2.1
C3
AS 1003
172.2.3.1
Quagga
BGP
Tarpan
BGP + Wiser
Path Cost

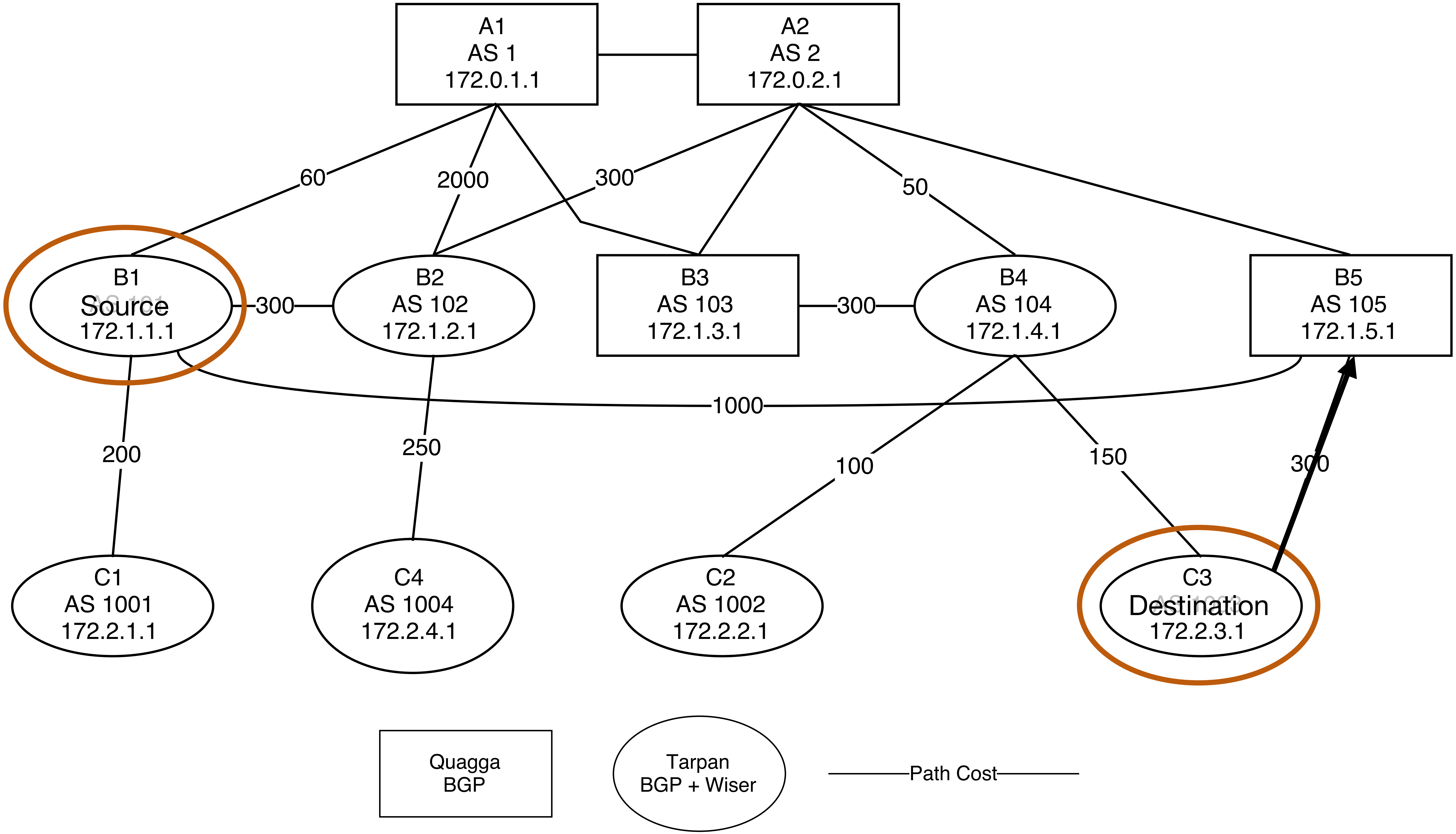

A1
AS 1
172.0.1.1
A2
AS 2
172.0.2.1
60
2000
300
50
B1
Source
172.1.1.1
300
B2
AS 102
172.1.2.1
B3
AS 103
172.1.3.1
300
B4
AS 104
172.1.4.1
B5
AS 105
172.1.5.1
1000
200
250
100
150
300
C1
AS 1001
172.2.1.1
C4
AS 1004
172.2.4.1
C2
AS 1002
172.2.2.1
C3
Destination
172.2.3.1
Quagga
BGP
Tarpan
BGP + Wiser
Path Cost

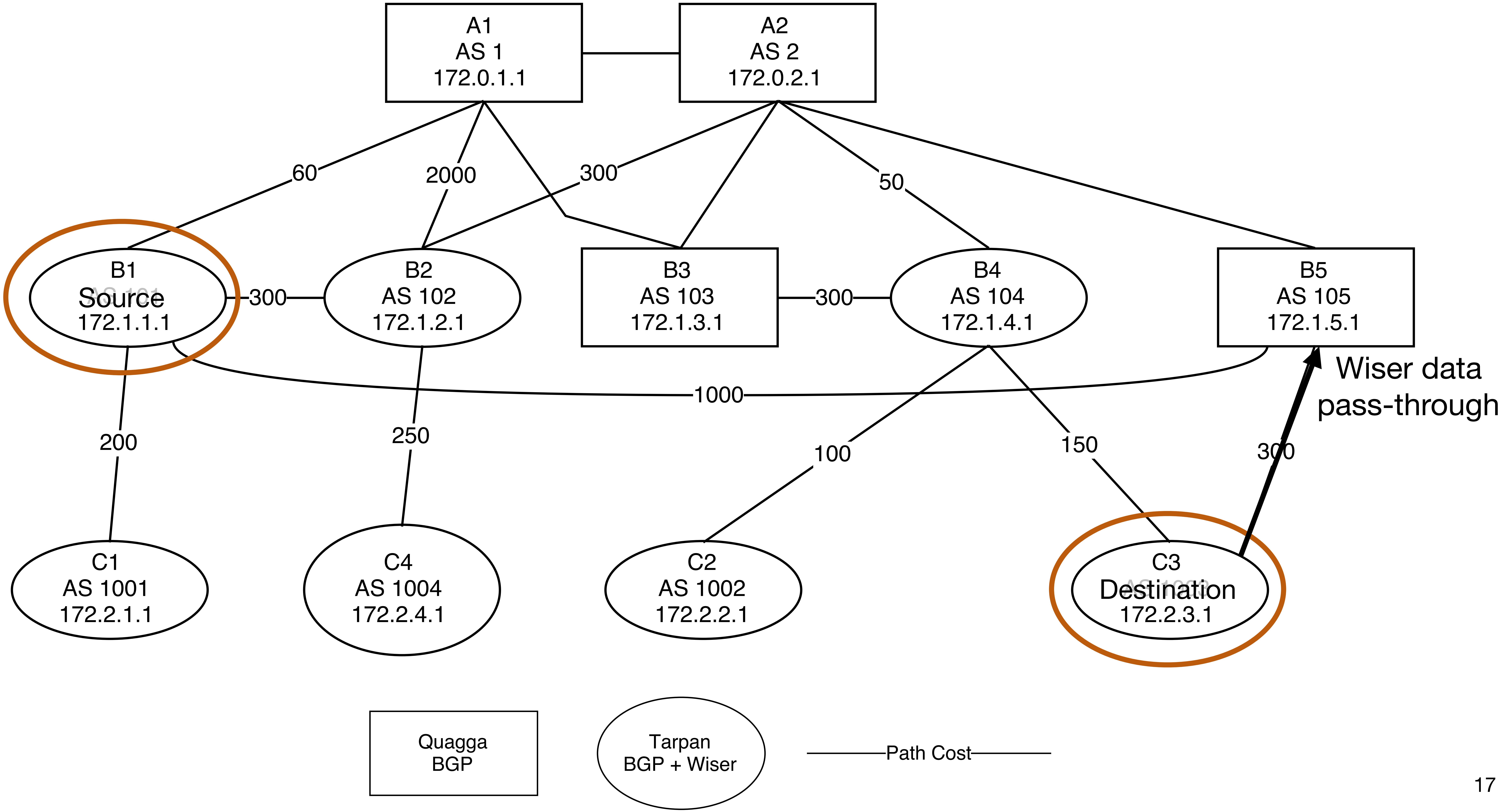

A1
AS 1
172.0.1.1
A2
AS 2
172.0.2.1
60
2000
300
50
B1
Source
172.1.1.1
300
B2
AS 102
172.1.2.1
B3
AS 103
172.1.3.1
300
B4
AS 104
172.1.4.1
B5
AS 105
172.1.5.1
1000
200
250
100
150
300
C1
AS 1001
172.2.1.1
C4
AS 1004
172.2.4.1
C2
AS 1002
172.2.2.1
C3
Destination
172.2.3.1
Quagga
BGP
Tarpan
BGP + Wiser
Path Cost

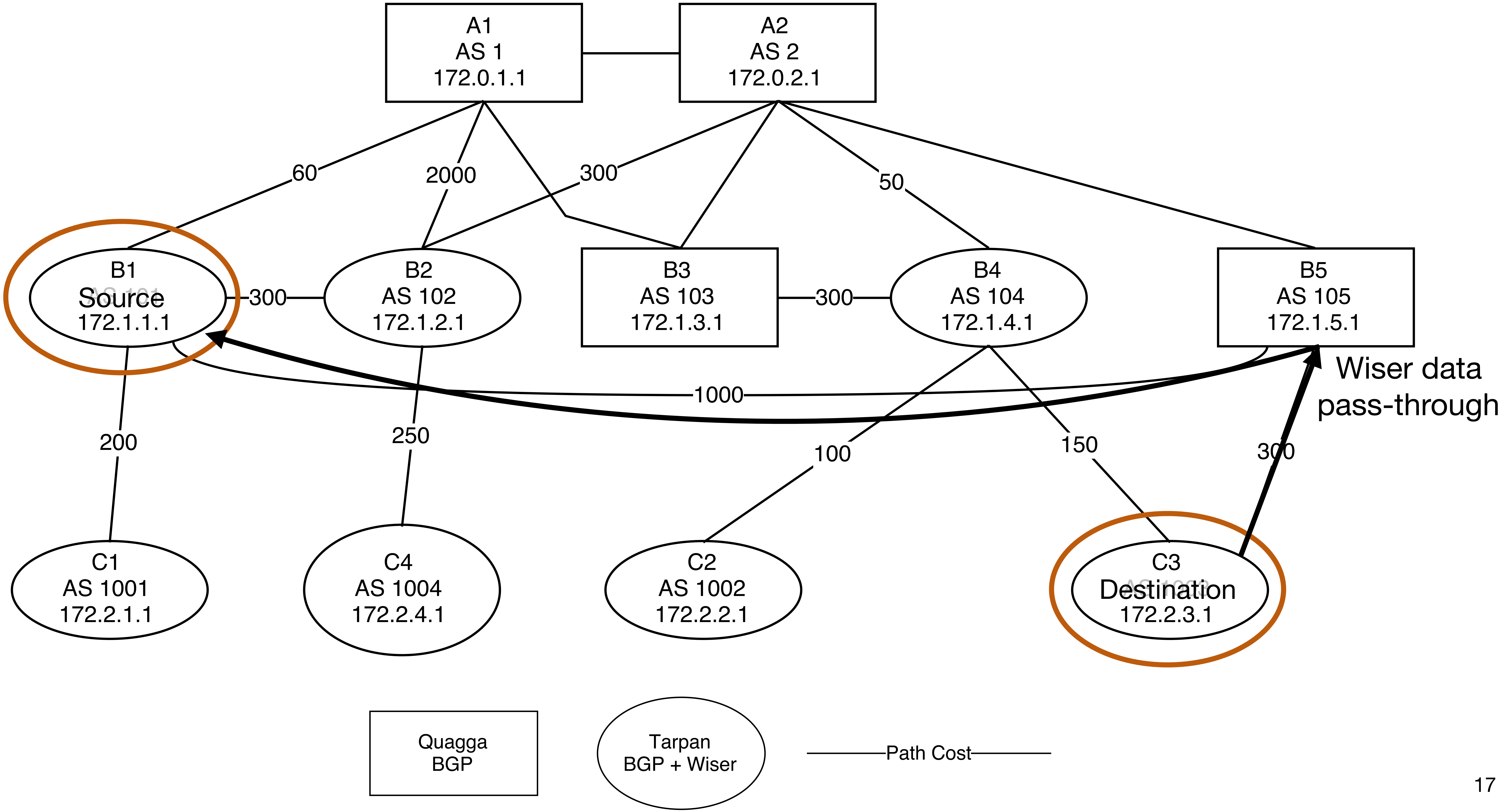


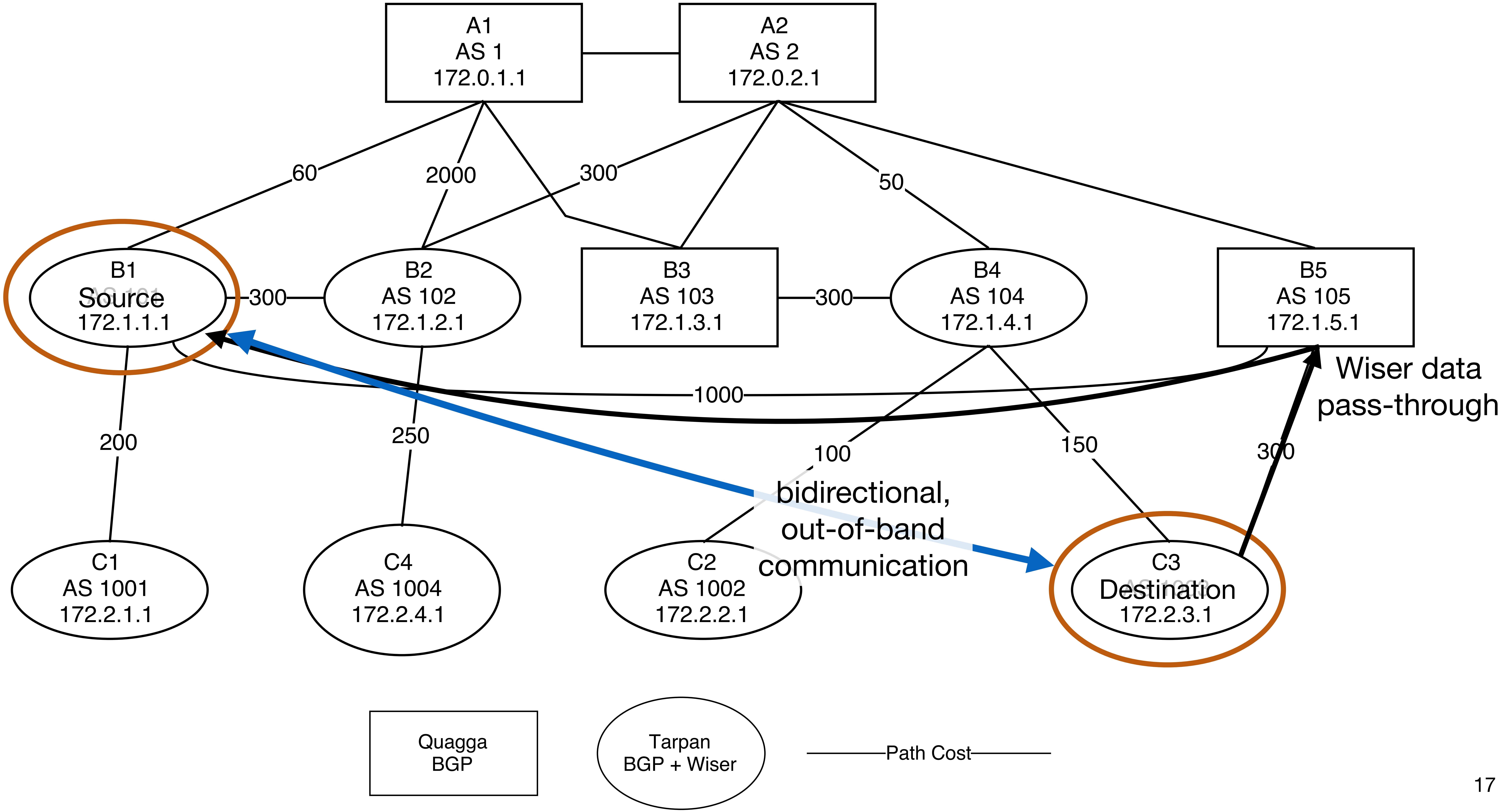
A1
AS 1
172.0.1.1
A2
AS 2
172.0.2.1
60
2000
300
50
B1
Source
172.1.1.1
300
B2
AS 102
172.1.2.1
B3
AS 103
172.1.3.1
300
B4
AS 104
172.1.4.1
B5
AS 105
172.1.5.1
1000
200
250
100
150
300
C1
AS 1001
172.2.1.1
C4
AS 1004
172.2.4.1
C2
AS 1002
172.2.2.1
C3
Destination
172.2.3.1
Quagga
BGP
Tarpan
BGP + Wiser
Path Cost


A1
AS 1
172.0.1.1
A2
AS 2
172.0.2.1
60
2000
300
50
B1
Source
172.1.1.1
300
B2
AS 102
172.1.2.1
B3
AS 103
172.1.3.1
300
B4
AS 104
172.1.4.1
B5
AS 105
172.1.5.1
1000
200
250
100
150
300
C1
AS 1001
172.2.1.1
C4
AS 1004
172.2.4.1
C2
AS 1002
172.2.2.1
C3
Destination
172.2.3.1
Quagga
BGP
Tarpan
BGP + Wiser
Path Cost


A1
AS 1
172.0.1.1
A2
AS 2
172.0.2.1
60
2000
300
50
B1
Source
172.1.1.1
300
B2
AS 102
172.1.2.1
B3
AS 103
172.1.3.1
300
B4
AS 104
172.1.4.1
B5
AS 105
172.1.5.1
1000
200
250
100
150
300
C1
AS 1001
172.2.1.1
C4
AS 1004
172.2.4.1
C2
AS 1002
172.2.2.1
C3
Destination
172.2.3.1
Quagga
BGP
Tarpan
BGP + Wiser
Path Cost

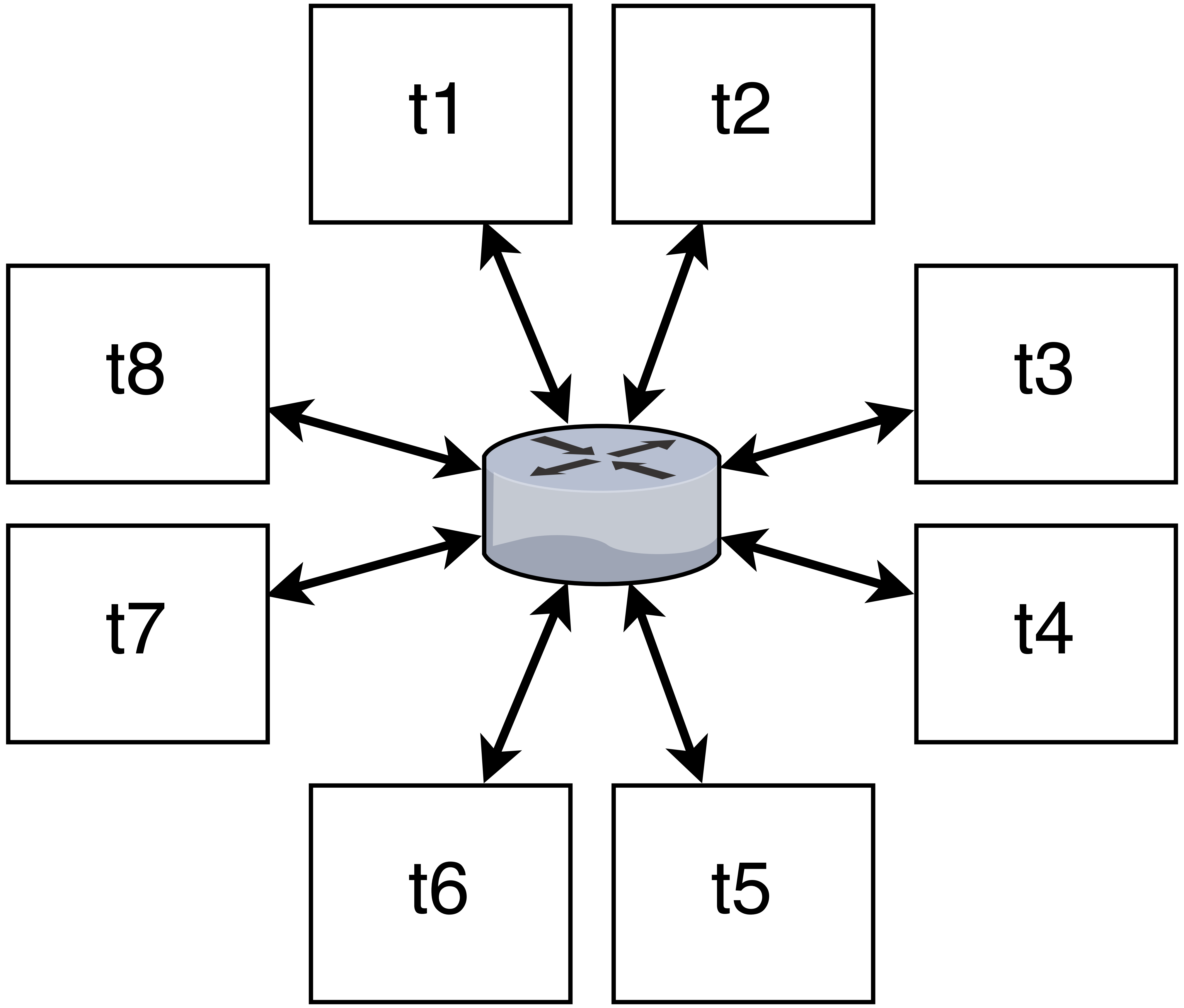

A1
AS 1
172.0.1.1
A2
AS 2
172.0.2.1
60
2000
300
50
B1
Source
172.1.1.1
300
B2
AS 102
172.1.2.1
B3
AS 103
172.1.3.1
300
B4
AS 104
172.1.4.1
B5
AS 105
172.1.5.1
Wiser data
pass-through
1000
200
250
100
150
300
C1
AS 1001
172.2.1.1
C4
AS 1004
172.2.4.1
C2
AS 1002
172.2.2.1
C3
Destination
172.2.3.1
Quagga
BGP
Tarpan
BGP + Wiser
Path Cost


A1
AS 1
172.0.1.1
A2
AS 2
172.0.2.1
60
2000
300
50
B1
Source
172.1.1.1
300
B2
AS 102
172.1.2.1
B3
AS 103
172.1.3.1
300
B4
AS 104
172.1.4.1
B5
AS 105
172.1.5.1
Wiser data
pass-through
1000
200
250
100
150
300
C1
AS 1001
172.2.1.1
C4
AS 1004
172.2.4.1
C2
AS 1002
172.2.2.1
C3
Destination
172.2.3.1
Quagga
BGP
Tarpan
BGP + Wiser
Path Cost


A1
AS 1
172.0.1.1
A2
AS 2
172.0.2.1
60
2000
300
50
B1
Source
172.1.1.1
300
B2
AS 102
172.1.2.1
B3
AS 103
172.1.3.1
300
B4
AS 104
172.1.4.1
B5
AS 105
172.1.5.1
Wiser data
pass-through
1000
200
250
100
150
300
bidirectional,
out-of-band
communication
C1
AS 1001
172.2.1.1
C4
AS 1004
172.2.4.1
C2
AS 1002
172.2.2.1
C3
Destination
172.2.3.1
Quagga
BGP
Tarpan
BGP + Wiser
Path Cost

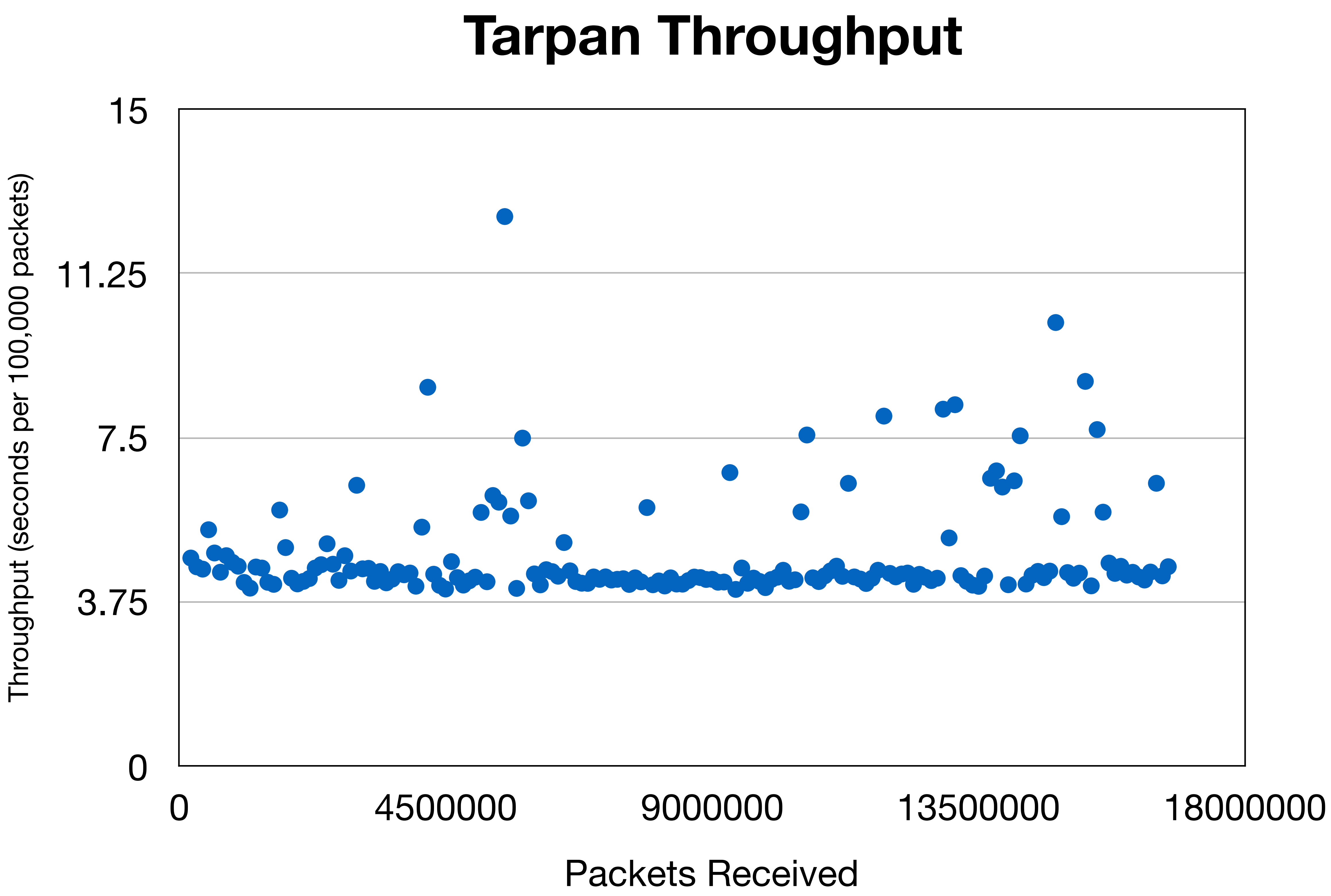
Throughput Setup

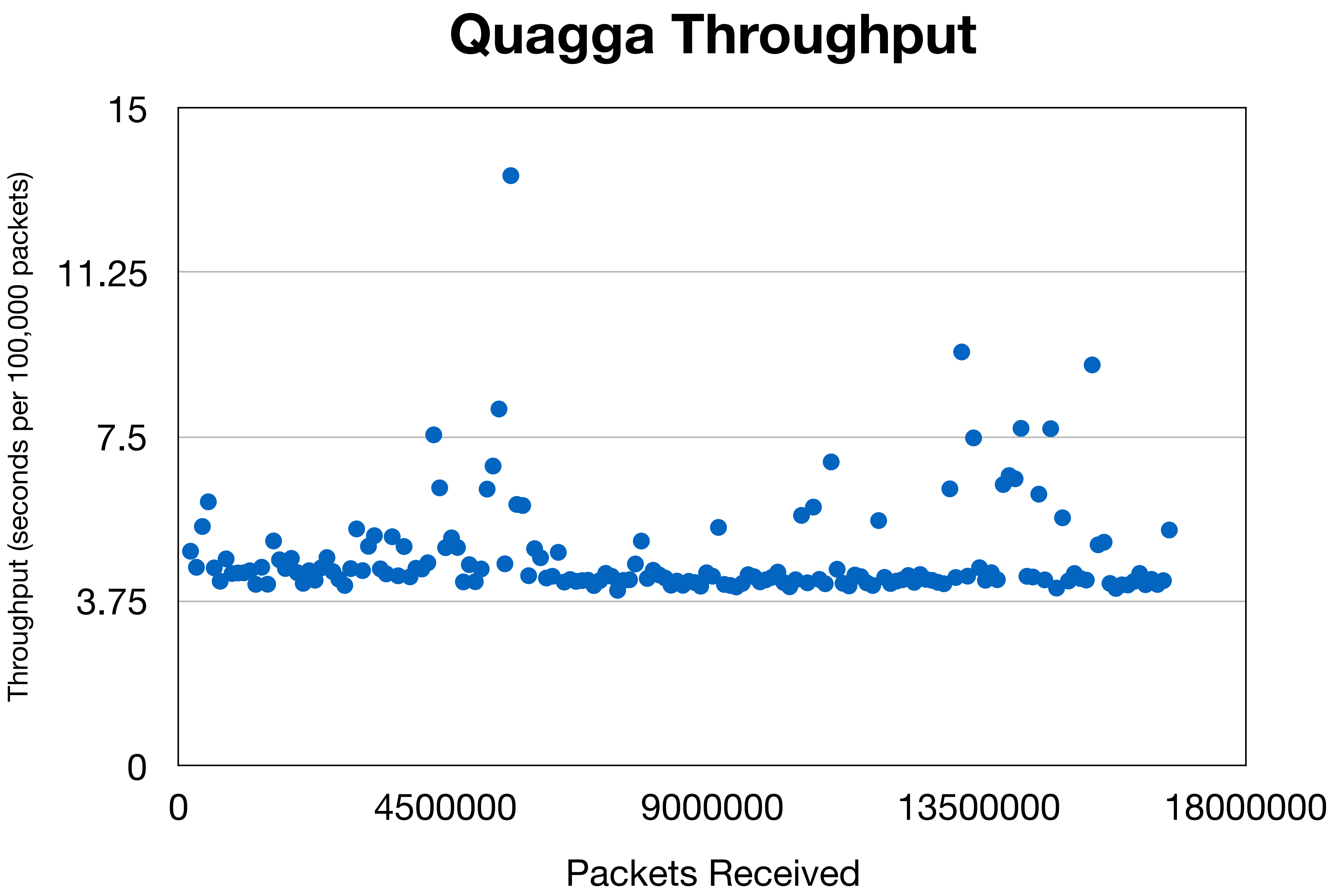
- Two virtual switches in miniNExT
- 8 bgpsimple scripts send actual routing tables into the router
- The router is either Tarpan or Quagga, with instrumentation for timing



Throughput Evaluation

- Graphs show inverted throughput (lower is better)





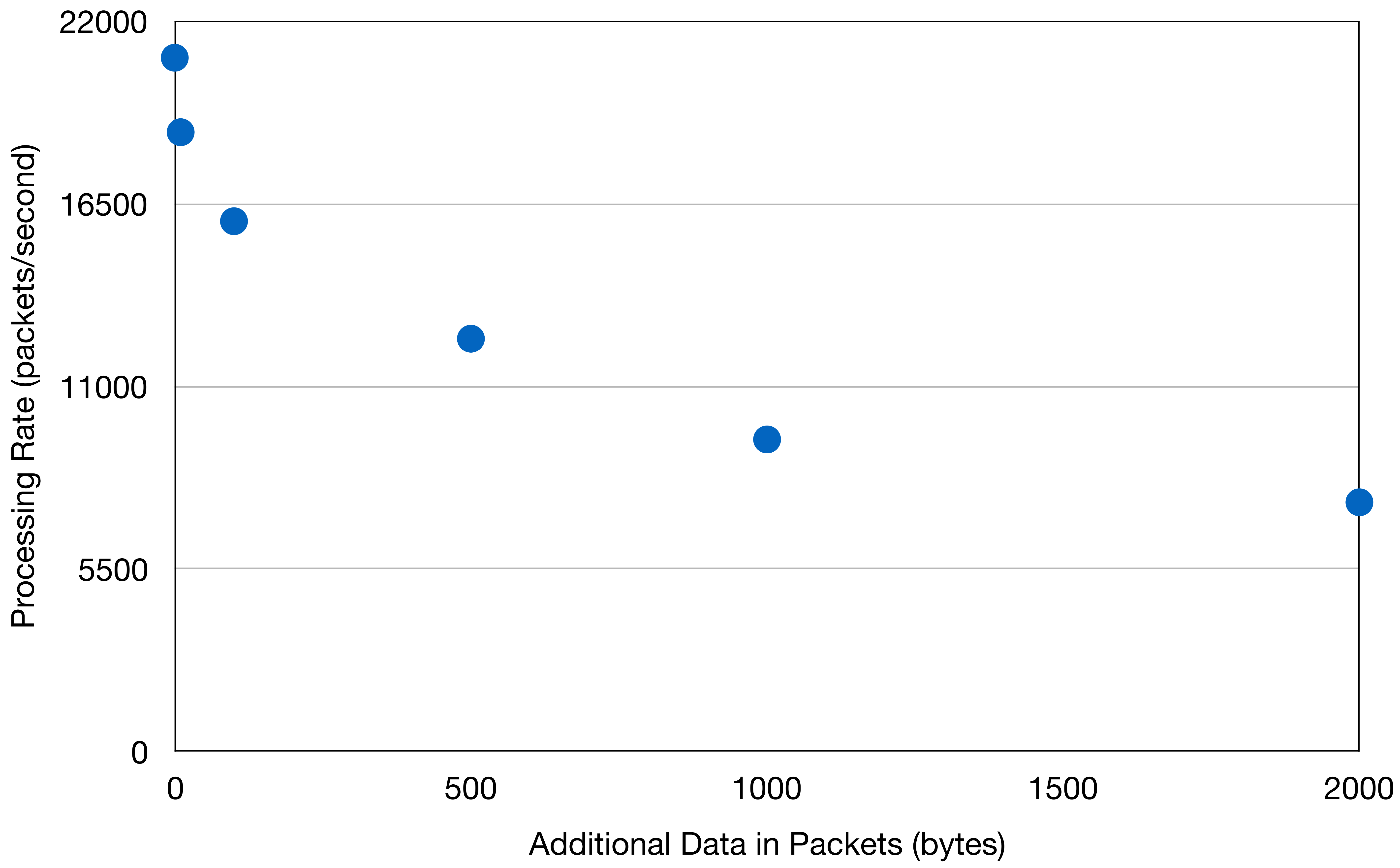
Avg.: 20,787 packets/sec

Avg.: 21,026 packets/sec

Large Payloads

- Test the effect of sending larger payloads to routers
- Modified bgpsimple script that sends a string of certain length in an attribute

Large Payloads



Future Work

- Convergence properties when running multiple protocols
- Exploring incremental deployment
- Further performance and memory usage improvements

Summary

- BGP is too rigid - cannot support deployment of new protocols across gulfs
- Tarpan allows new protocols to be deployed across gulfs by sending information in-band with BGP advertisements
- Wiser implemented using Tarpan’s API
- Low performance overhead

Acknowledgements

- **Raja Sambasivan** - Mentor
- **Massachusetts Open Cloud** - Large virtual machines for testing
- **MIT PRIMES** Program