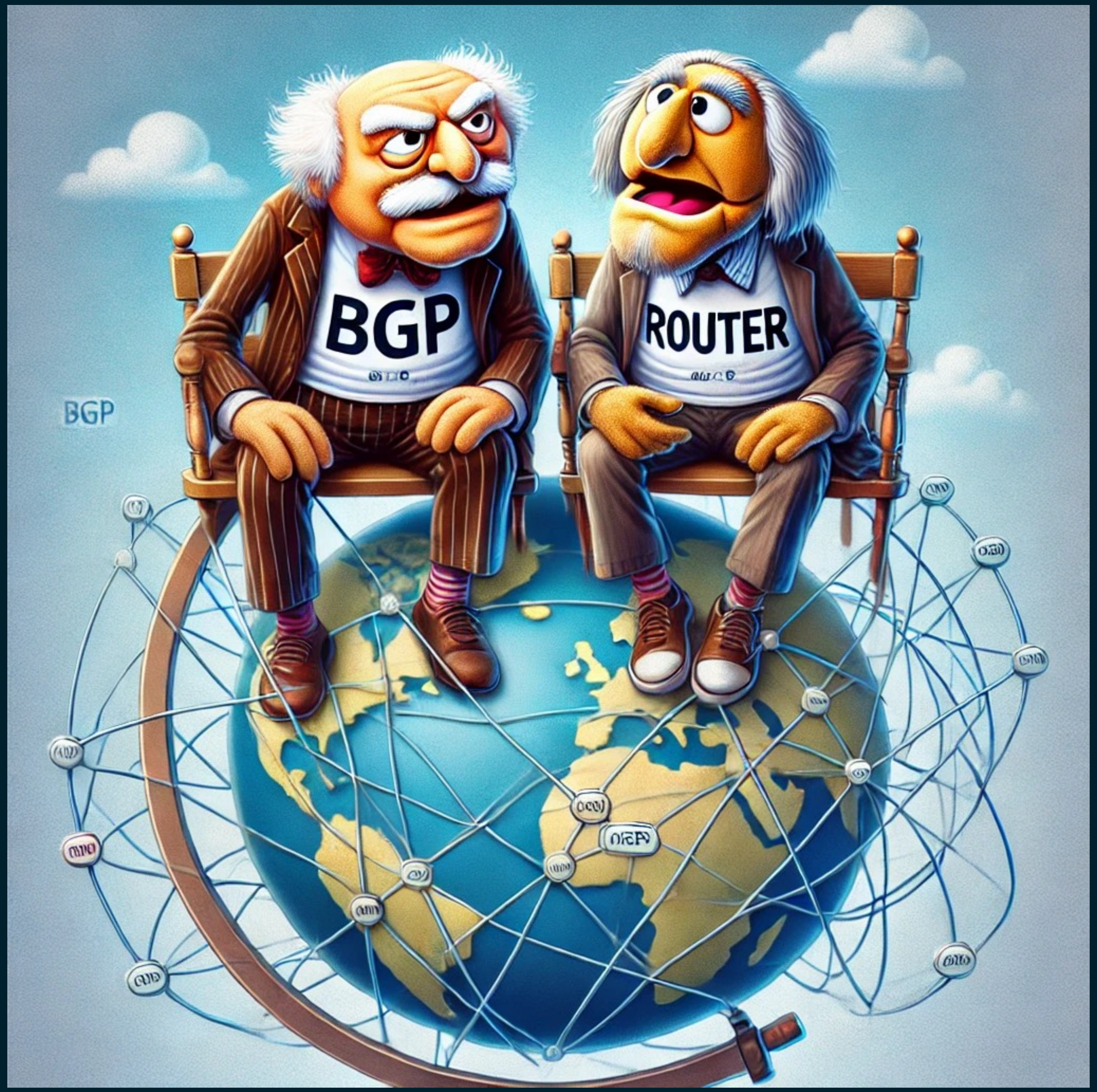
Navigating Internet Research with P4: Solutions for Performance and Security



Oct 3 2024

Maria Apostolaki
netsyn.princeton.edu

What do we want from the Internet?

What do we want from the Internet?

Cyber-physical systems

Live streaming

Video conferencing

Online shopping

Online banking

Cryptocurrencies

What do we want from the Internet?

- Cyber-physical systems
- Live streaming
- Video conferencing
- Online shopping
- Online banking
- Cryptocurrencies

- Low latency
- Throughput
- Privacy
- Reliable connectivity

What do we want from the Internet?

Cyber-physical systems

Cryptocurrencies

Today’s Internet provides best-effort service

…leading to performance, privacy, and security problems

BGP
BGP
ROUTER

Internet research is hindered by both protocols and hardware

Internet research is hindered by both protocols and hardware

BGP….

- lack of route control
- suboptimal routing
- insecure routing
- lack of path diversity

…

Internet research is hindered by both protocols and hardware

BGP….

- lack of route control
- suboptimal routing
- insecure routing
- lack of path diversity

…

Internet Routers….

- fixed-headers support
- no cryptographic operation
- lack of performance visibility
- no DDoS support

….

What can you do with a couple of programmable points in the Internet?

What can you do with a couple of programmable points in the Internet?

Tango: performance-driven routing system

NSDI’24

SABRE: secure overlay for BTC block propagation

NDSS’19

What can you do with a couple of programmable points in the Internet?

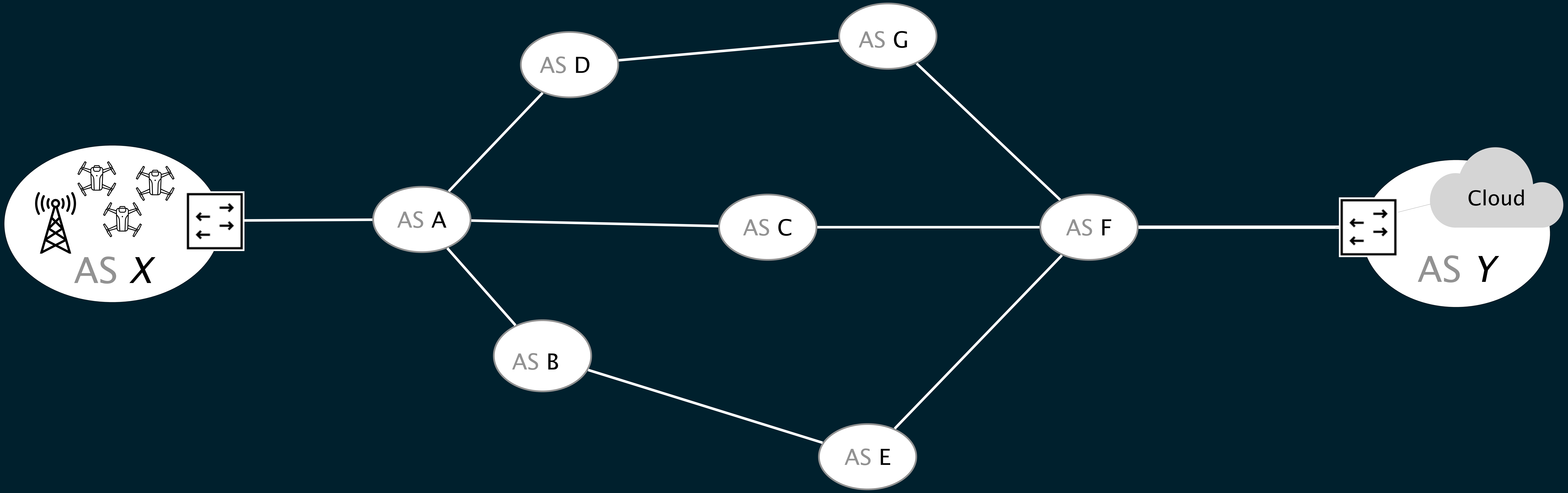
Tango: performance-driven routing system

NSDI'24

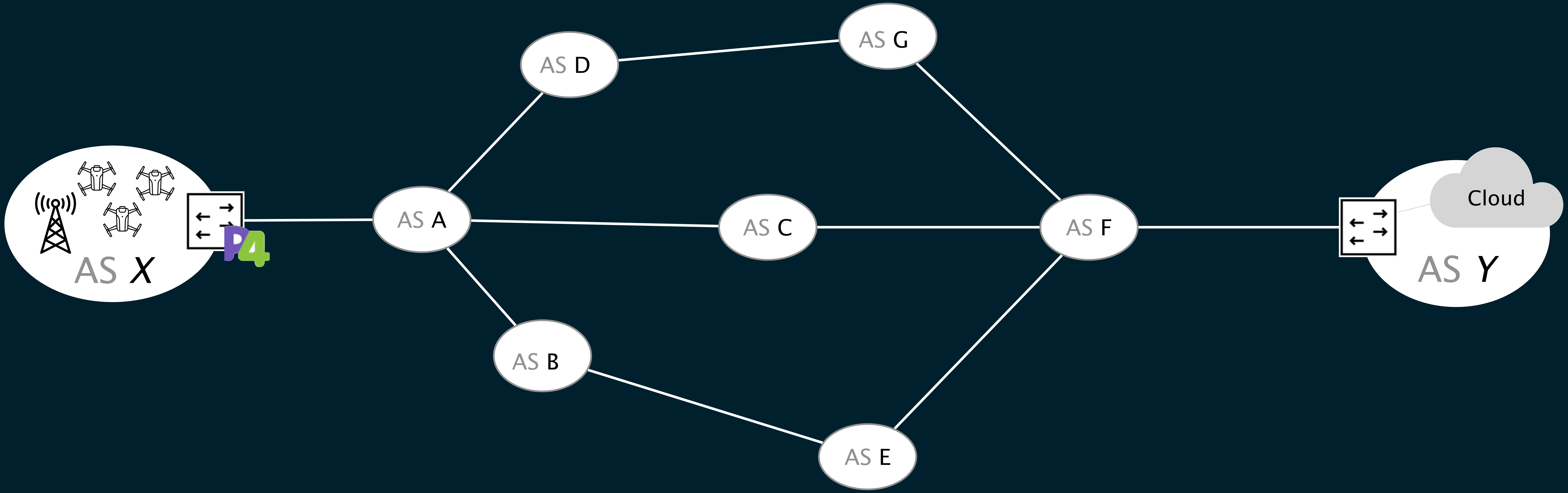
SABRE: secure overlay for BTC block propagation

NDSS'19

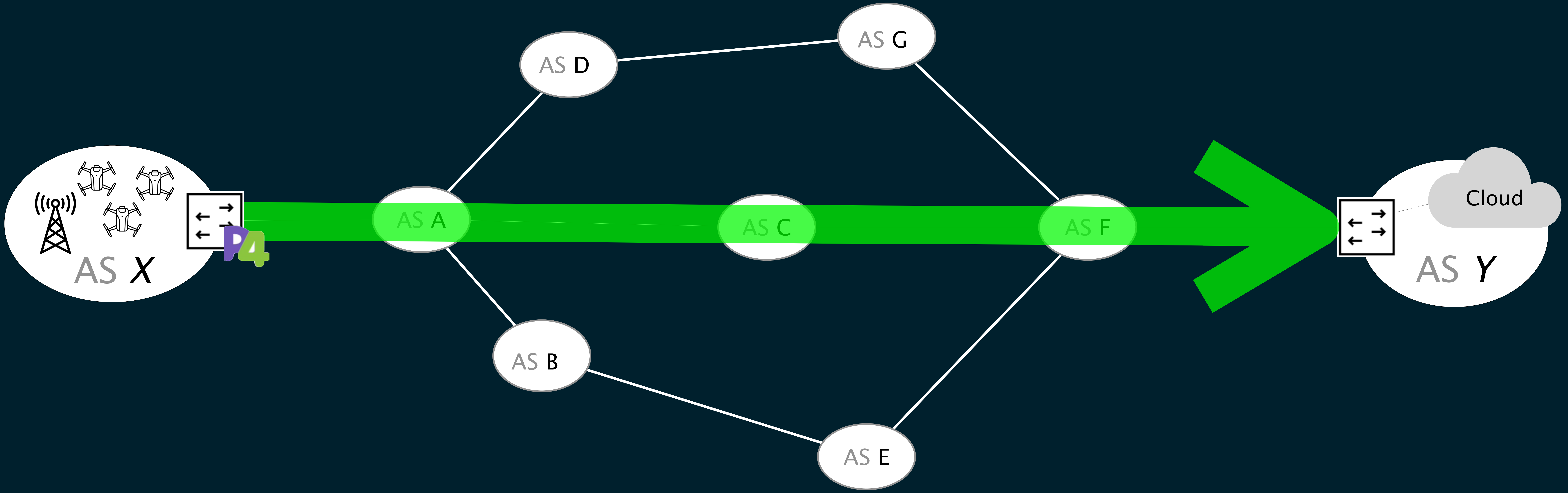
To communicate with ASY, ASX



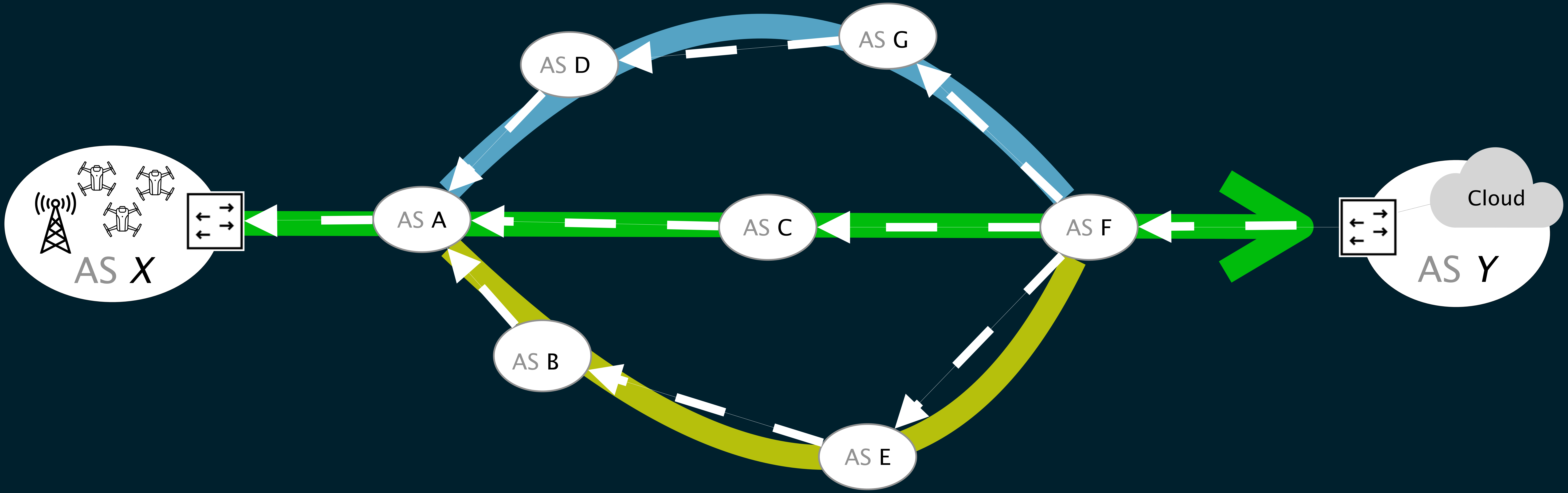
To communicate with ASY, ASX



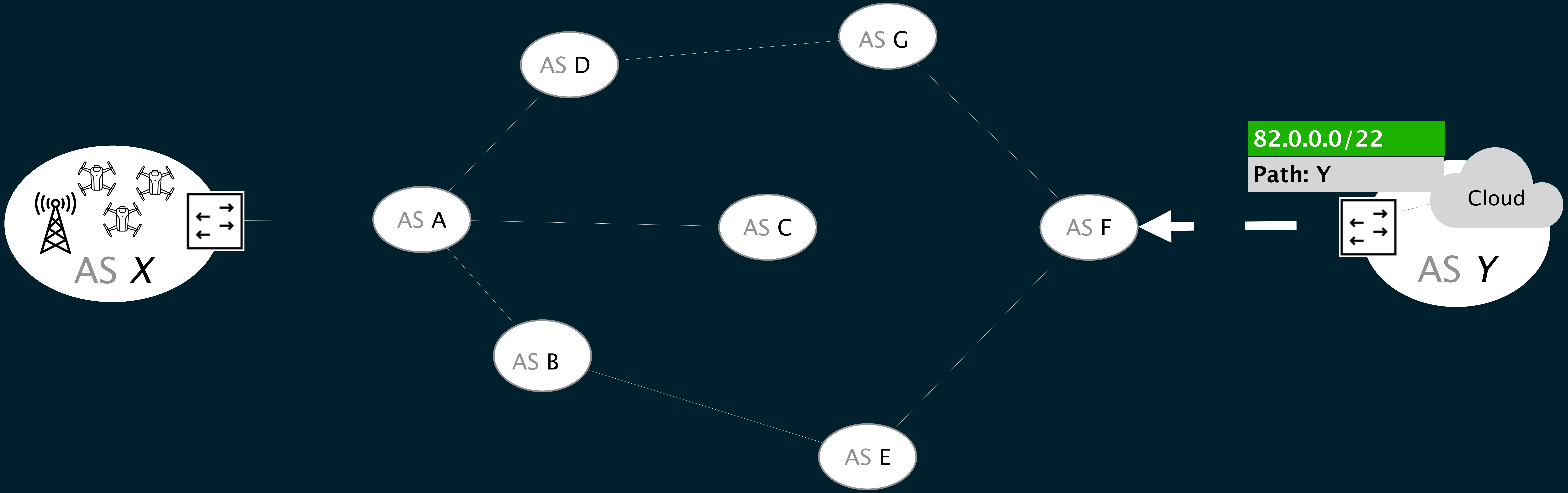
To communicate with ASY, ASX can only use one path



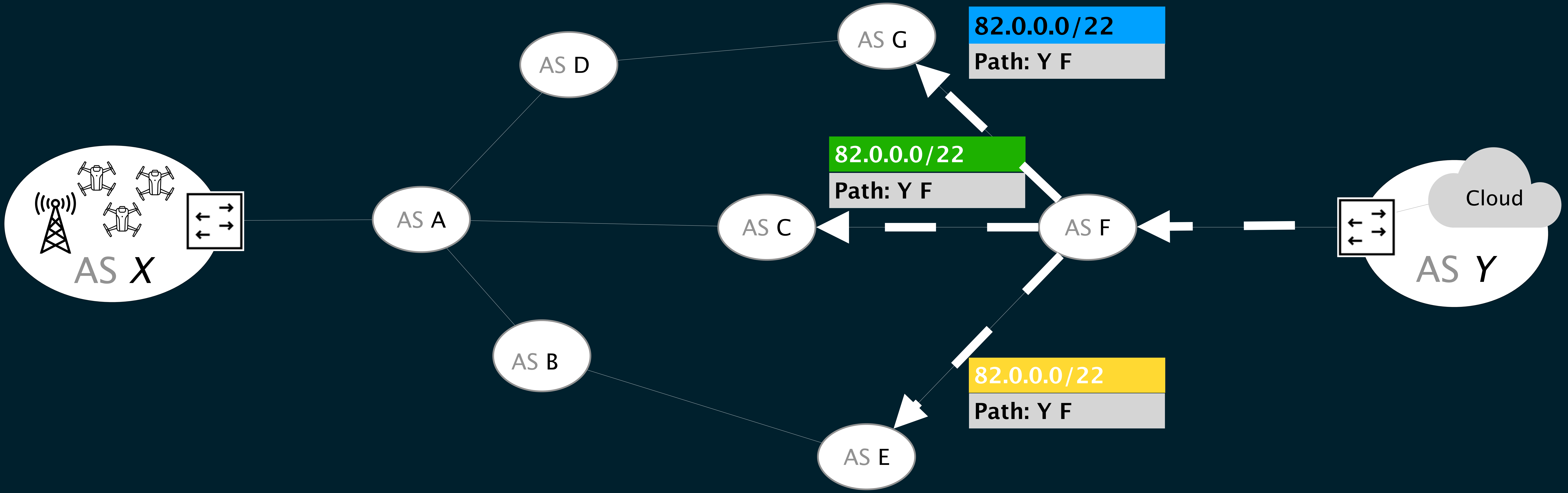
To communicate with ASY, ASX can only use one path despite the path diversity, and independently of performance



BGP
BGP
ROUTER


AS G
AS D
82.0.0.0/22
Path: Y
Cloud
AS X
AS A
AS C
AS F
AS Y
AS B
AS E

The BGP advertisement is propagated via multiple paths

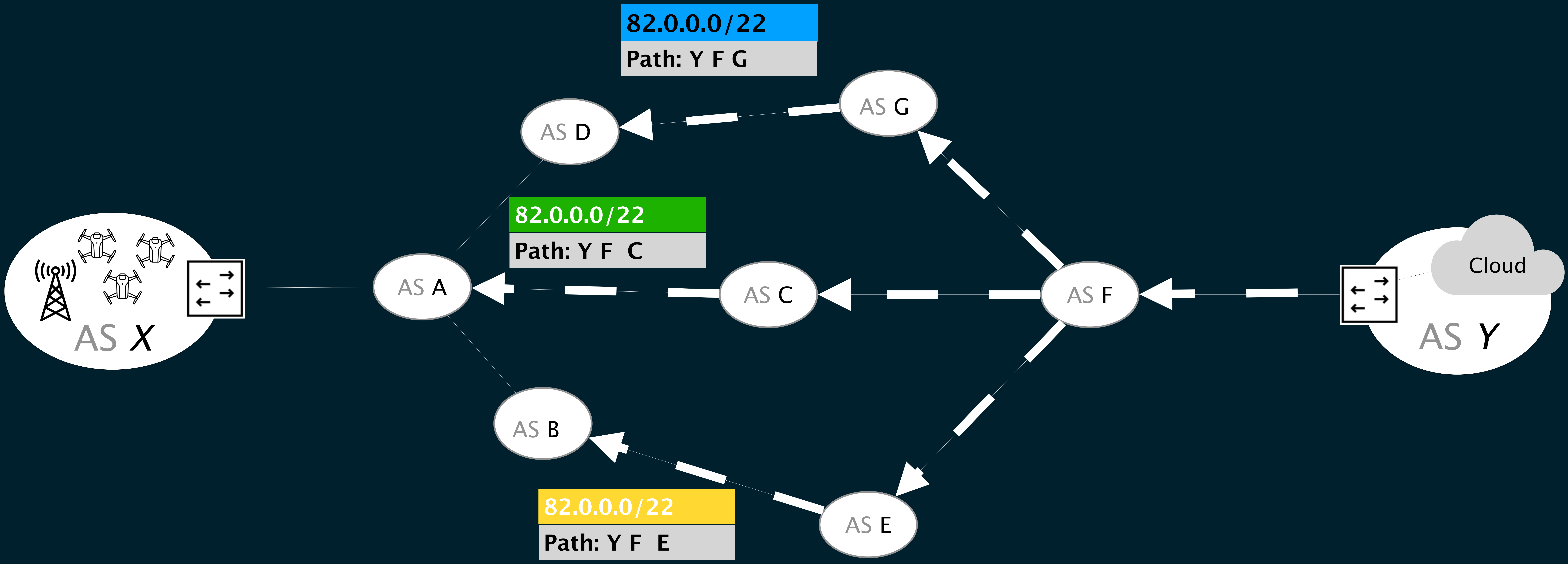


82.0.0.0/22
Path: Y F

82.0.0.0/22
Path: Y F

82.0.0.0/22
Path: Y F

The BGP advertisement is propagated via multiple paths



The BGP advertisement is propagated via multiple paths
But only a single advertisement reaches the sender

AS G
AS D
AS X
82.0.0.0/22
Path: Y F C A
AS A
AS C
AS F
Cloud
AS Y
AS B
AS E

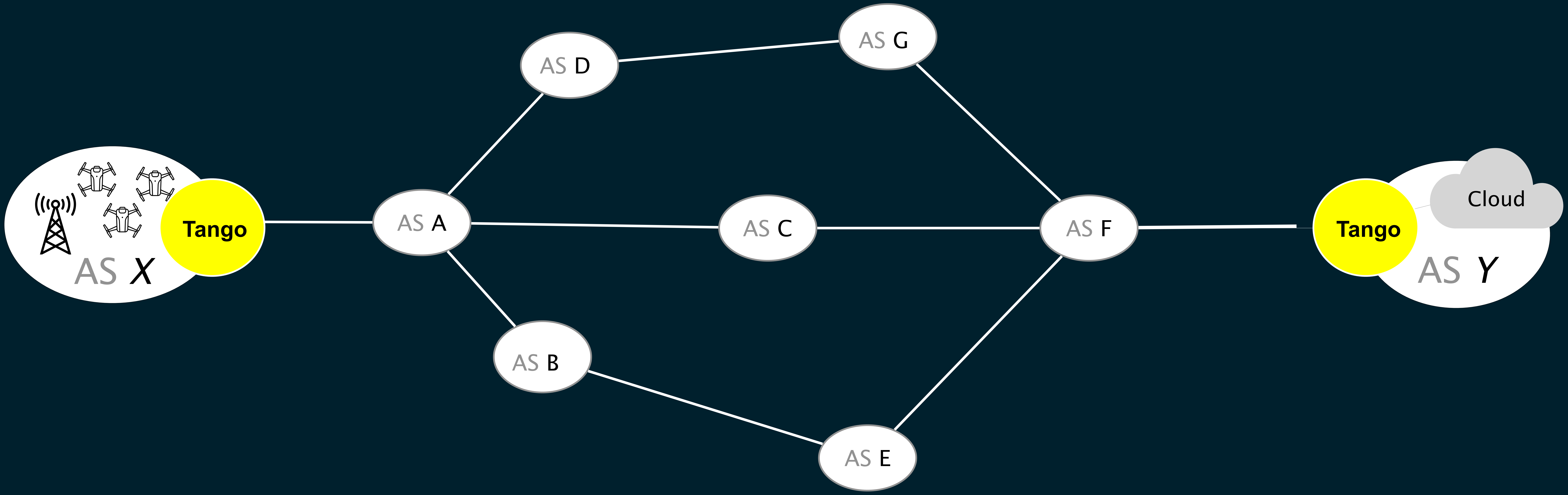
What can you do with a couple of programmable points in the Internet?

Tango: performance-driven routing system

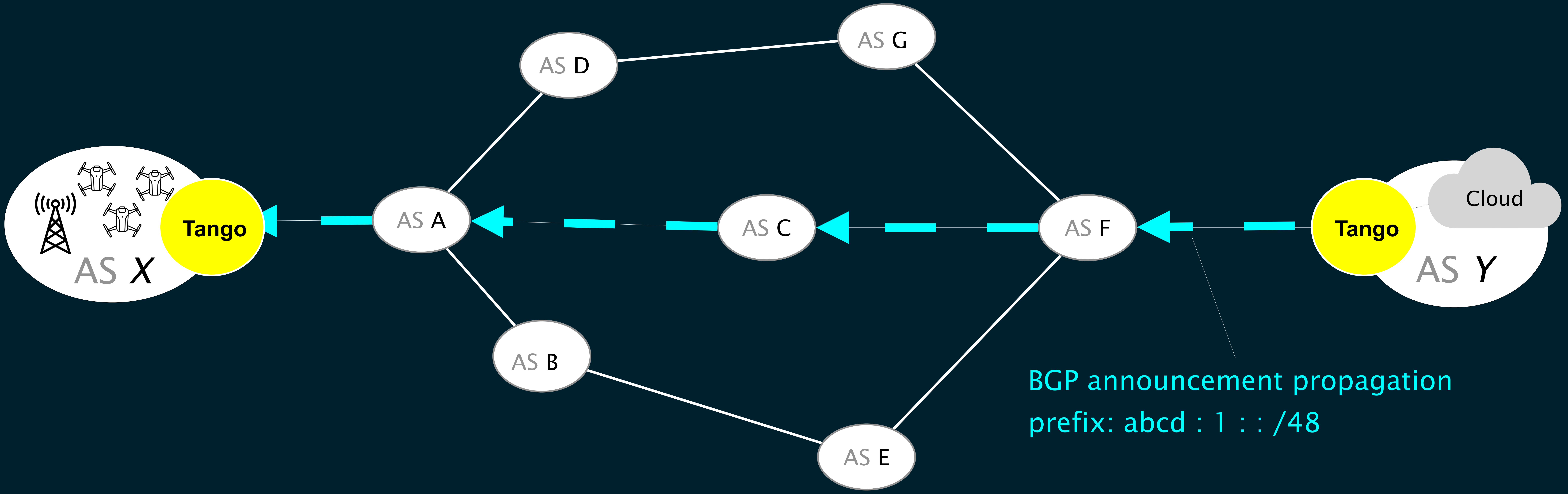
NSDI'24

SABRE: secure overlay for BTC block propagation

NDSS'19

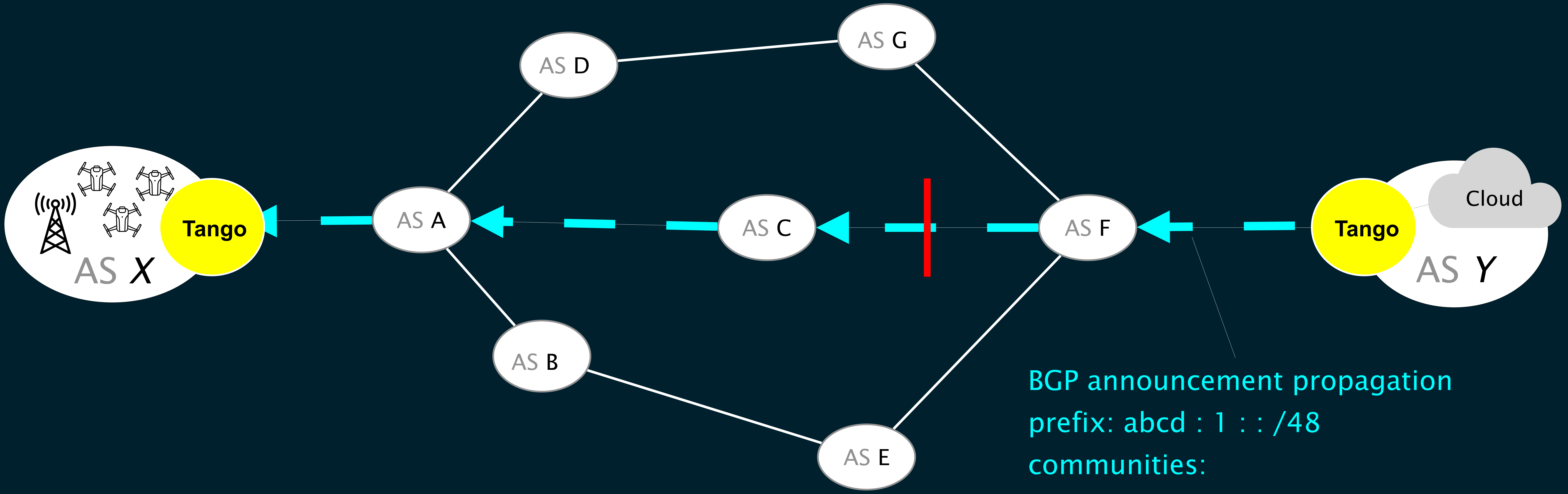

AS D
AS G
Tango
AS X
AS A
AS C
AS F
Tango
Cloud
AS Y
AS B
AS E

ASX only sees a single path, exported by its upstream AS



BGP announcement propagation
prefix: abcd : 1 : : /48

The Tango receiver advertises its IP prefix while suppressing the propagation of the default path



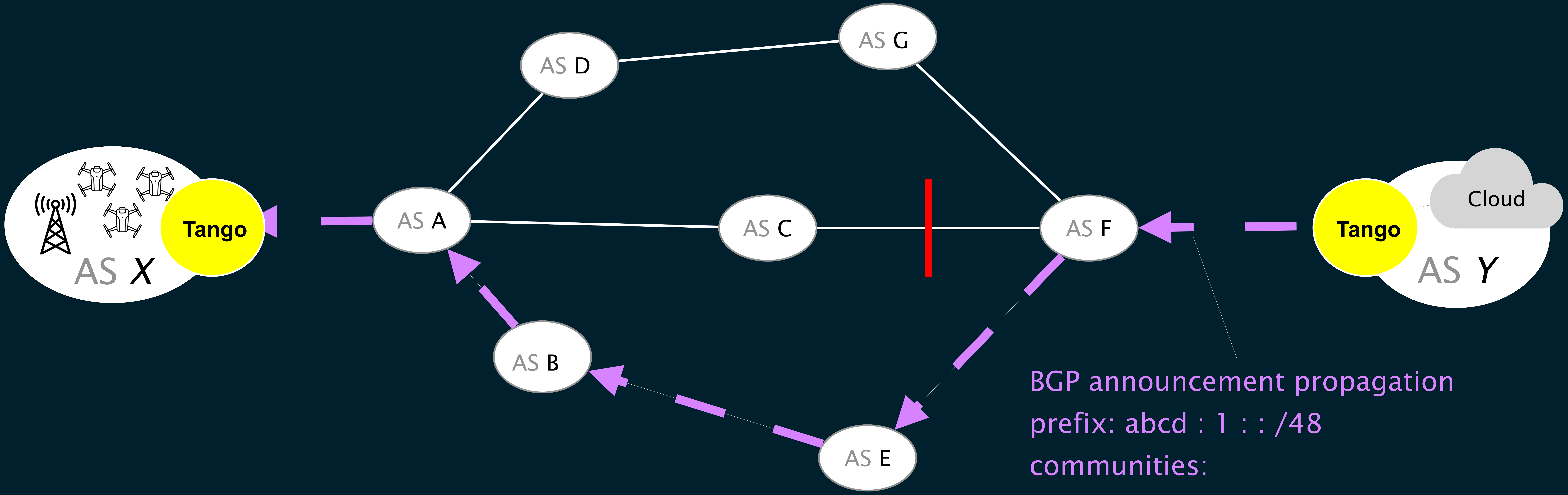
BGP announcement propagation

prefix: abcd : 1 : : /48

communities:

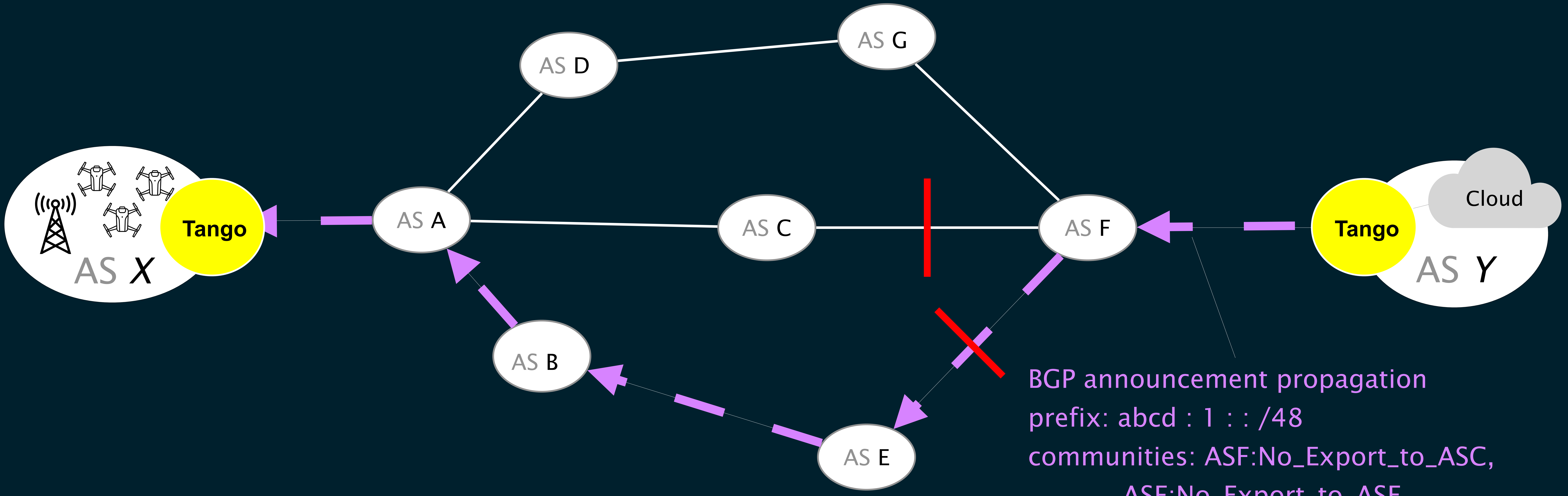
ASF:No_Export_to_ASC

The Tango receiver finds a new path through AS E



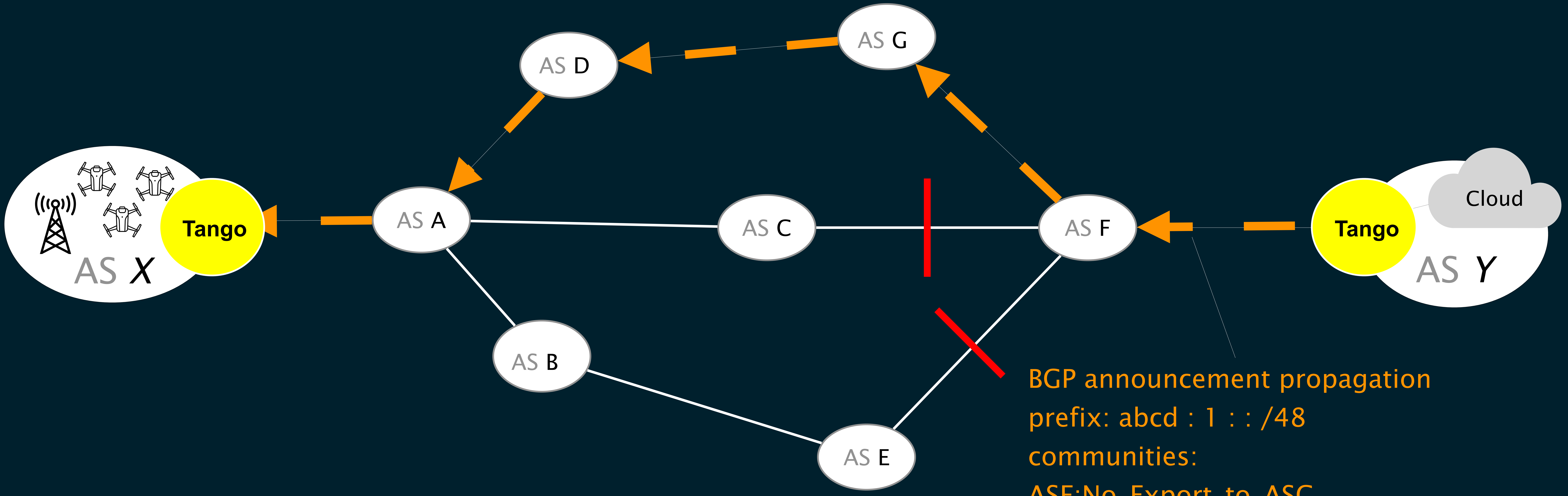
BGP announcement propagation
prefix: abcd : 1 : : /48
communities:
ASF:No_Export_to_ASC

The Tango receiver finds a new path through AS E which it will again suppress



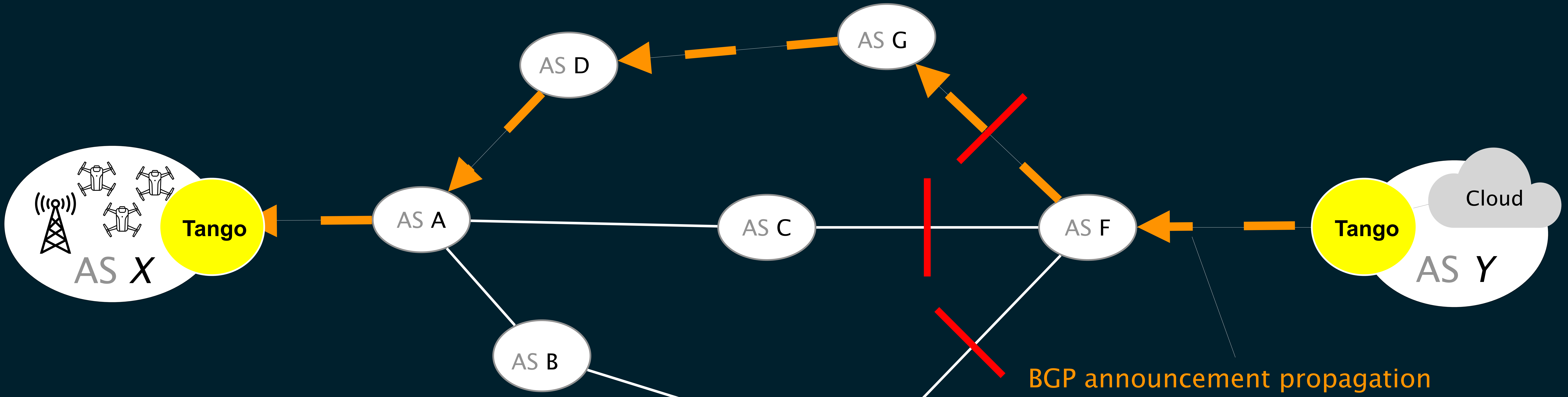
BGP announcement propagation
prefix: abcd : 1 : : /48
communities: ASF:No_Export_to_ASC,
ASF:No_Export_to_ASE

The Tango receiver finds a new path through AS E which it will again suppress to find yet another path through AS G



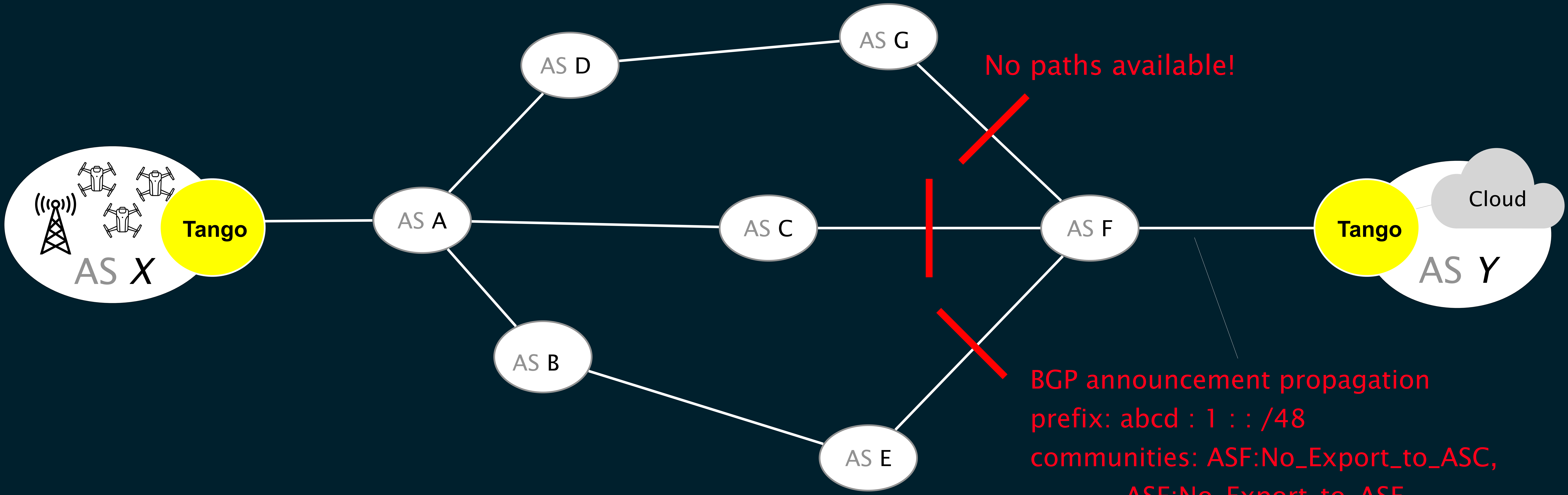
BGP announcement propagation
prefix: abcd : 1 : : /48
communities:
ASF:No_Export_to_ASC,
ASF:No_Export_to_ASE

The Tango receiver stops when there are no new paths



BGP announcement propagation
prefix: abcd : 1 : : /48
communities:
ASF:No_Export_to_ASC,
ASF:No_Export_to_ASE,
ASF:No_Export_to_ASG

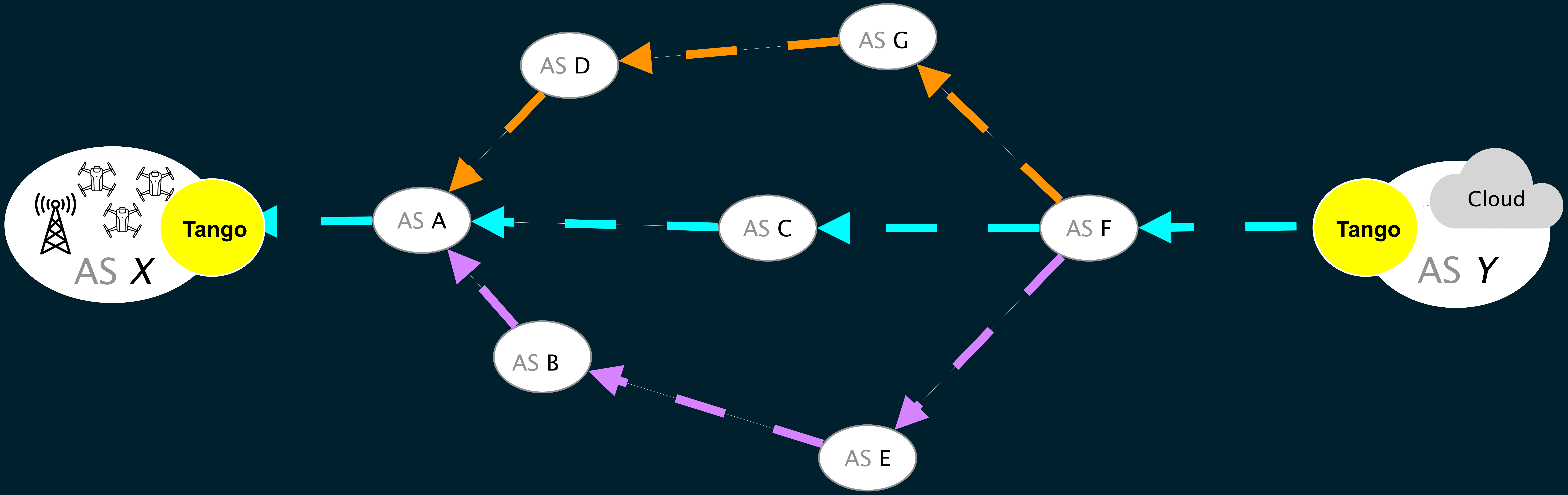
The Tango receiver stops when there are no new paths



No paths available!

BGP announcement propagation
prefix: abcd : 1 : : /48
communities: ASF:No_Export_to_ASC,
ASF:No_Export_to_ASE,
ASF:No_Export_to_ASG

AS Y announces different IP prefixes along different paths



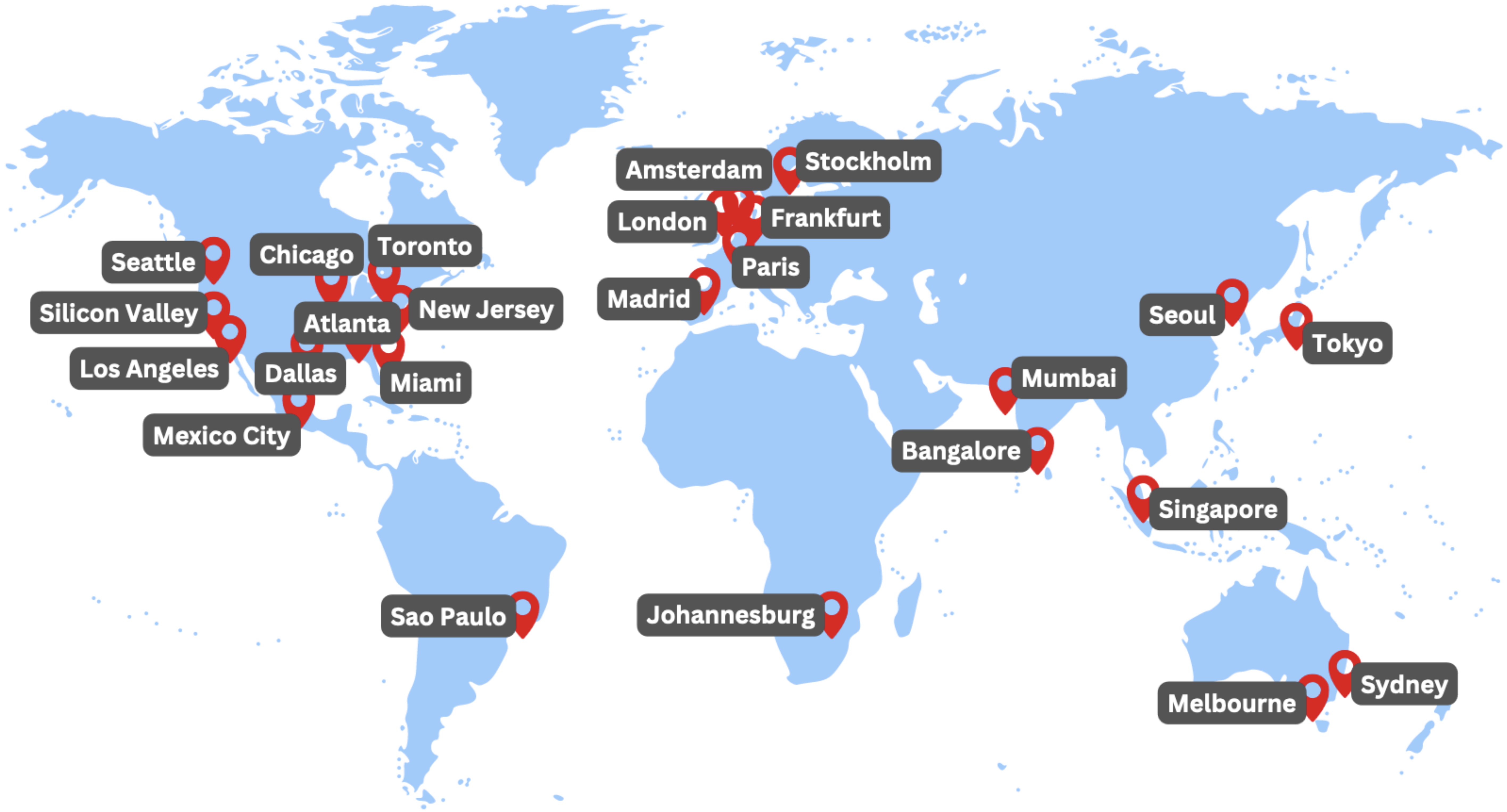
BGP
BGP
ROUTER

Global testbed

Run Tango- Pathfinder from 23 nodes hosted by Vultr to exposes Internet paths

Routed traffic over the exposed and default paths to two destinations: LA and Stockholm

Collected latency and loss measurements every 10ms, over roughly 32 hours



Tango-paths outperform the default path

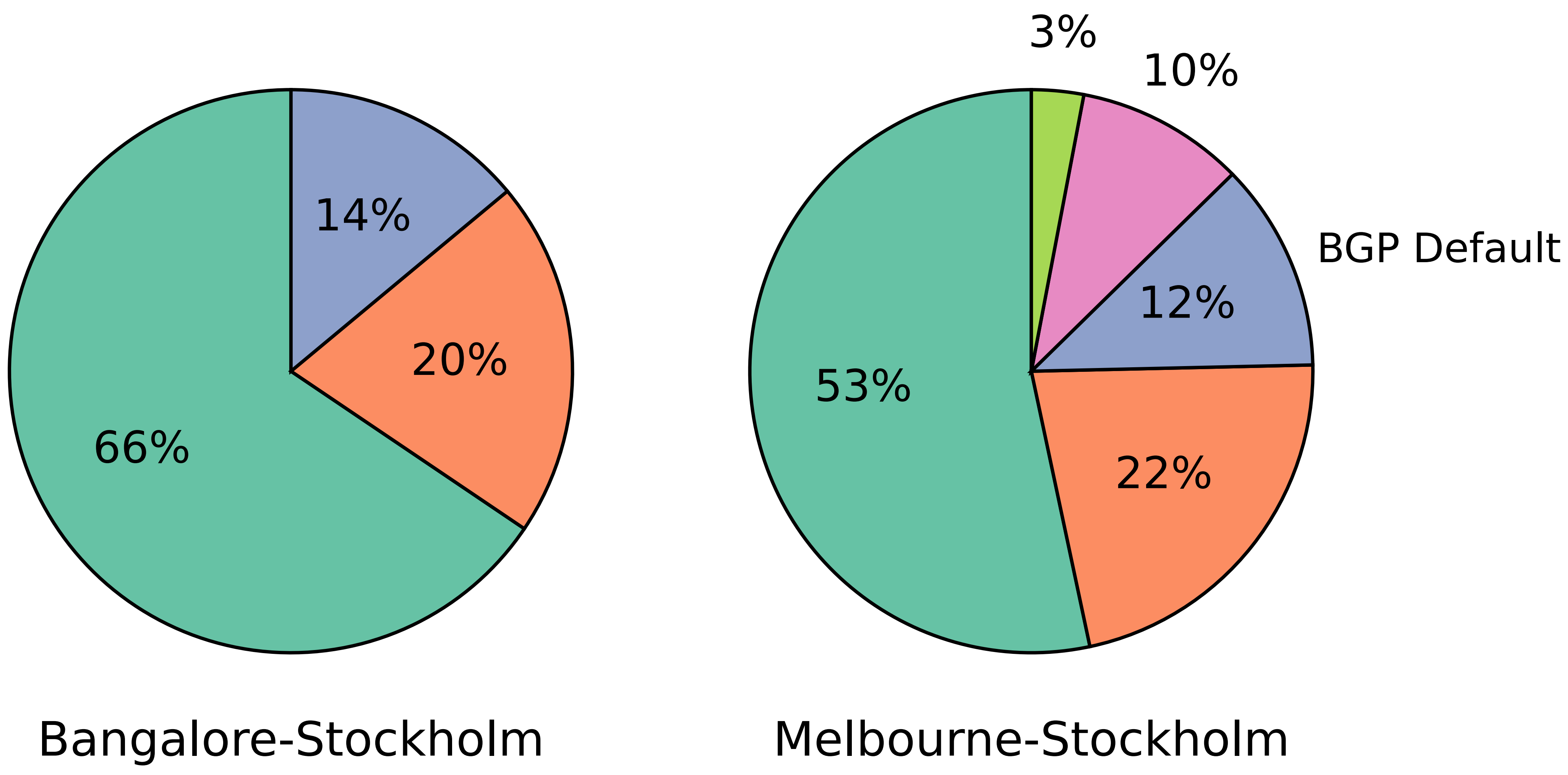
Across 23 measured pairs, 20 pairs had alternative paths that outperformed the default:

100% of the time for 15 pairs
75-88% of the time for 5 pairs

Bangalore-Stockholm: BGP default beaten by alternative paths 100% of the time

Melbourne-Stockholm: BGP default beaten by alternative paths 88% of the time

Breakdown of best paths for two pairs



BGP
BGP
ROUTER

Tango’s design requirements for performance-driven routing

Route Control

Tango senders need to control which path traffic will use.

Accurate Measurements

Measurements should not be affected by irrelevant conditions e.g., slow receivers, Wi-Fi.

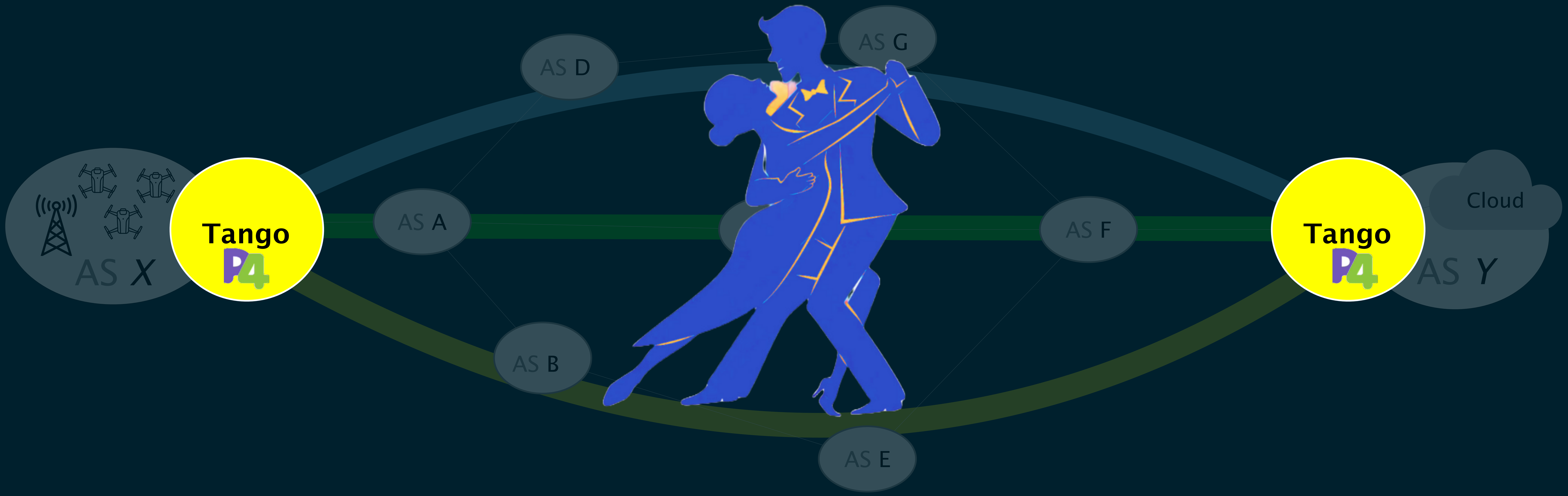
Trustworthy Measurements

An on-path attacker should not be able to distort measurements to their advantage.

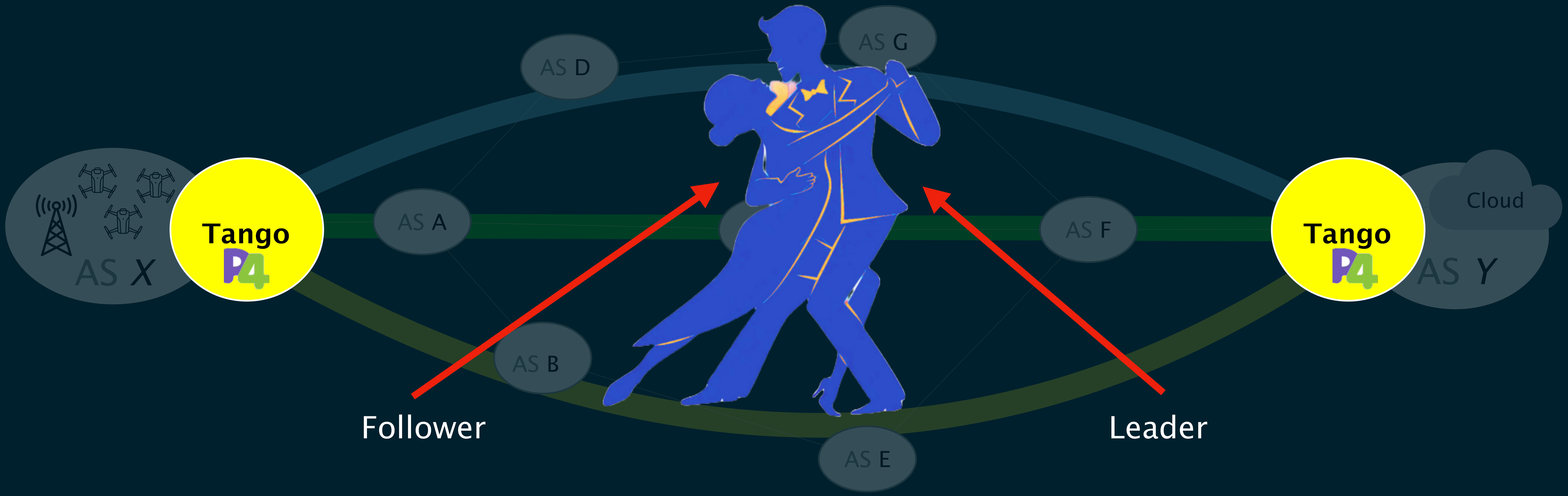
Dynamic & Secure Rerouting

Tango should allow dynamic performance-driven and safe reroutes.

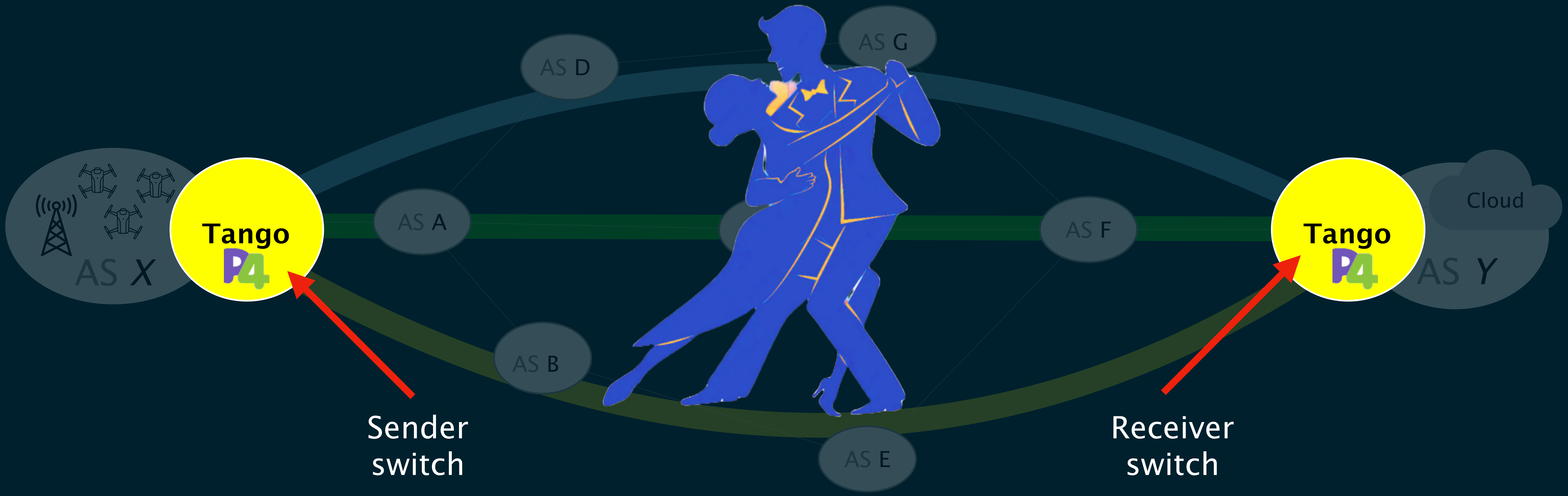
Tango solves these challenges with P4 and co-operation



Tango solves these challenges with P4 and co-operation



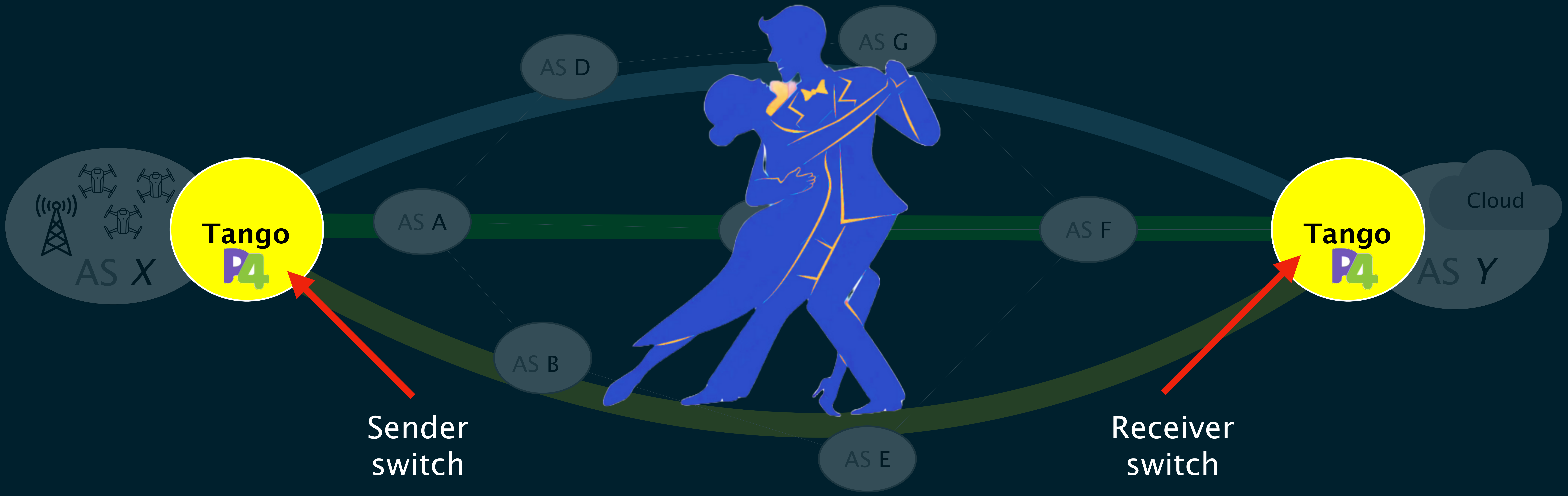
Tango solves these challenges with P4 and co-operation



Sender switch

Receiver switch

Tango solves these challenges with P4 and co-operation



The sender switch performs the move that the receiver has signaled.

Tango’s design requirements for performance-driven routing

Route Control

Tango senders need to control which path traffic will use.

Accurate Measurements

Measurements should not be affected by irrelevant conditions e.g., slow receivers, Wi–Fi.

Trustworthy Measurements

An on–path attacker should not be able to distort measurement to their advantage.

Dynamic & Secure Rerouting

Tango should allow dynamic performance–driven and safe reroutes.

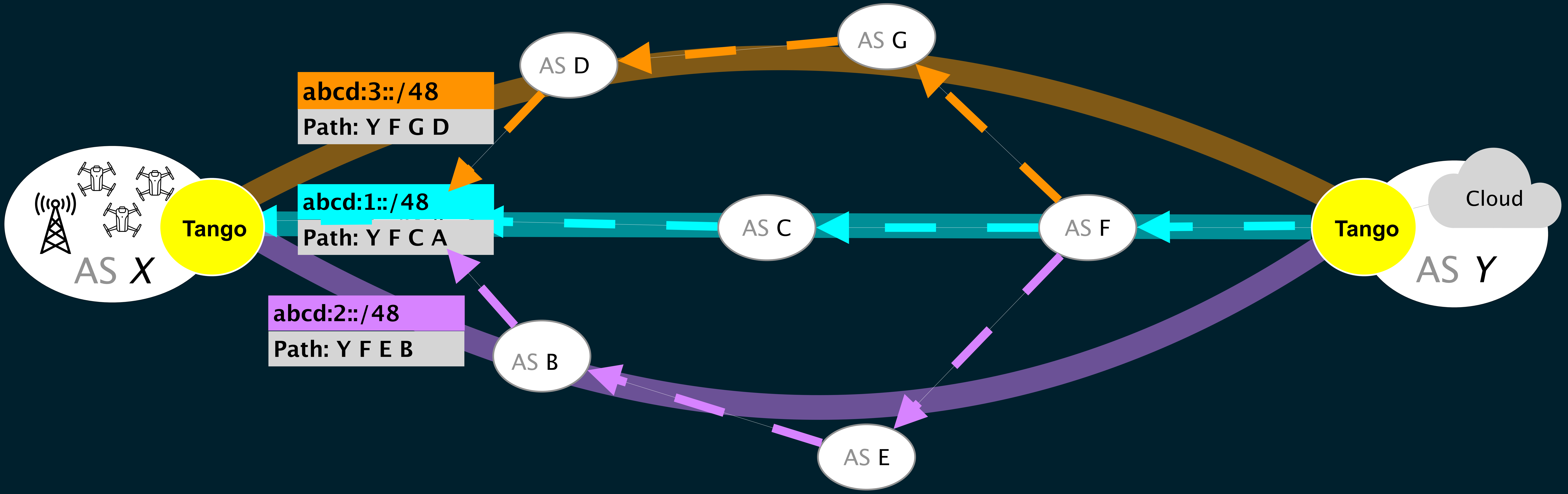
AS Y announces different IP prefixes along different paths

AS G
AS D
abcd:1::/48
Path: Y
Cloud
Tango
AS Y
abcd:1::/48
Path: Y F C A
Tango
AS X
AS A
AS C
AS F
AS B
AS E

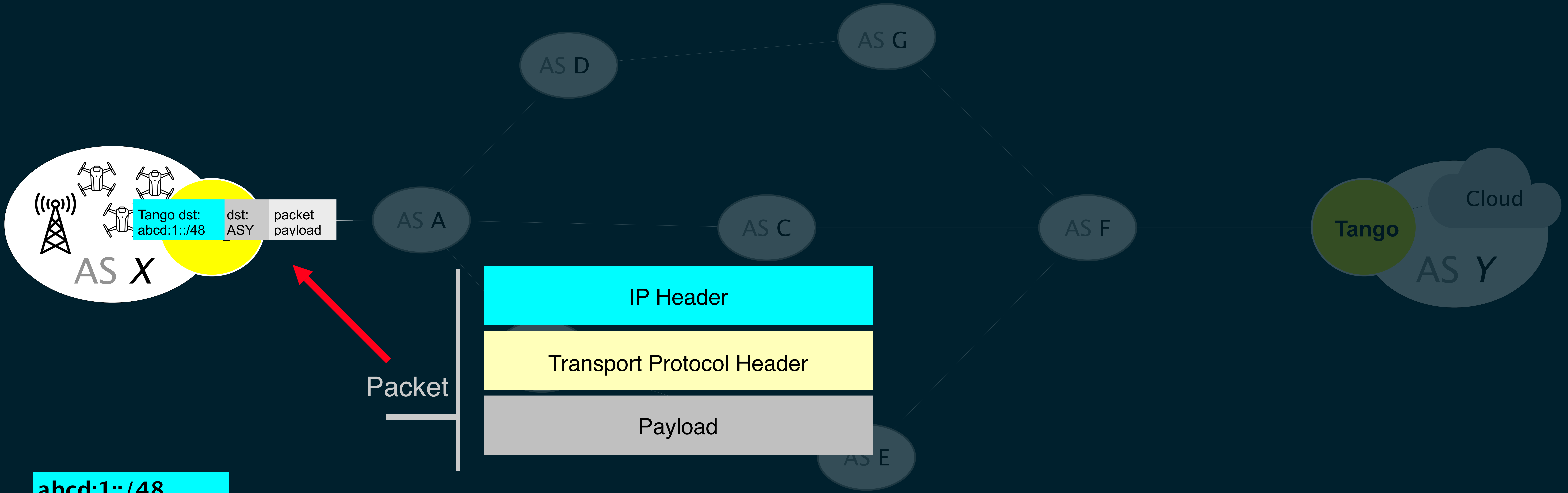
AS Y announces different IP prefixes along different paths

AS G
AS D
abcd:3::/48
Path: Y
abcd:1::/48
Path: Y
Cloud
abcd:1::/48
Path: Y F C A
Tango
AS A
AS C
AS F
Tango
AS X
AS Y
abcd:2::/48
Path: Y
AS B
AS E

AS Y announces different IP prefixes along different paths

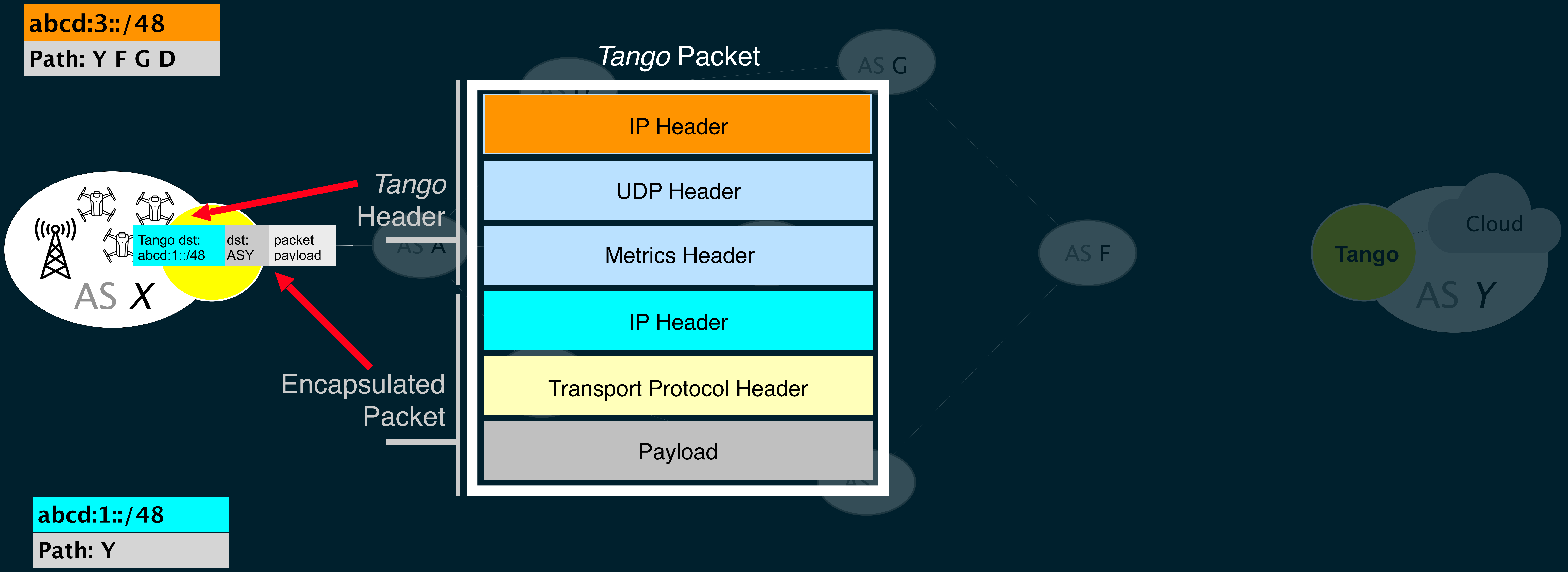


Upon reception of a packet,



abcd:1::/48
Path: Y

Upon reception of a packet, the sender encapsulates it with a destination within the prefix that correspond to the path of choice



abcd:3::/48
Path: Y F G D

Tango Packet

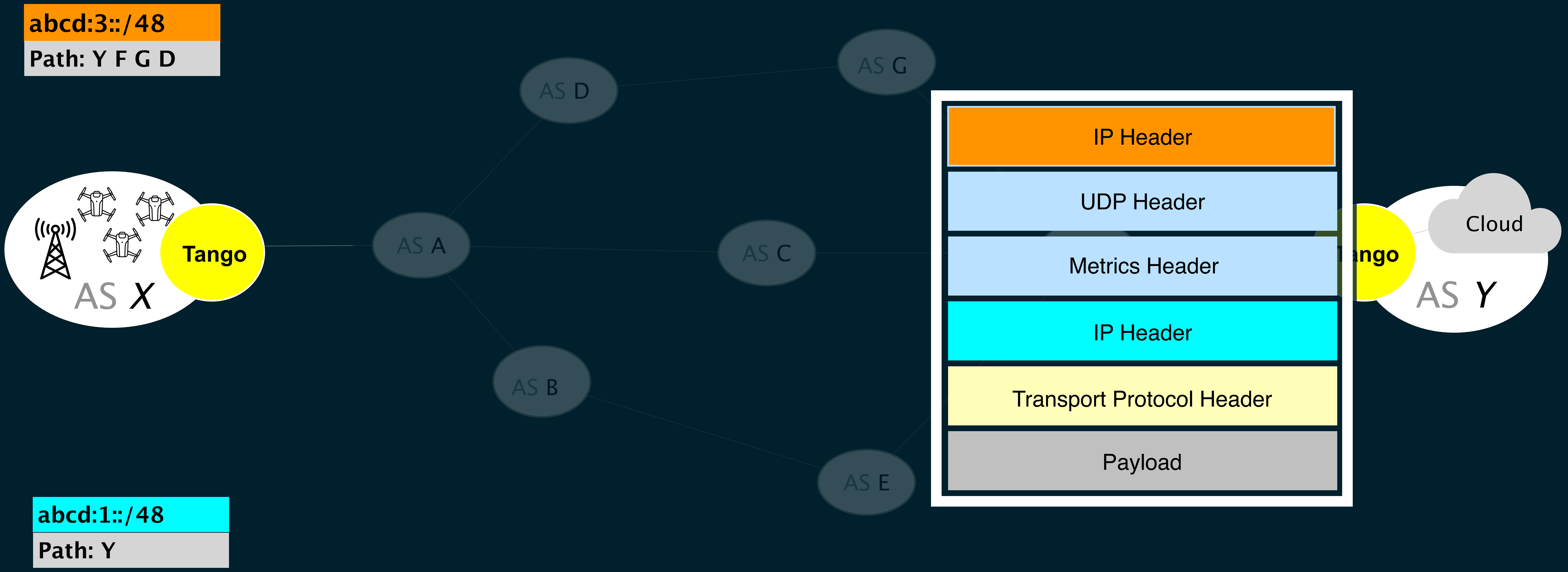
- IP Header
- UDP Header
- Metrics Header
- IP Header
- Transport Protocol Header
- Payload

Tango Header

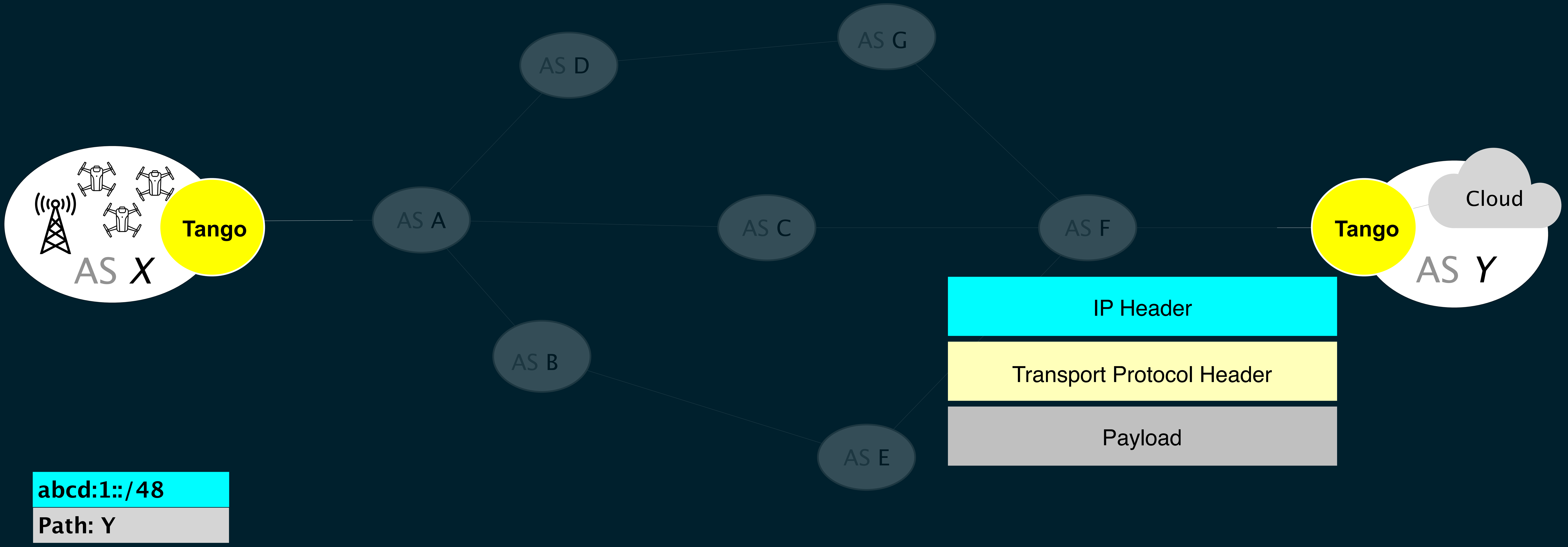
Encapsulated Packet

abcd:1::/48
Path: Y

The receiver decapsulates packets before letting them reach their destination



The receiver decapsulates packets before letting them reach their destination



AS D

AS G

AS A

AS C

AS F

AS B

AS E

Tango

AS *X*

Tango

Cloud

AS *Y*

IP Header

Transport Protocol Header

Payload

abcd:1::/48

Path: Y

Tango's design requirements for performance-driven routing

Route Control

Tango senders need to control which path traffic will use.

Accurate Measurements

Measurements should not be affected by irrelevant conditions e.g., slow receivers, Wi-Fi.

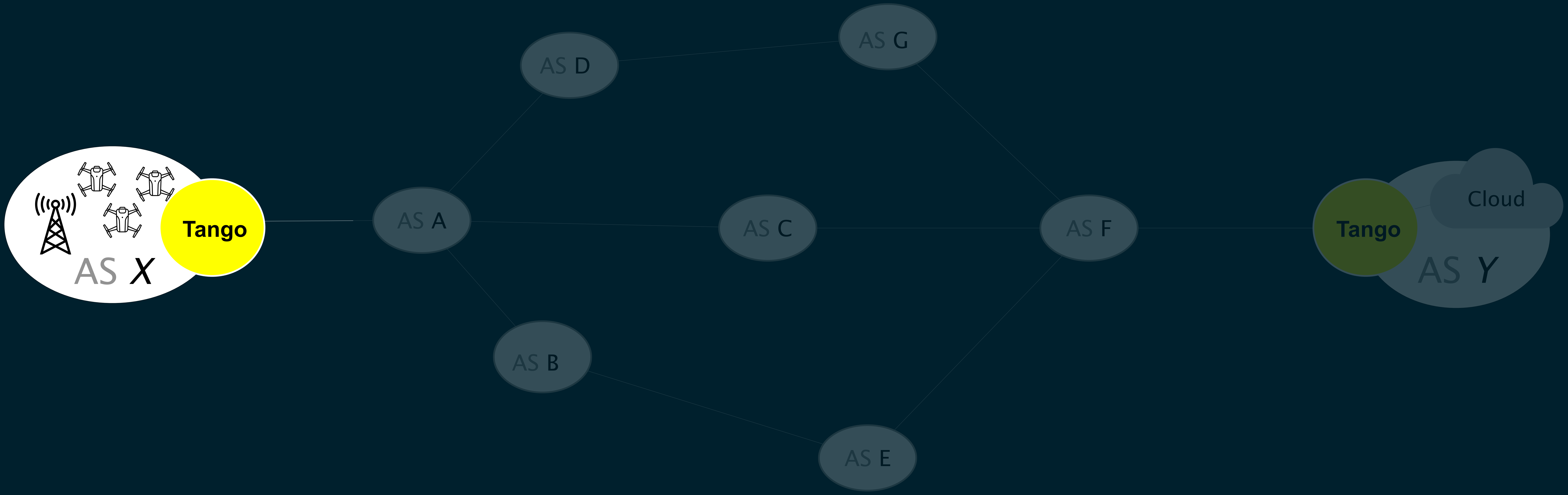
Trustworthy Measurements

An on-path attacker should not be able to distort measurements to their advantage.

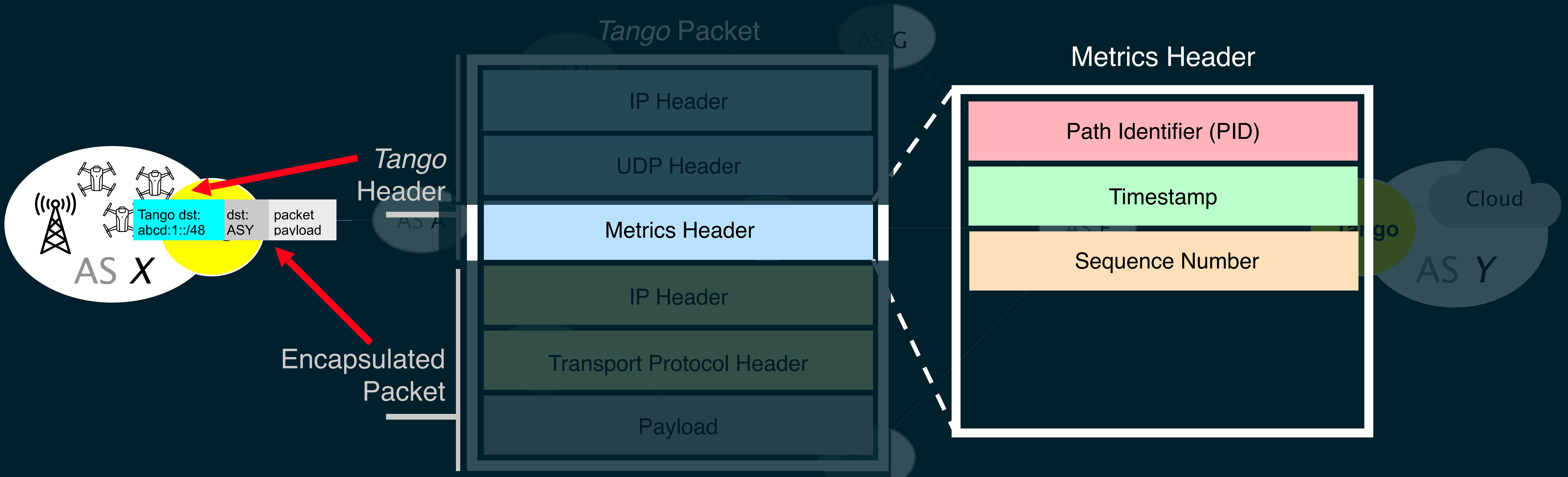
Dynamic & Secure Rerouting

Tango should allow dynamic performance-driven and safe reroutes.

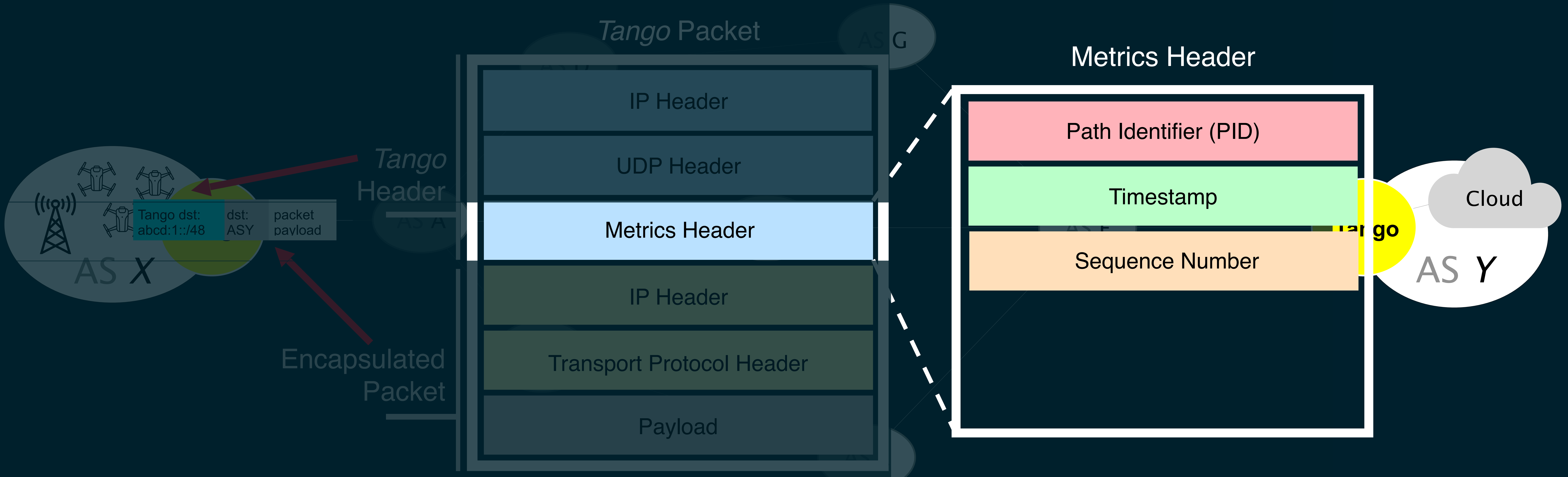
Conventional active round-trip measurements are inaccurate and can be easily manipulated



The sender includes the timestamp of each packet’s departure, and a per-path sequence number in the Tango header.



The receiver calculates one-way latency and loss for each path avoiding the noise of the access networks



Tango's design requirements for performance-driven routing

Route Control

Tango senders need to control which path traffic will use.

Accurate Measurements

Measurements should not be affected by irrelevant conditions e.g., slow receivers, Wi-Fi.

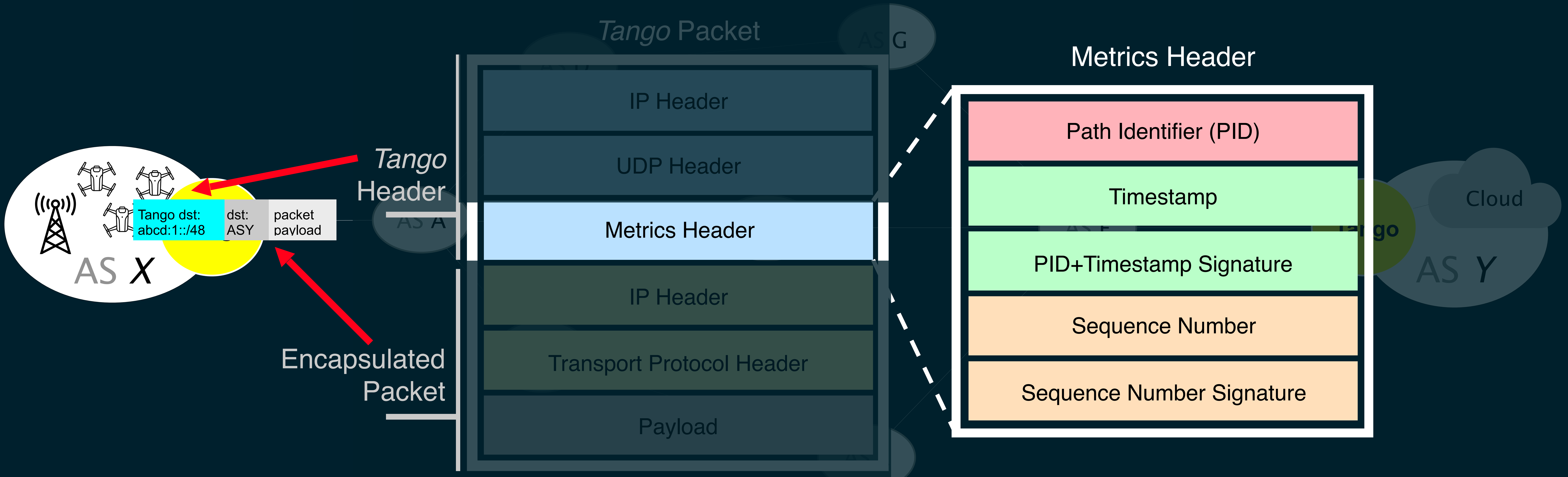
Trustworthy Measurements

An on-path attacker should not be able to distort measurements to their advantage.

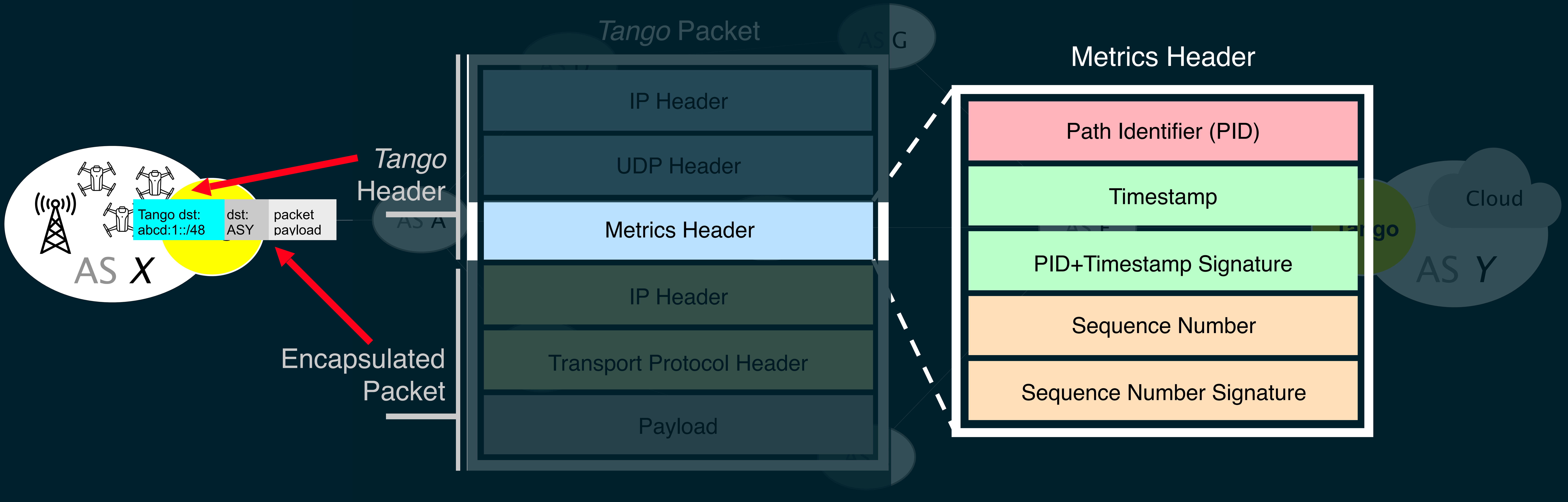
Dynamic & Secure Rerouting

Tango should allow dynamic performance-driven and safe reroutes.

Tango sender adds a path-specific signature to each ms timestamp, an attacker cannot manipulate or replay it to affect latency measurements



Tango sender adds one bit signature to each sequence number, an on-path attacker would need to guess multiple to affect loss rate



Tango’s design requirements for performance-driven routing

Route Control

Tango senders need to control which path traffic will use.

Accurate Measurements

Measurements should not be affected by irrelevant conditions e.g., slow receivers, Wi-Fi.

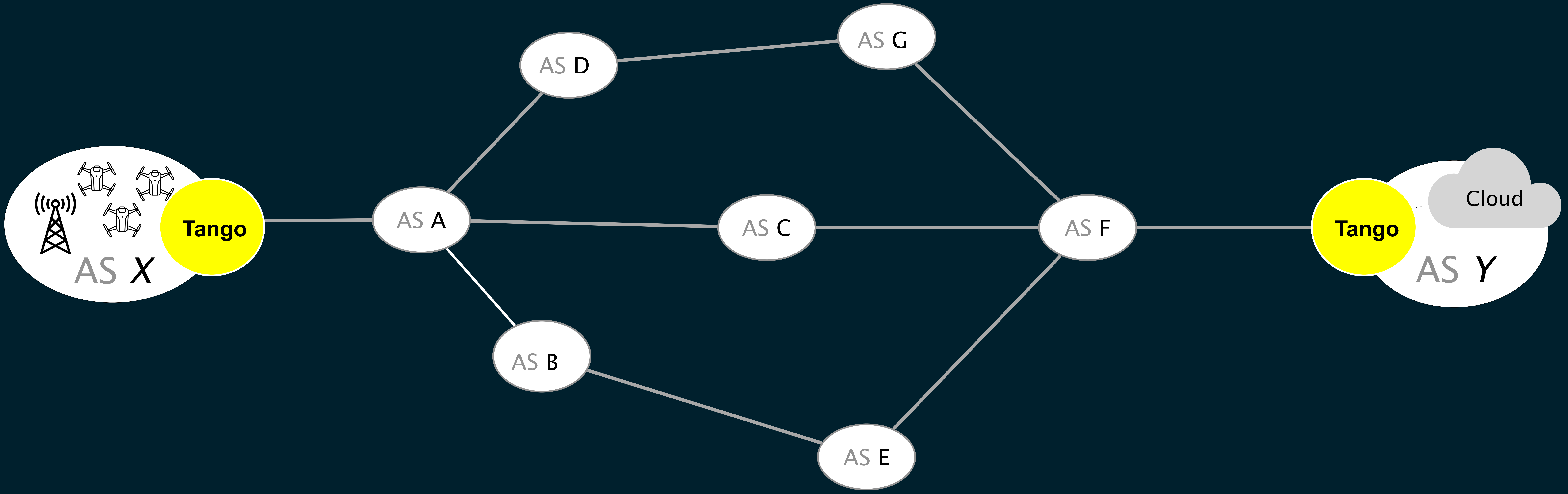
Trustworthy Measurements

An on-path attacker should not be able to distort measurements to their advantage.

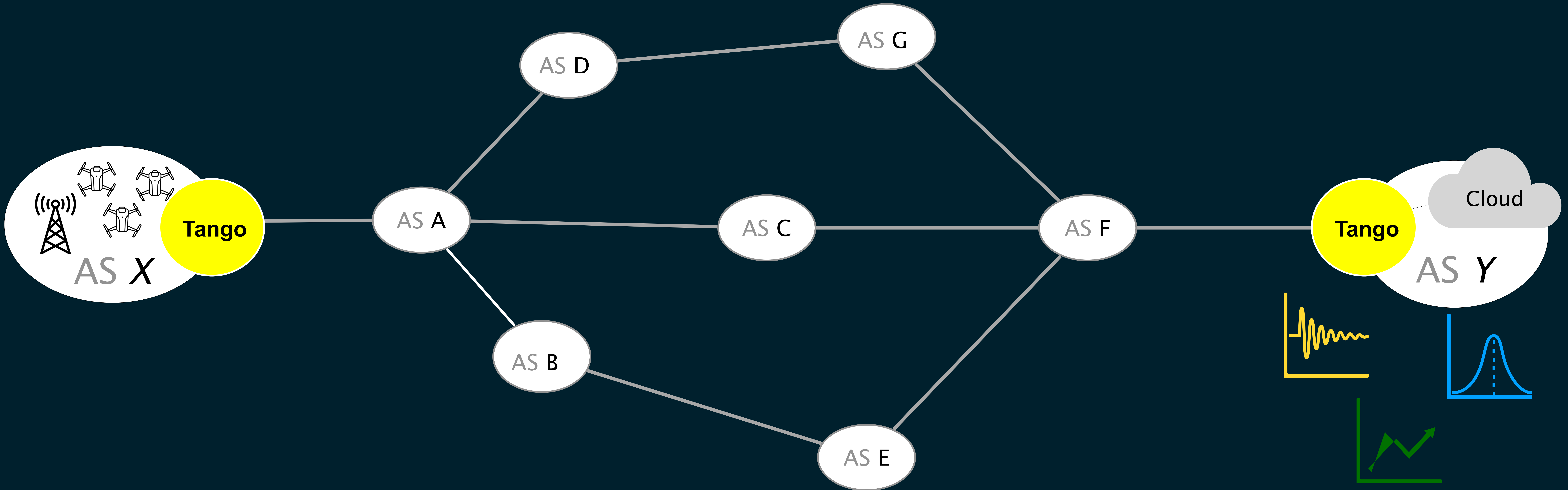
Dynamic & Secure Rerouting

Tango should allow dynamic performance-driven and safe reroutes.

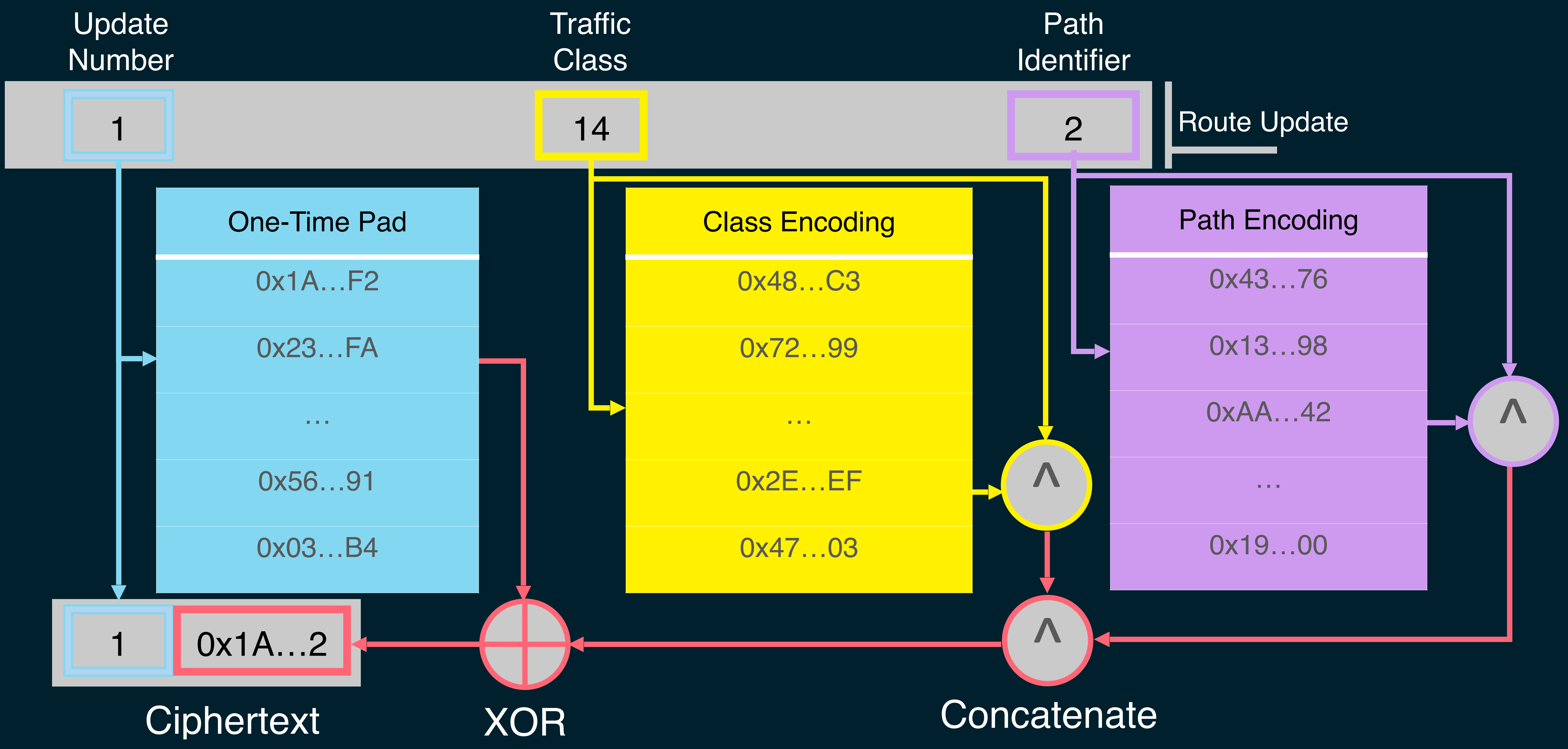
The Tango sender selects paths,



The Tango sender selects paths, but the Tango receiver collected the measurements



Tango protects reroute commands with one-time-pad



One-Time Pad
0x1A…F2
0x23…FA
…
0x56…91
0x03…B4

Class Encoding
0x48…C3
0x72…99
…
0x2E…EF
0x47…03

Path Encoding
0x43…76
0x13…98
0xAA…42
…
0x19…00

Internet

Princeton

Vultr -Stockholm

Cloud

Tango Sender

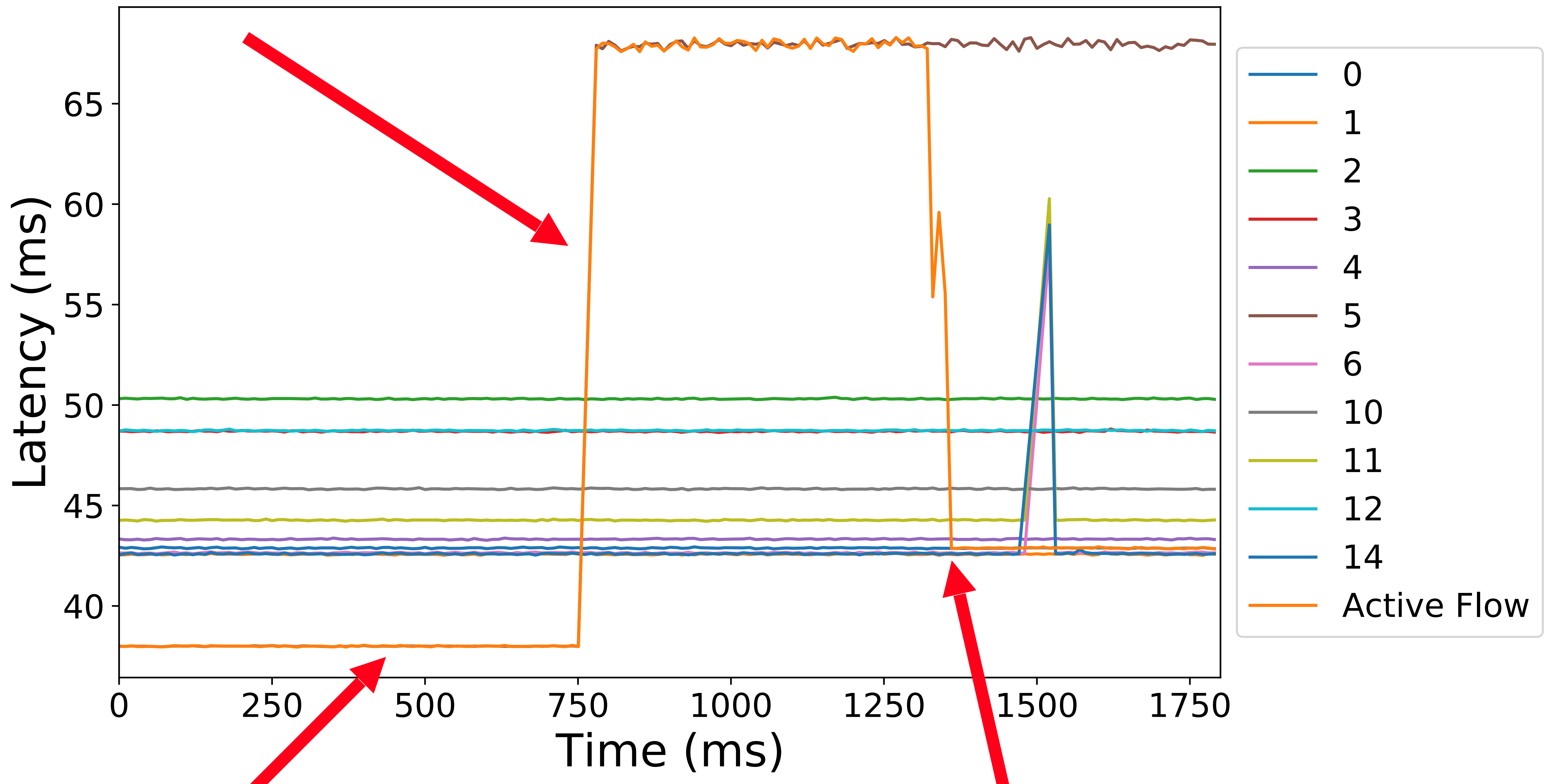
Tango Receiver

Real-world Testbed

We run Tango between Princeton and Stockholm!

Route update complete in <1s

delay spike



initial best path

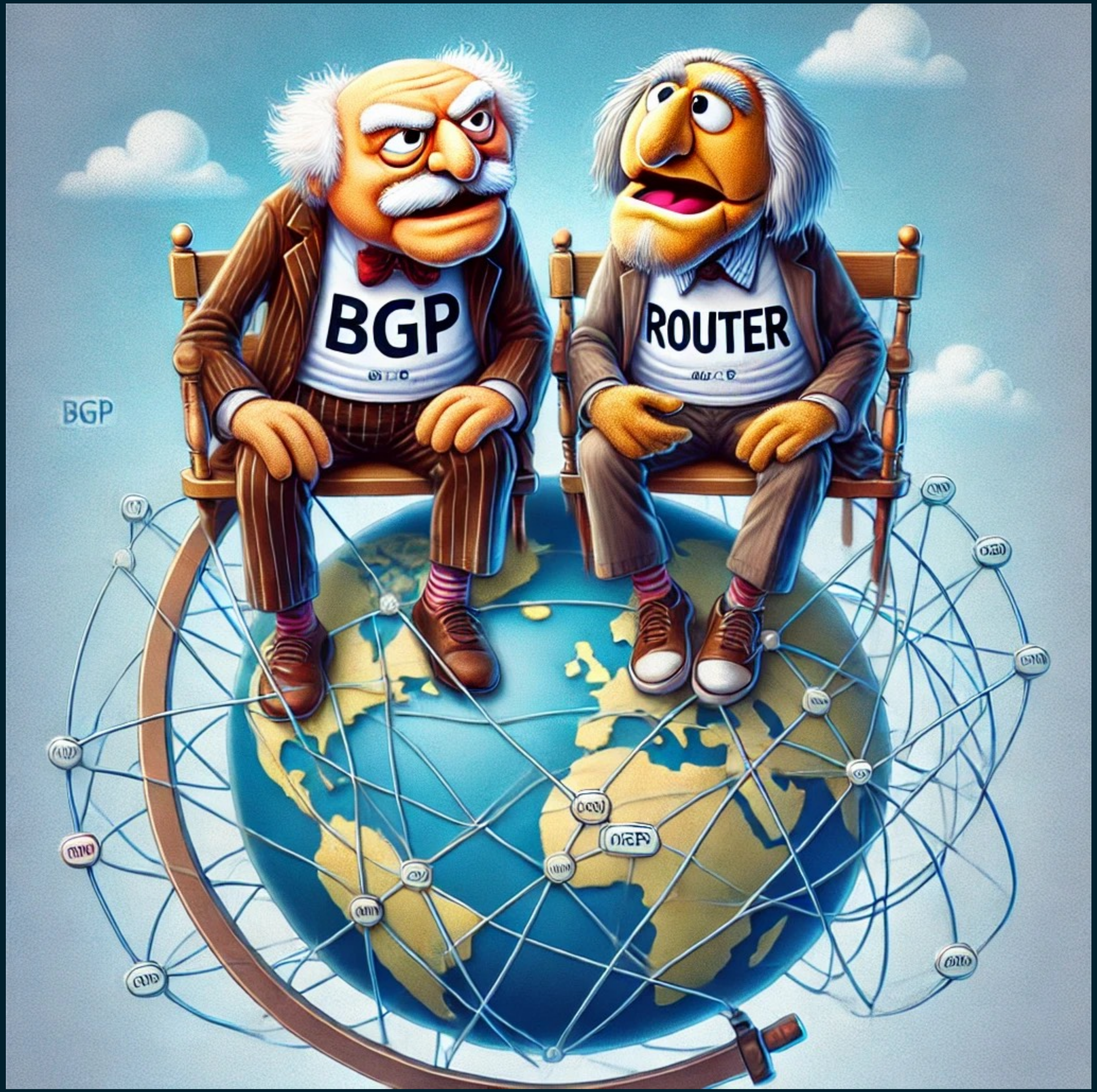
dynamically move to new best path

What can you do with a couple of programmable points in the Internet?

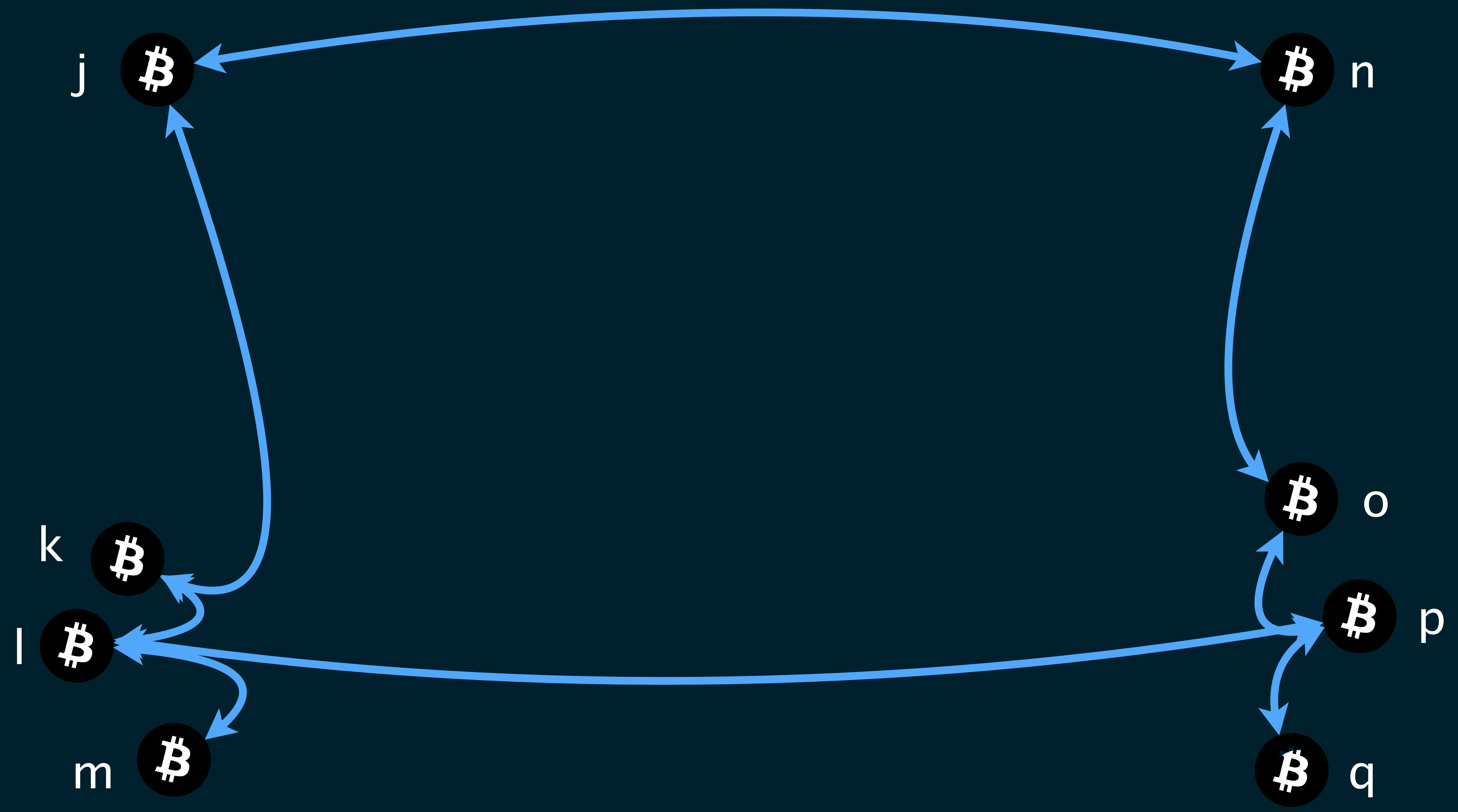
Tango: performance-driven routing system

SABRE: secure overlay for BTC block propagation

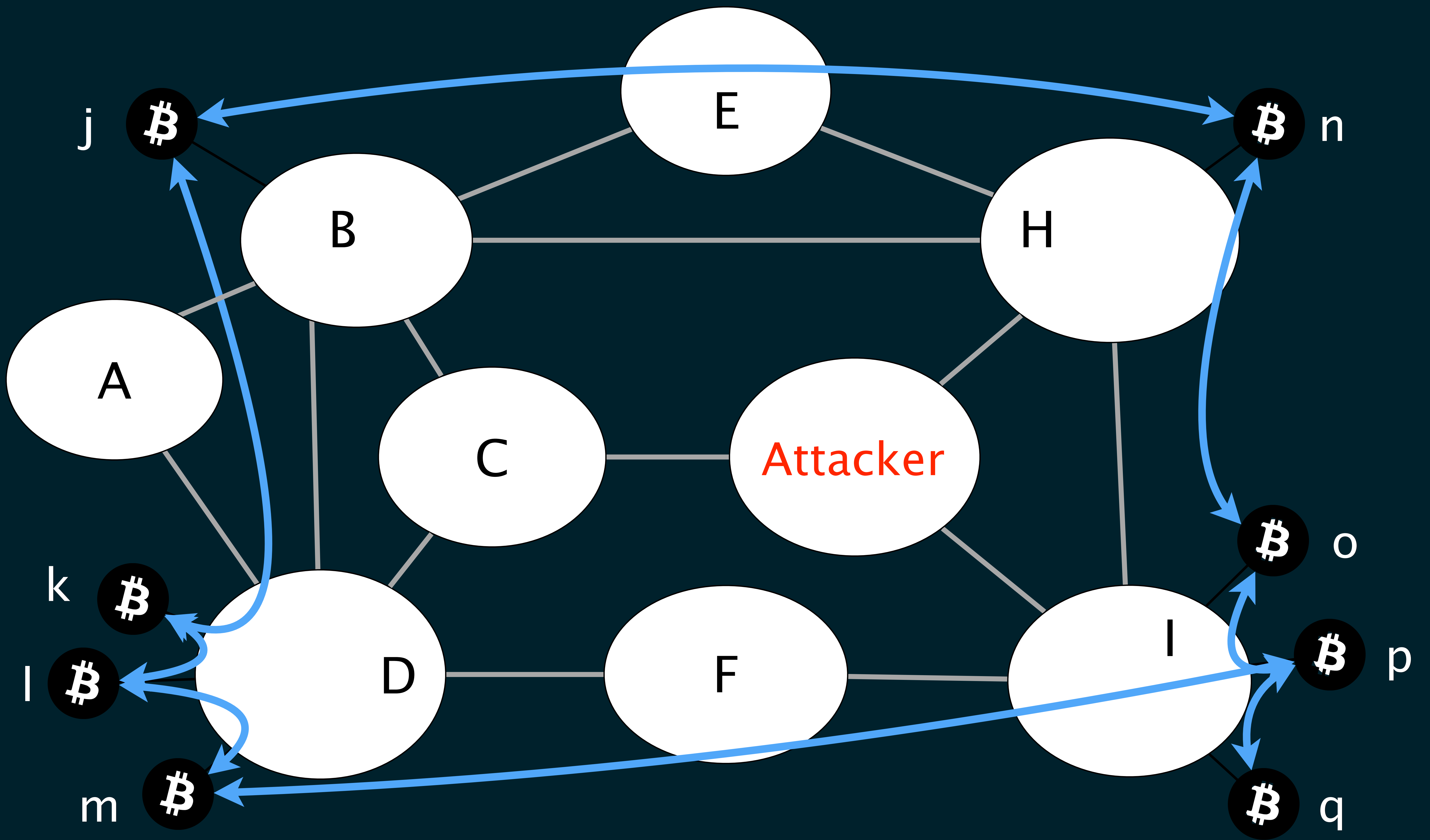
NDSS'19


BGP
ROUTER
BGP

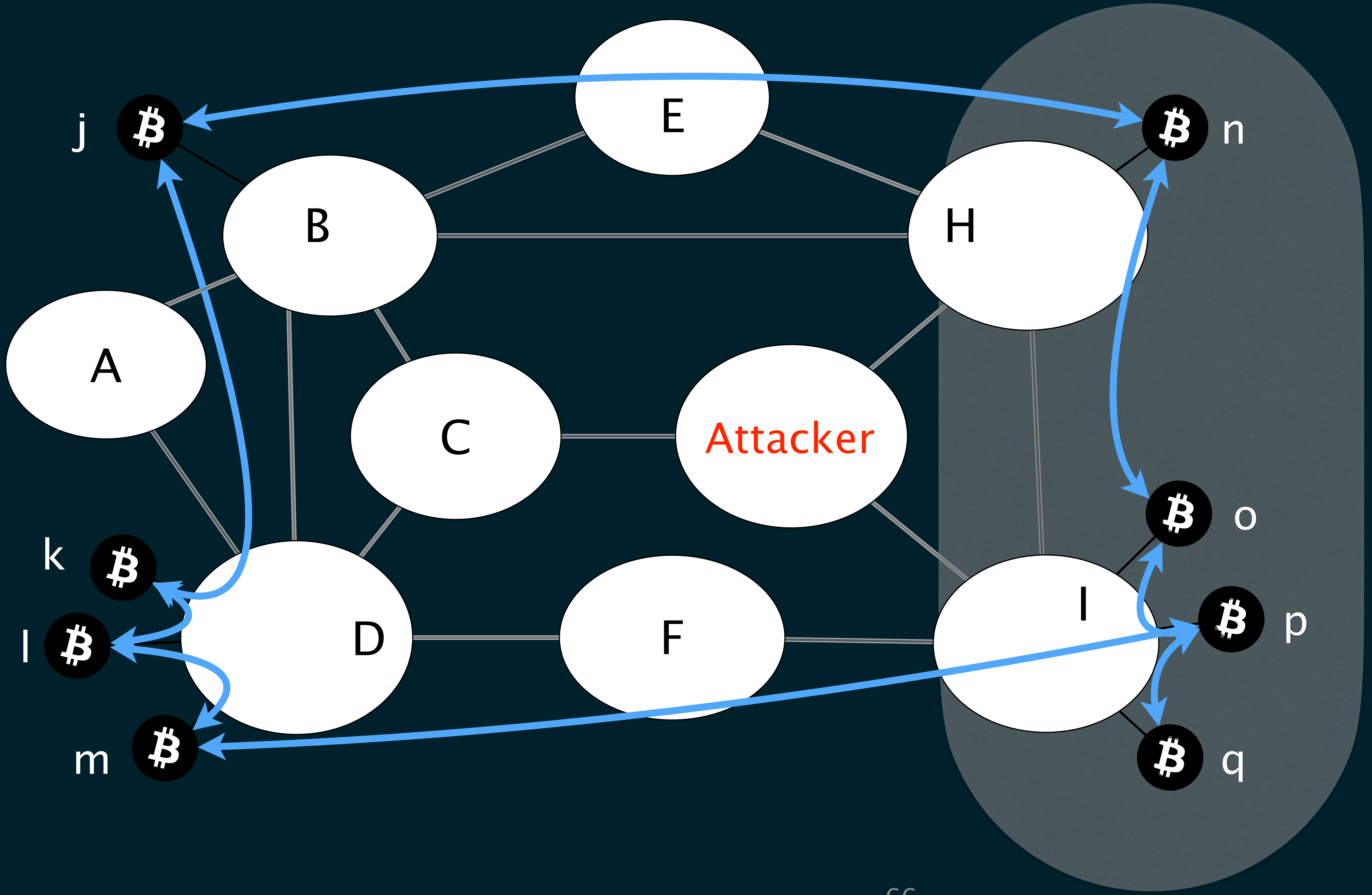
Bitcoin clients exchange Blocks
which contain the most recent transactions



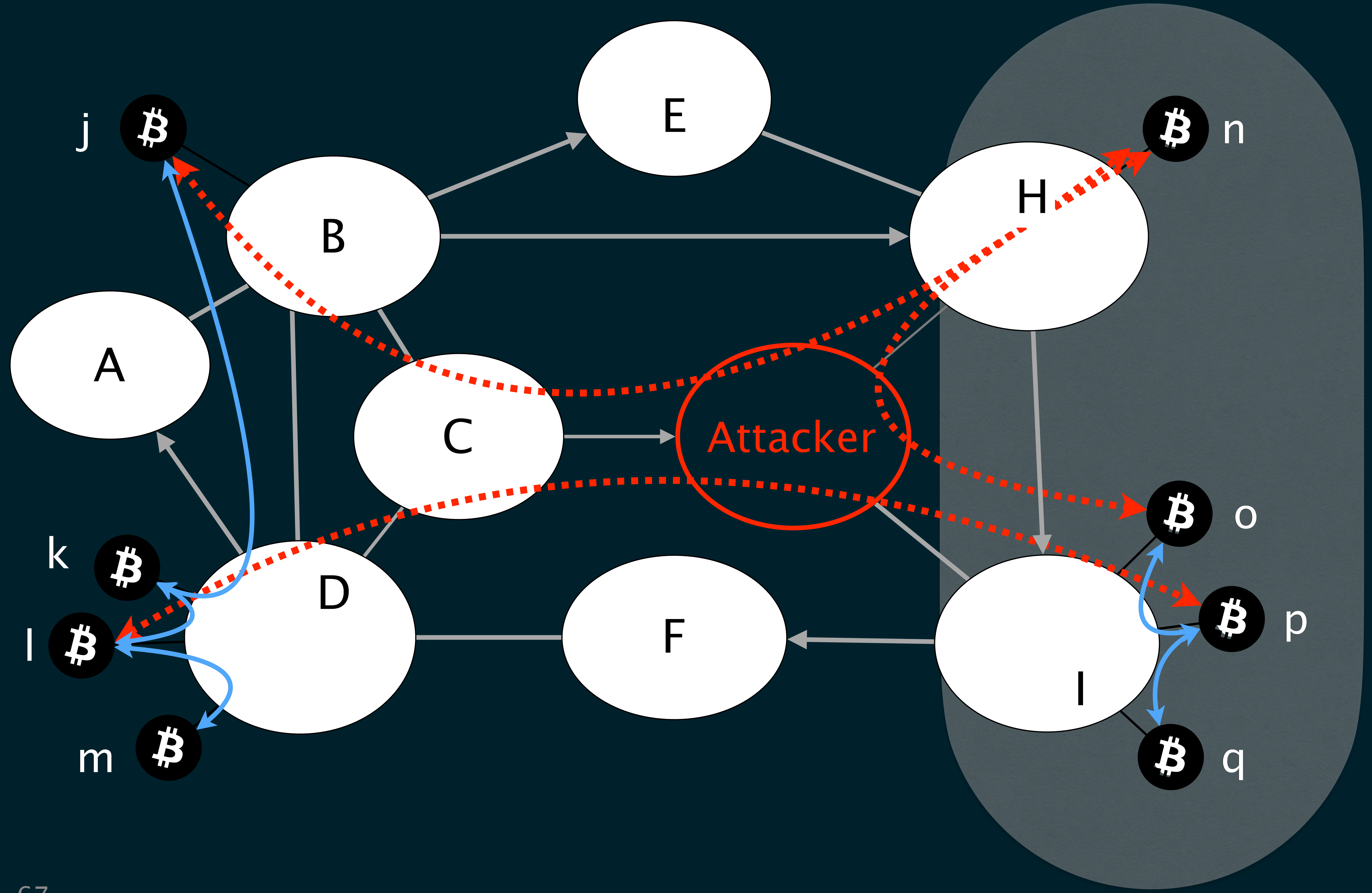
A malicious or compromised AS aims at isolating the grey zone



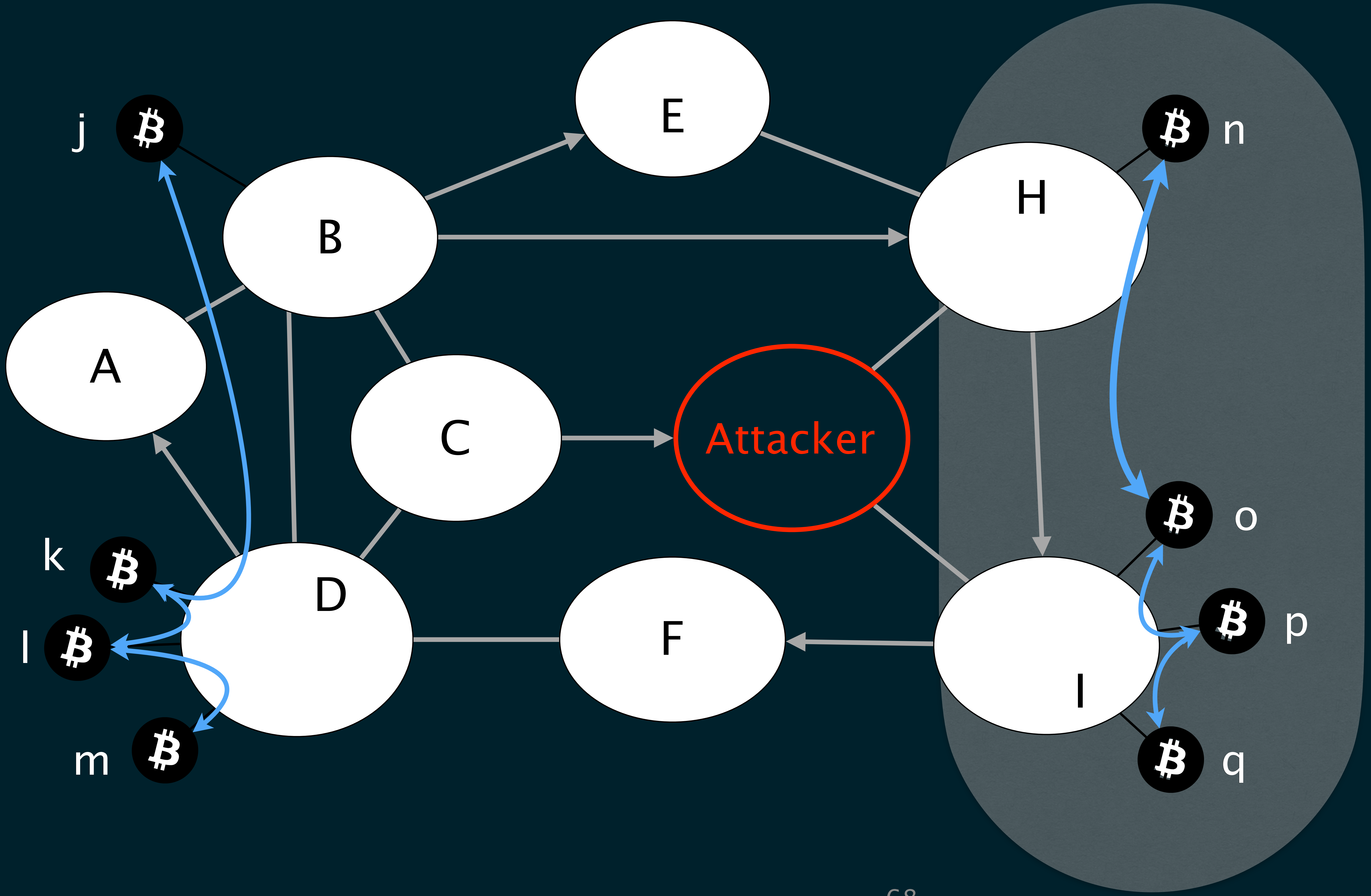
A malicious or compromised AS aims at isolating the grey zone



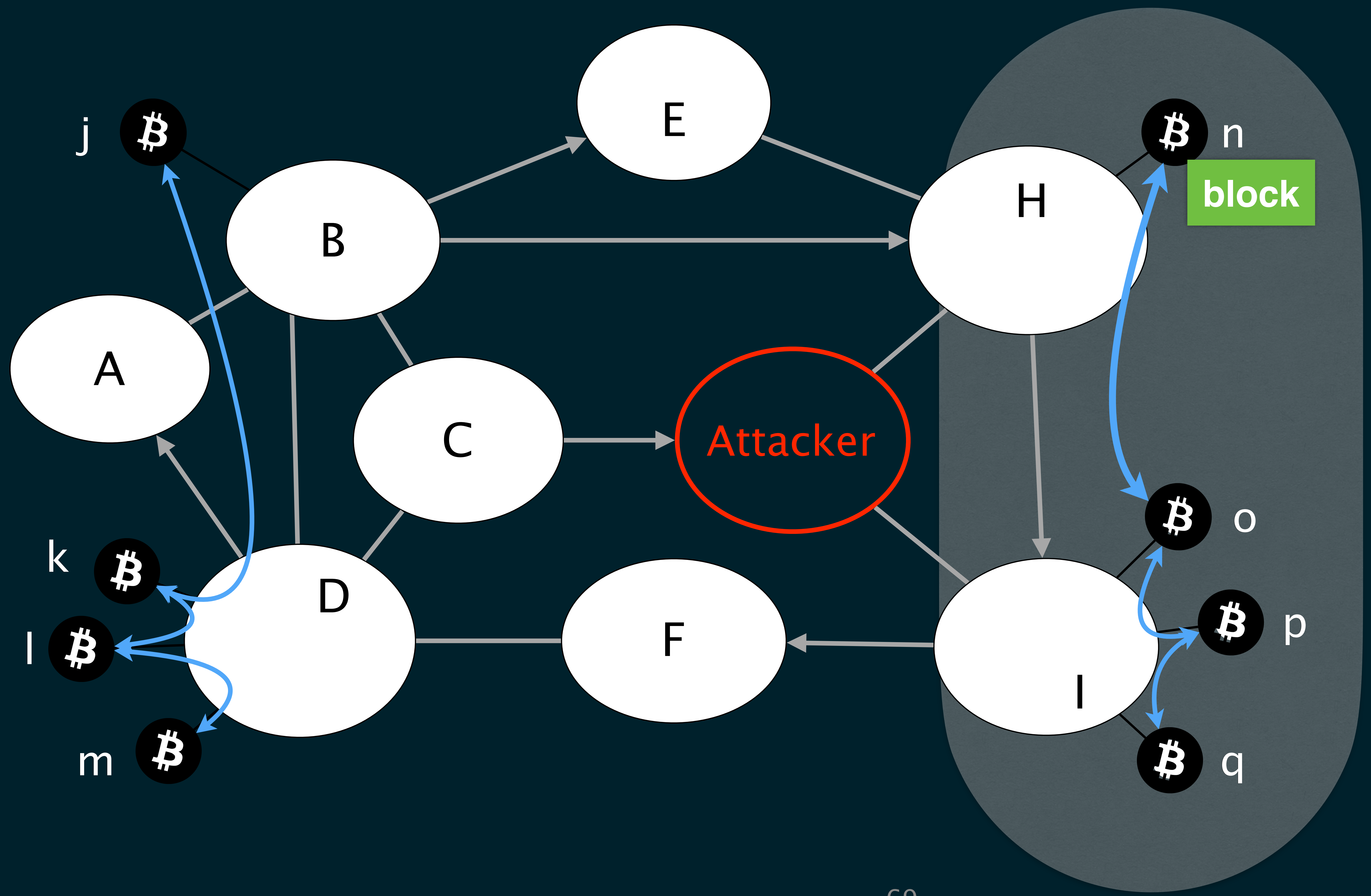
Attacker attracts connections with BGP hijacking



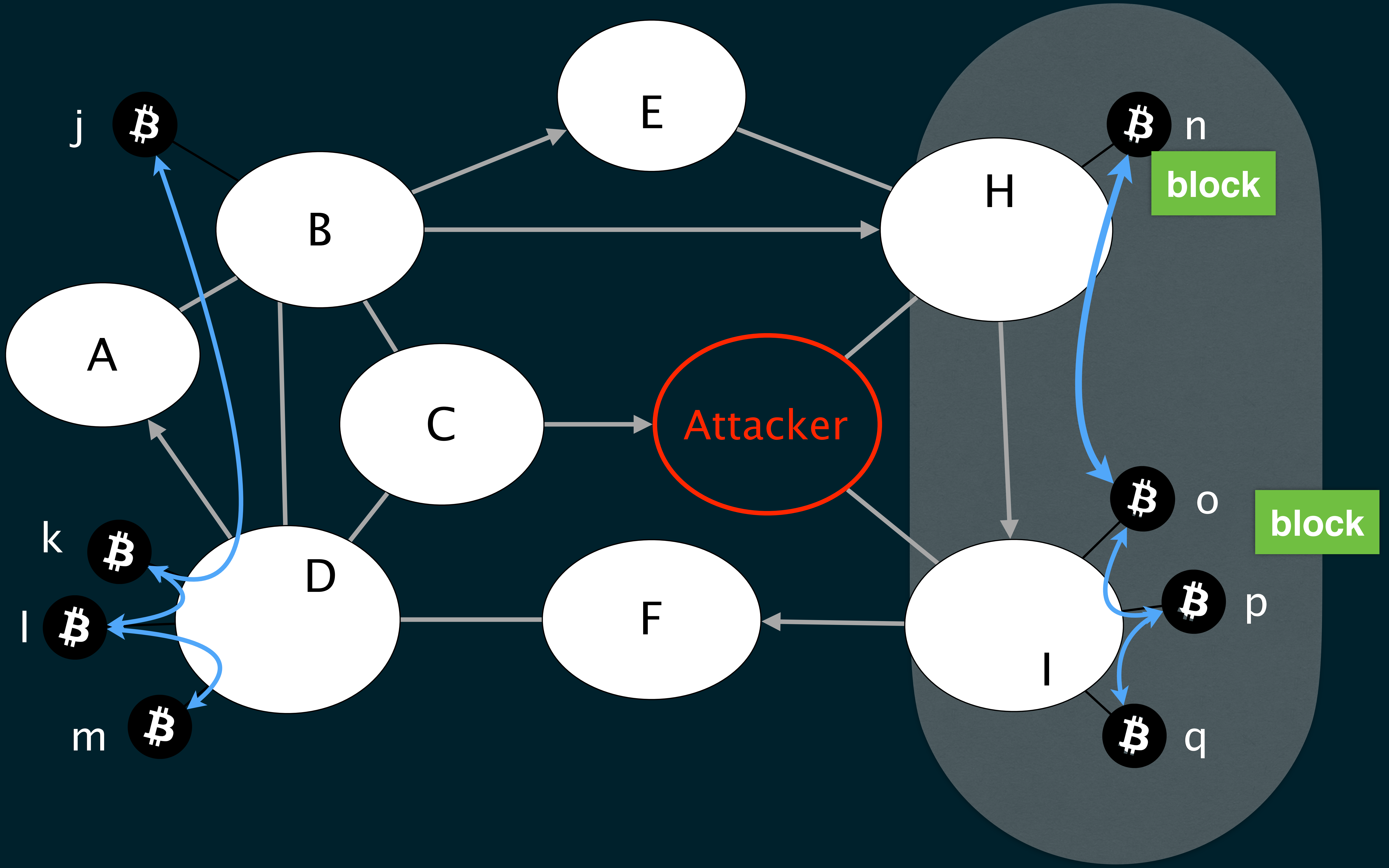
Attacker drops connections crossing the partition



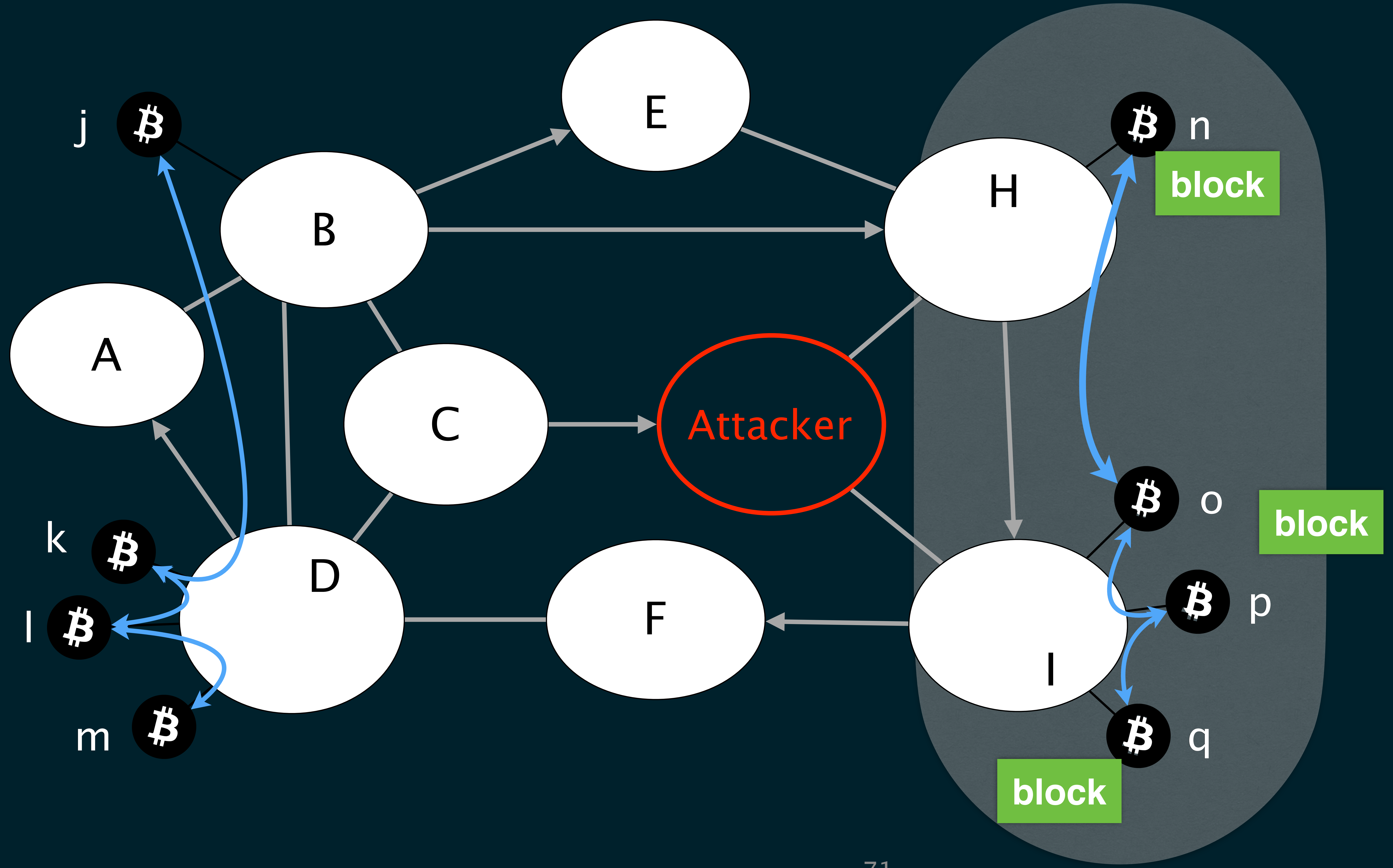
A new block in the grey zone cannot be propagated further



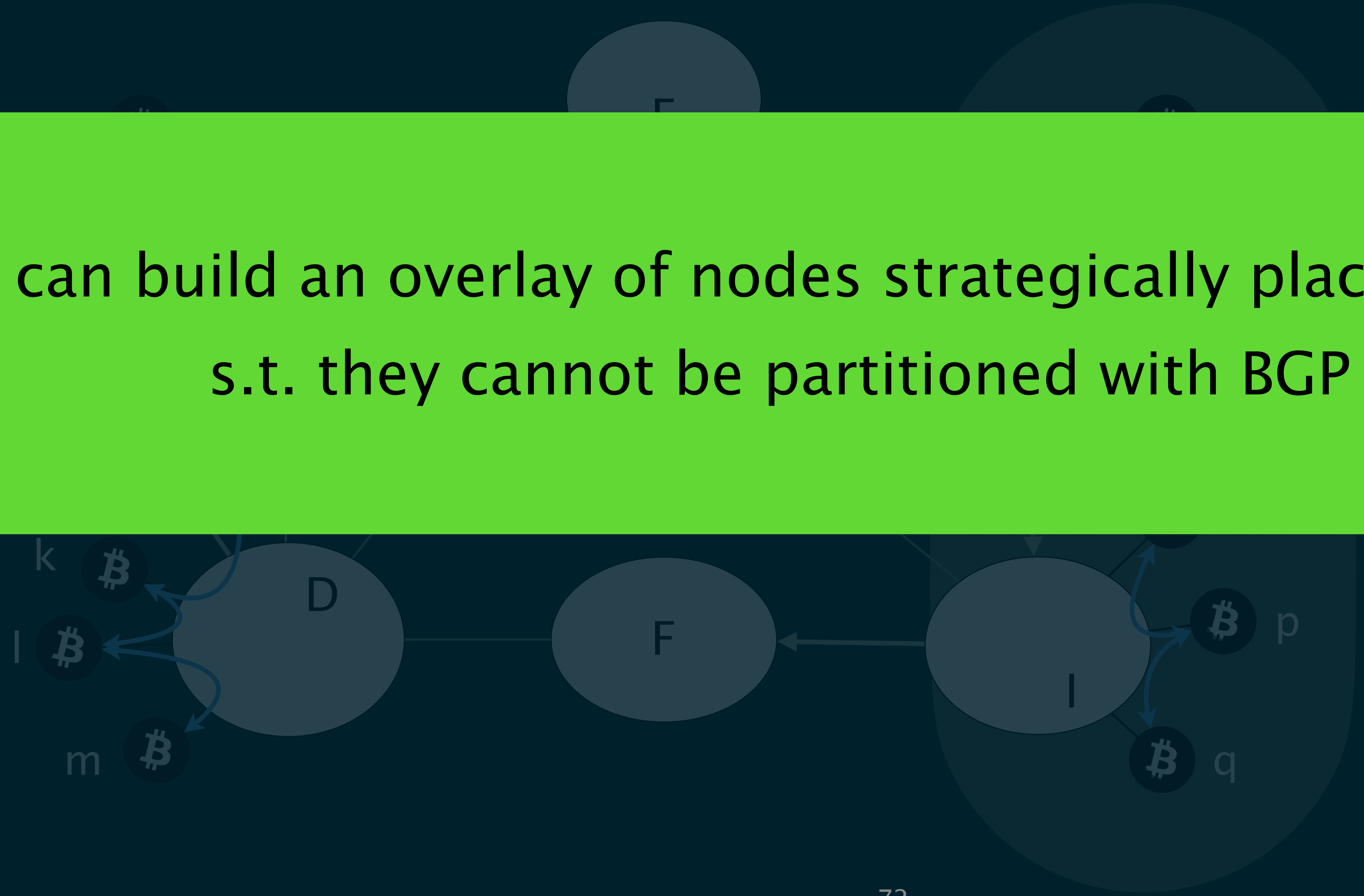
A new block in the grey zone cannot be propagated further



A new block in the grey zone cannot be propagated further

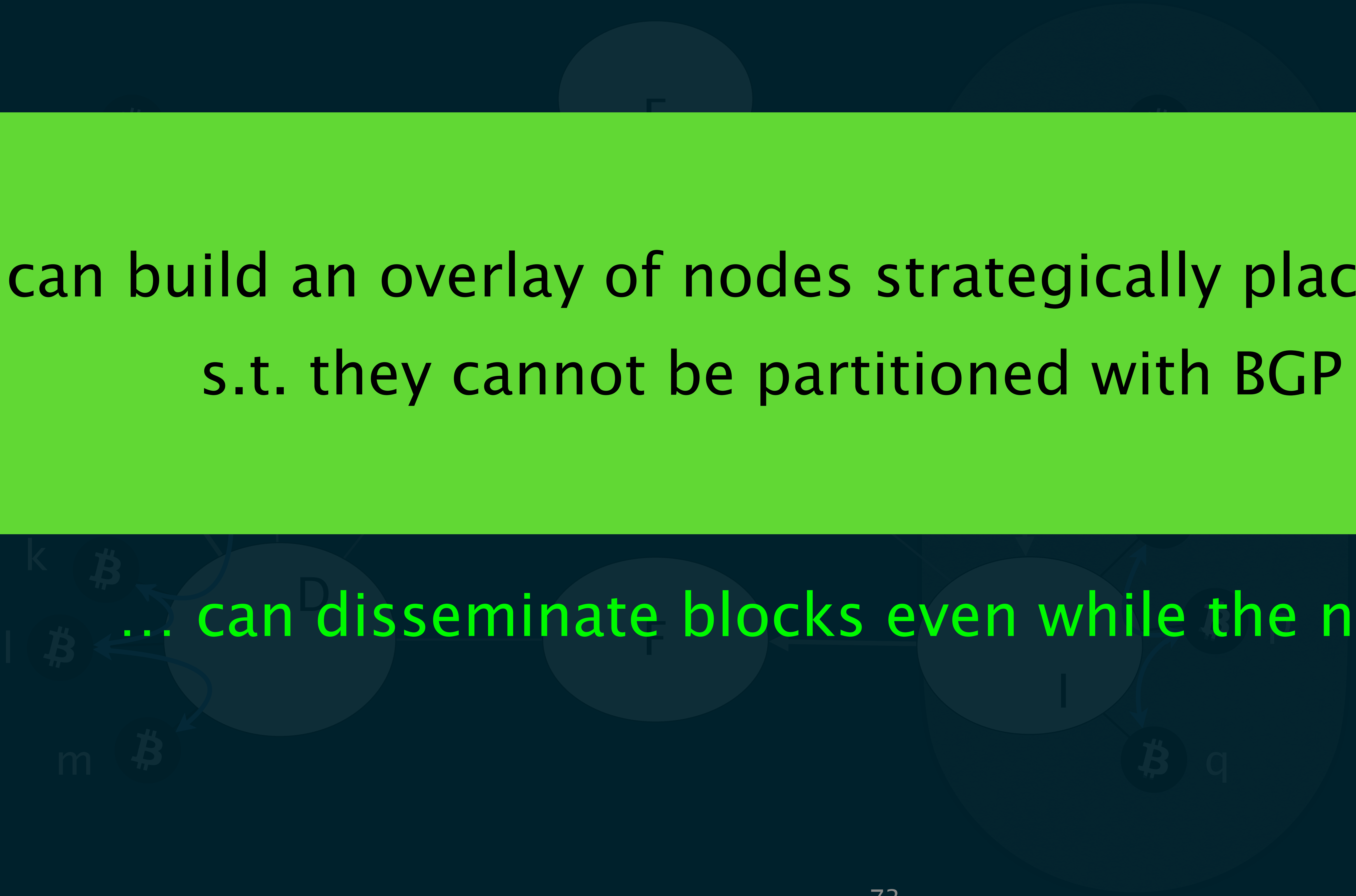


A new block in the grey zone cannot be propagated further



We can build an overlay of nodes strategically placed in the Internet s.t. they cannot be partitioned with BGP hijacks

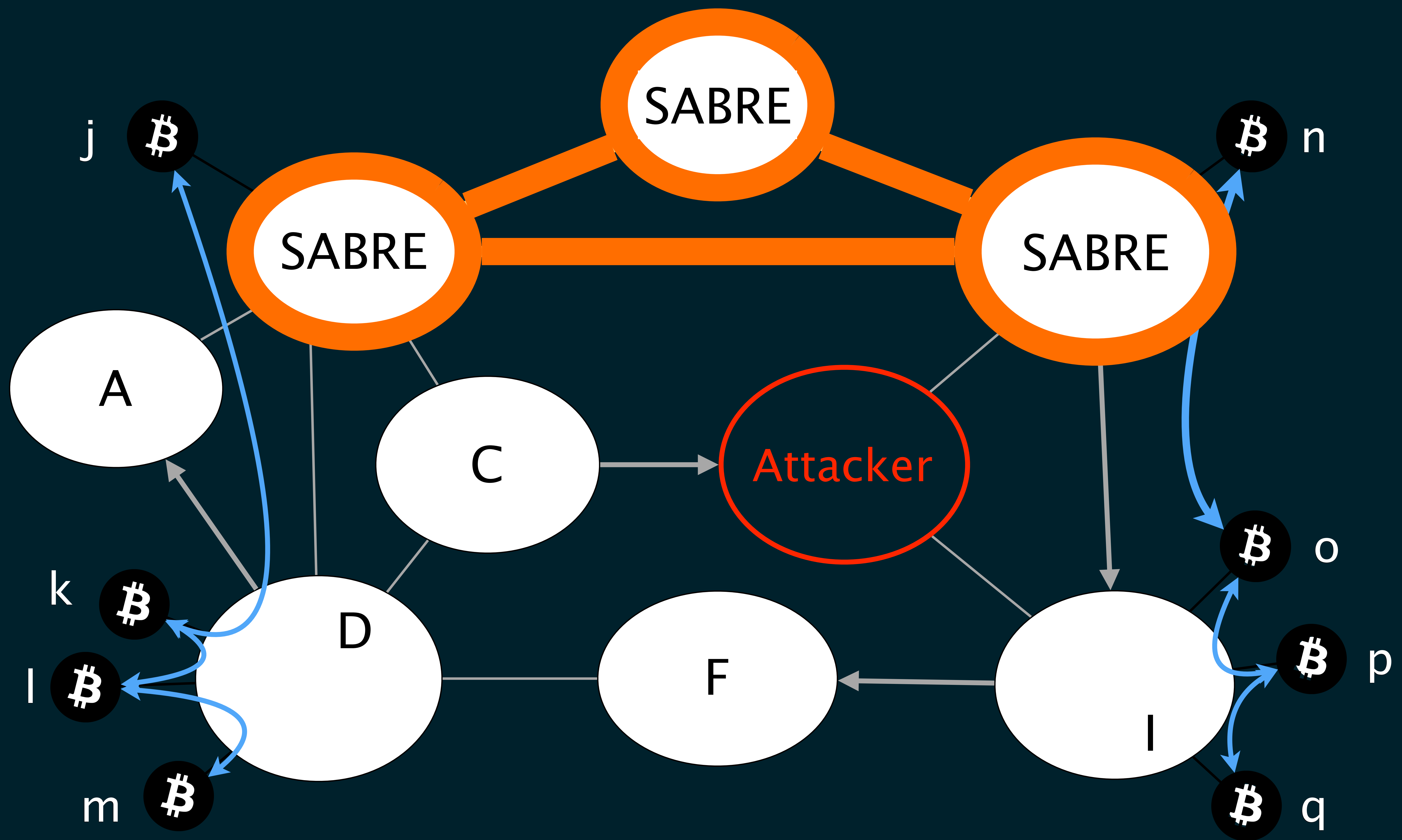
A new block in the grey zone cannot be propagated further



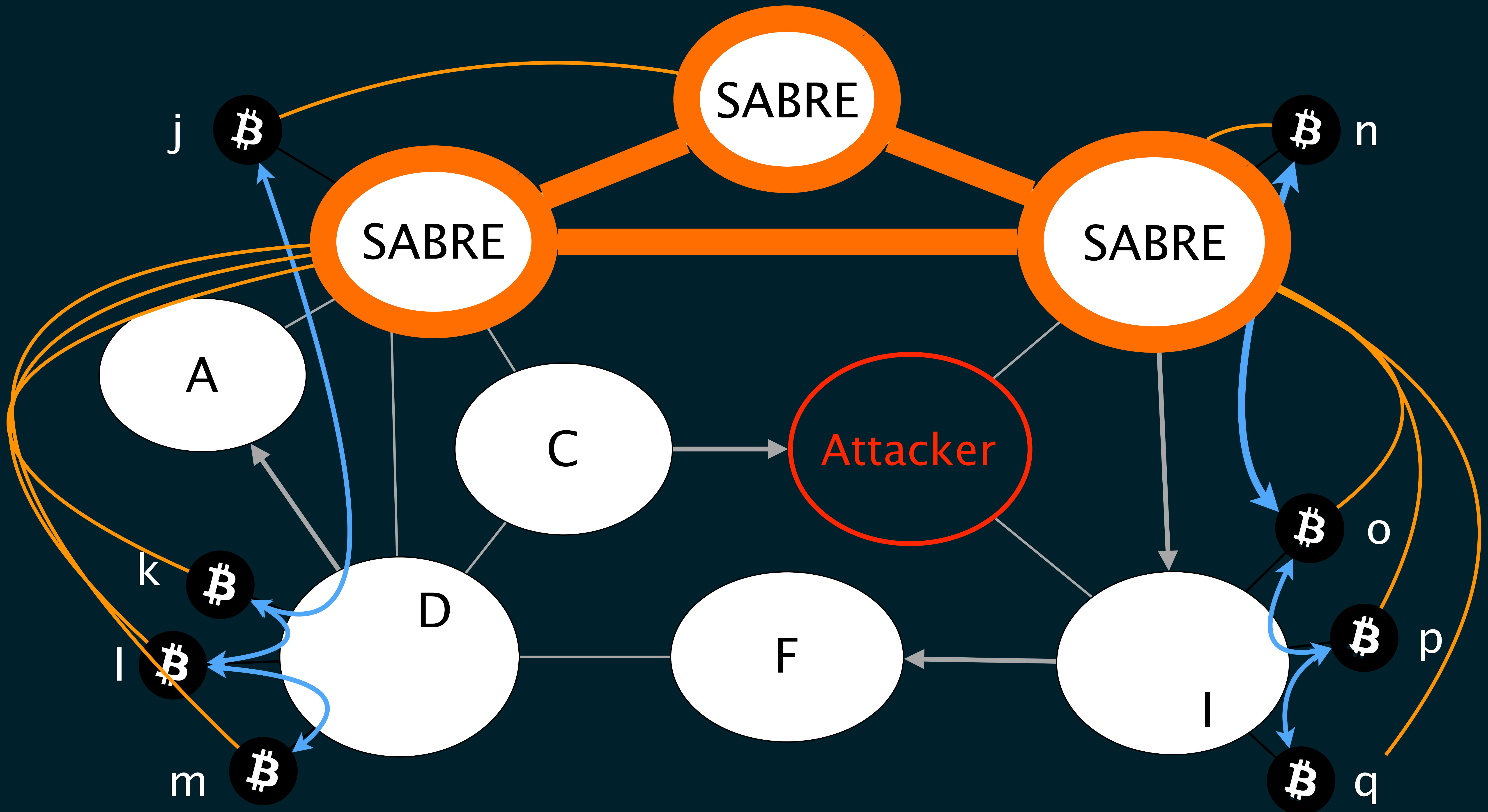
We can build an overlay of nodes strategically placed in the Internet s.t. they cannot be partitioned with BGP hijacks

… can disseminate blocks even while the network is partitioned

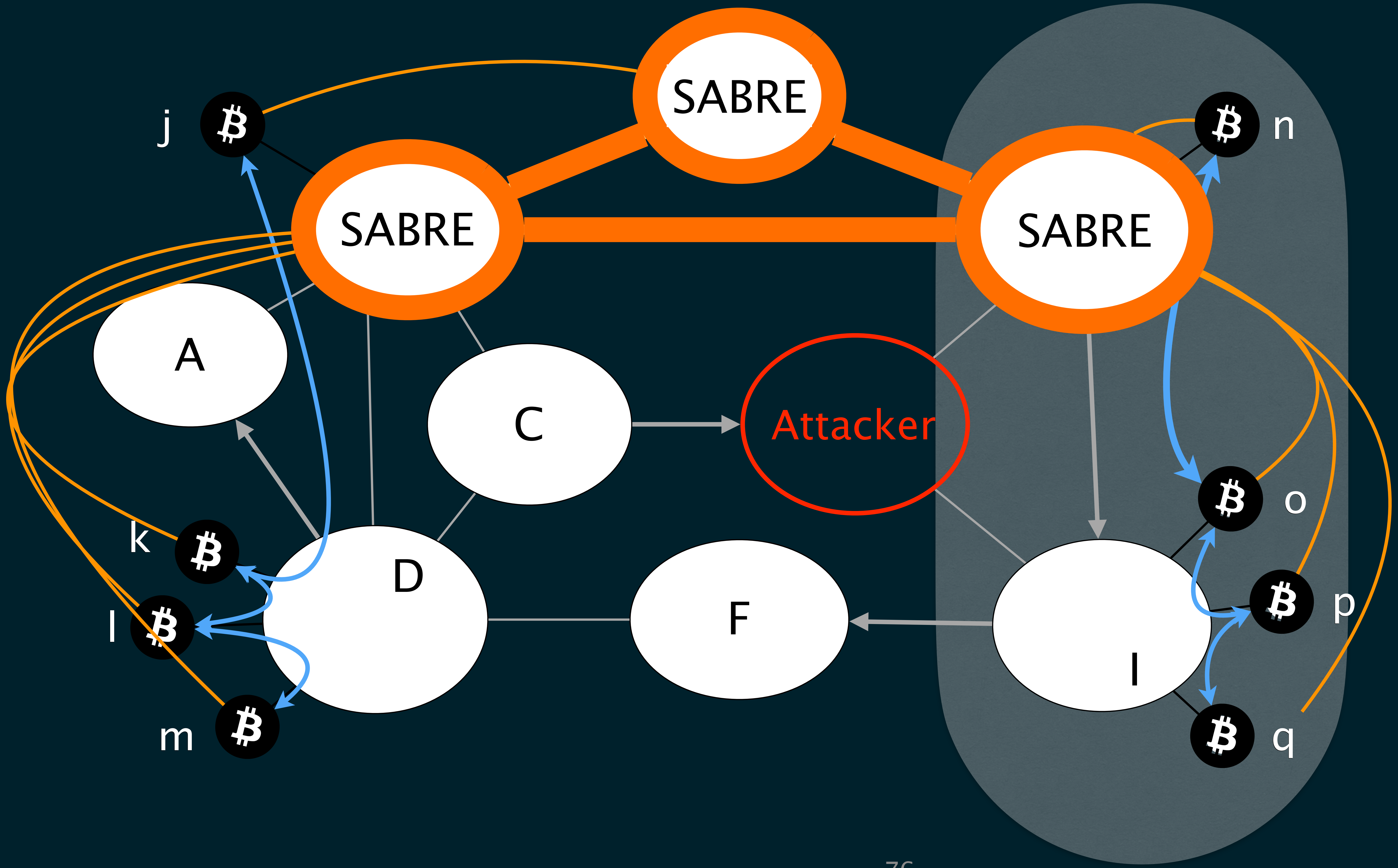
A strongly-connected overlay can disseminate blocks even while the network is partitioned



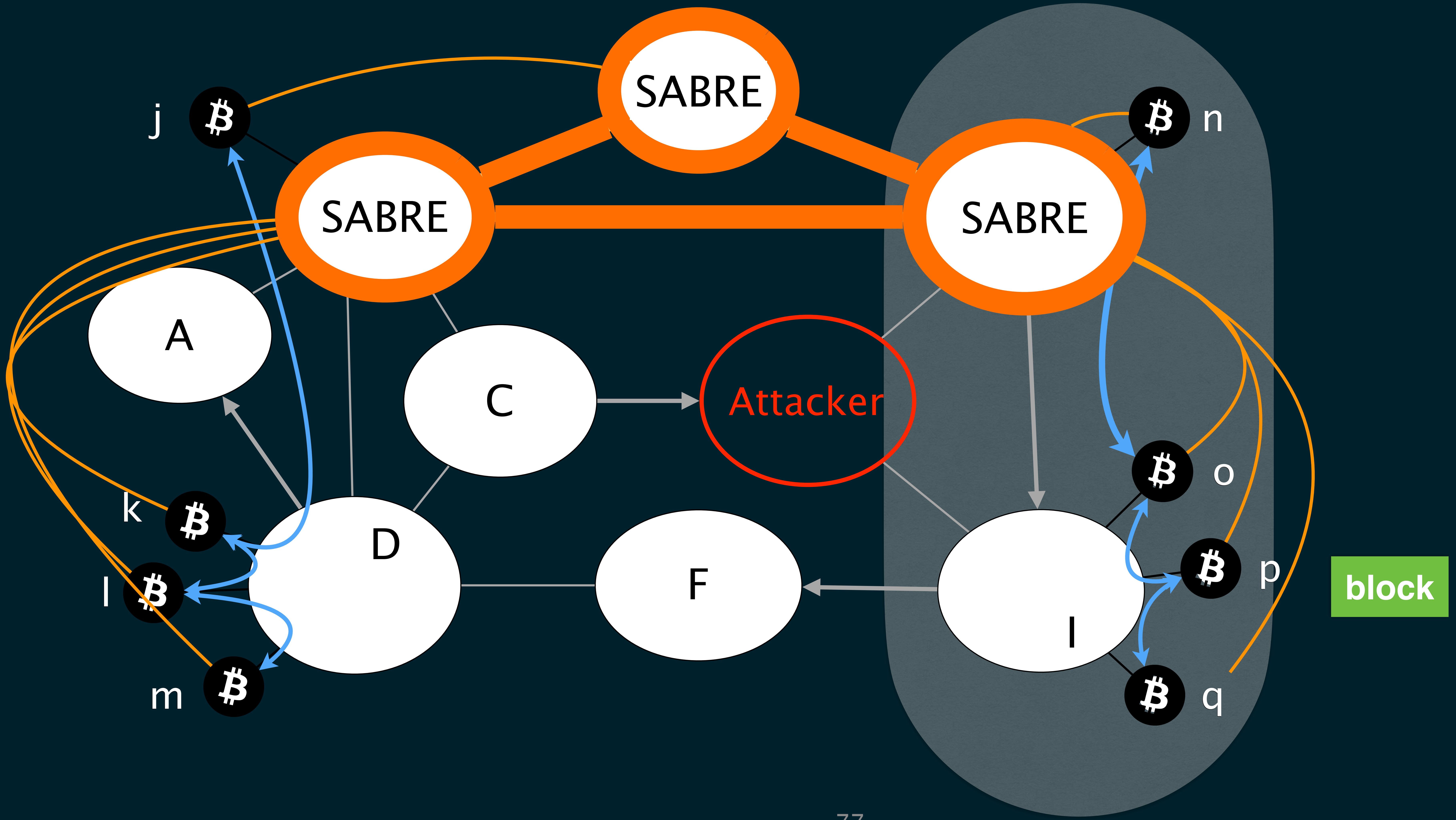
A strongly-connected overlay can disseminate blocks even while the network is partitioned



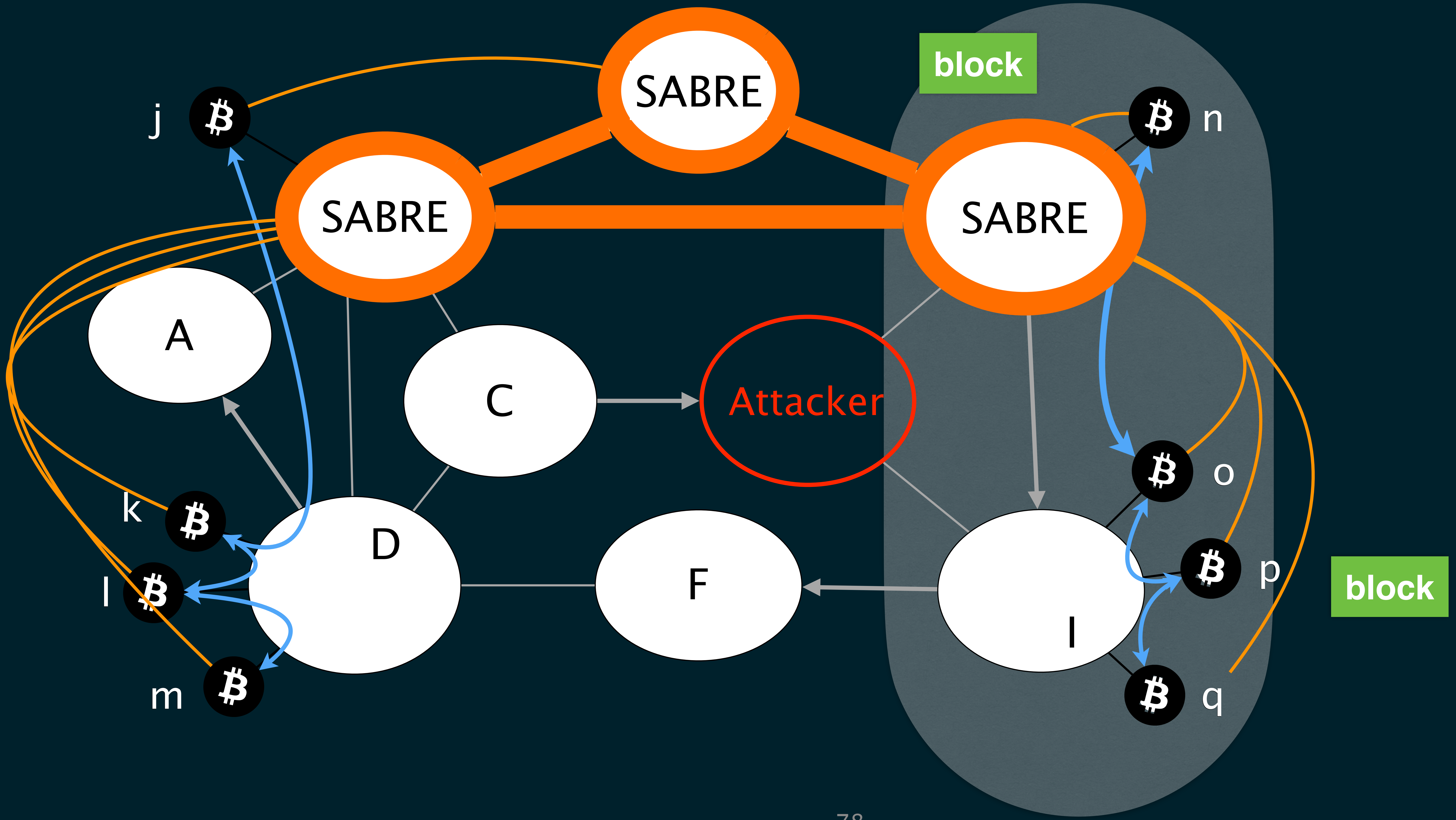
A strongly-connected overlay can disseminate blocks even while the network is partitioned



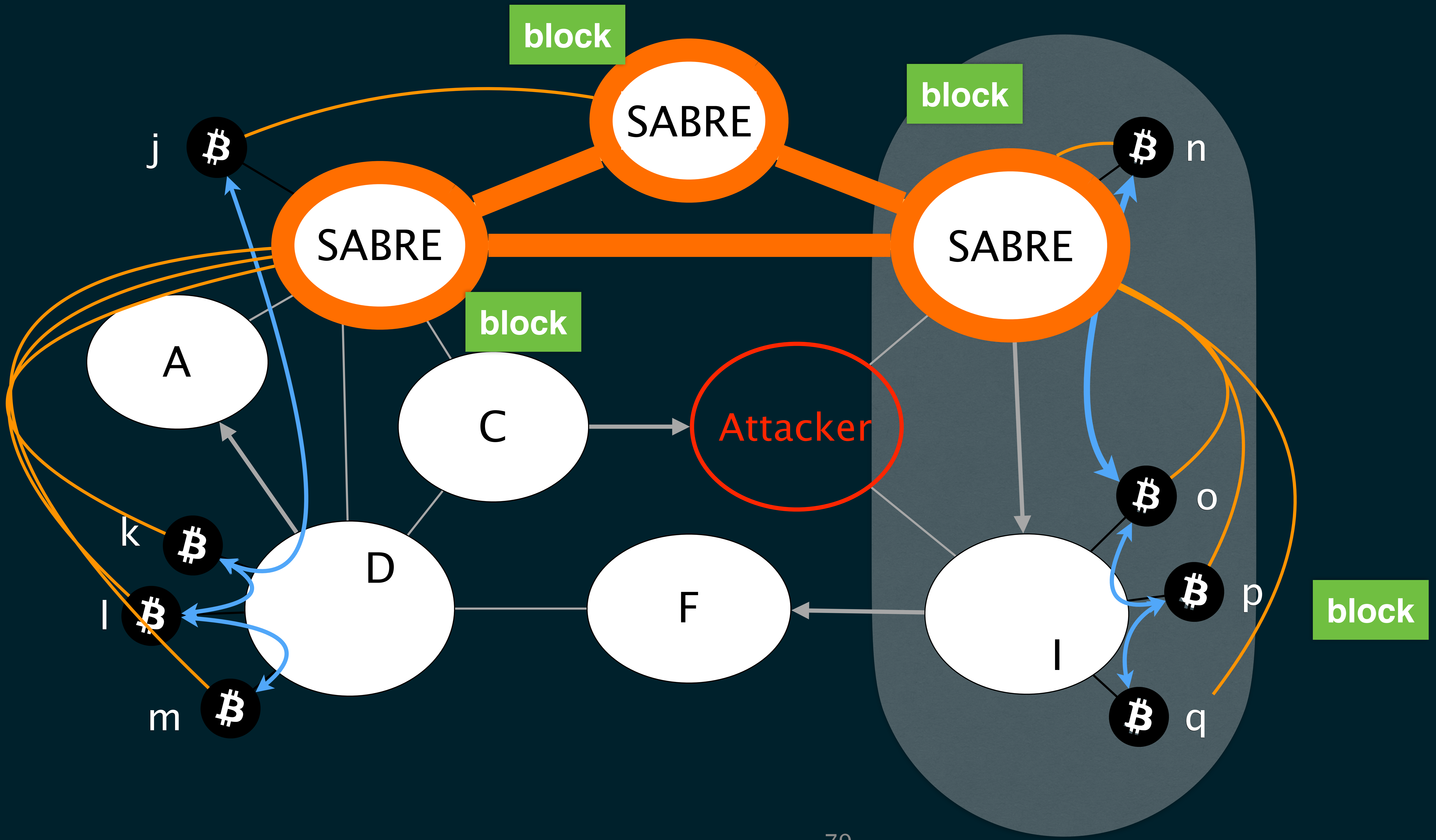
A strongly-connected overlay can disseminate blocks even while the network is partitioned



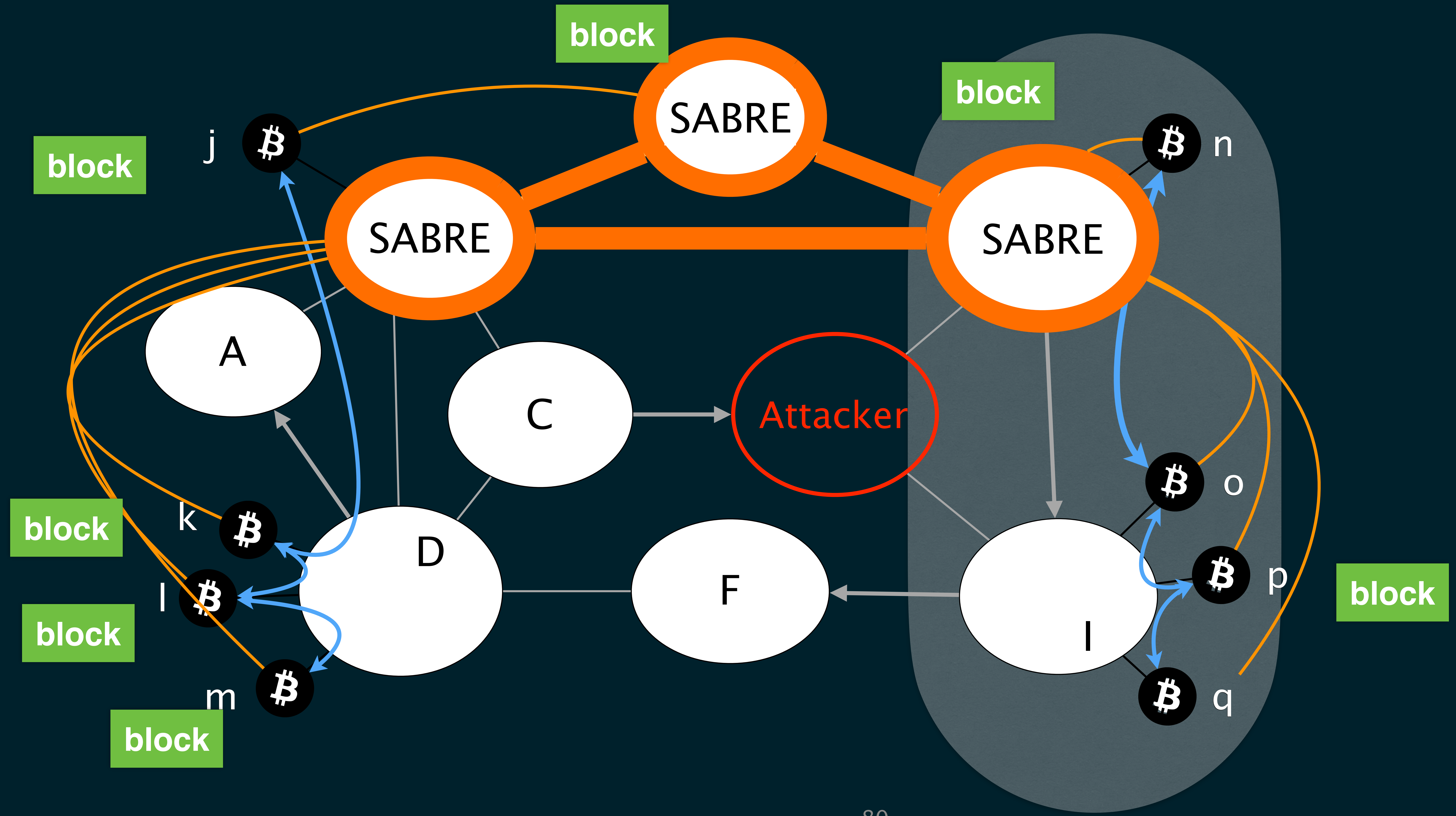
A strongly-connected overlay can disseminate blocks even while the network is partitioned



A strongly-connected overlay can disseminate blocks even while the network is partitioned



A strongly-connected overlay can disseminate blocks even while the network is partitioned



BGP
BGP
ROUTER

How should the SABRE nodes be implemented?

Public SABRE nodes need to scale

Public SABRE nodes need to scale

SABRE nodes need to…

- maintain thousands of (malicious) connections
- distinguish spoofing and malicious requests
- receive, verify and relay blocks fast

Public SABRE nodes need to scale

SABRE nodes need to…

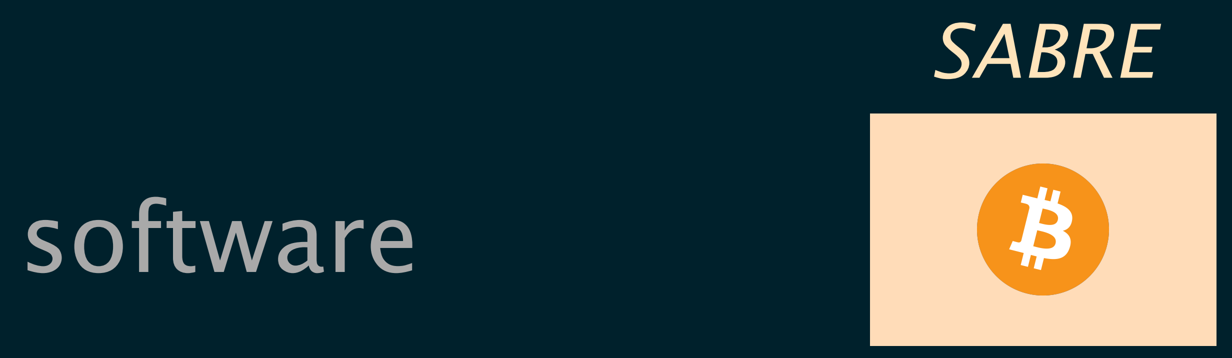
- maintain thousands of (malicious) connections
- distinguish spoofing and malicious requests
- receive, verify and relay blocks fast

Simple software implementation would not suffice!

SABRE can leverage programmable network devices

SABRE DP

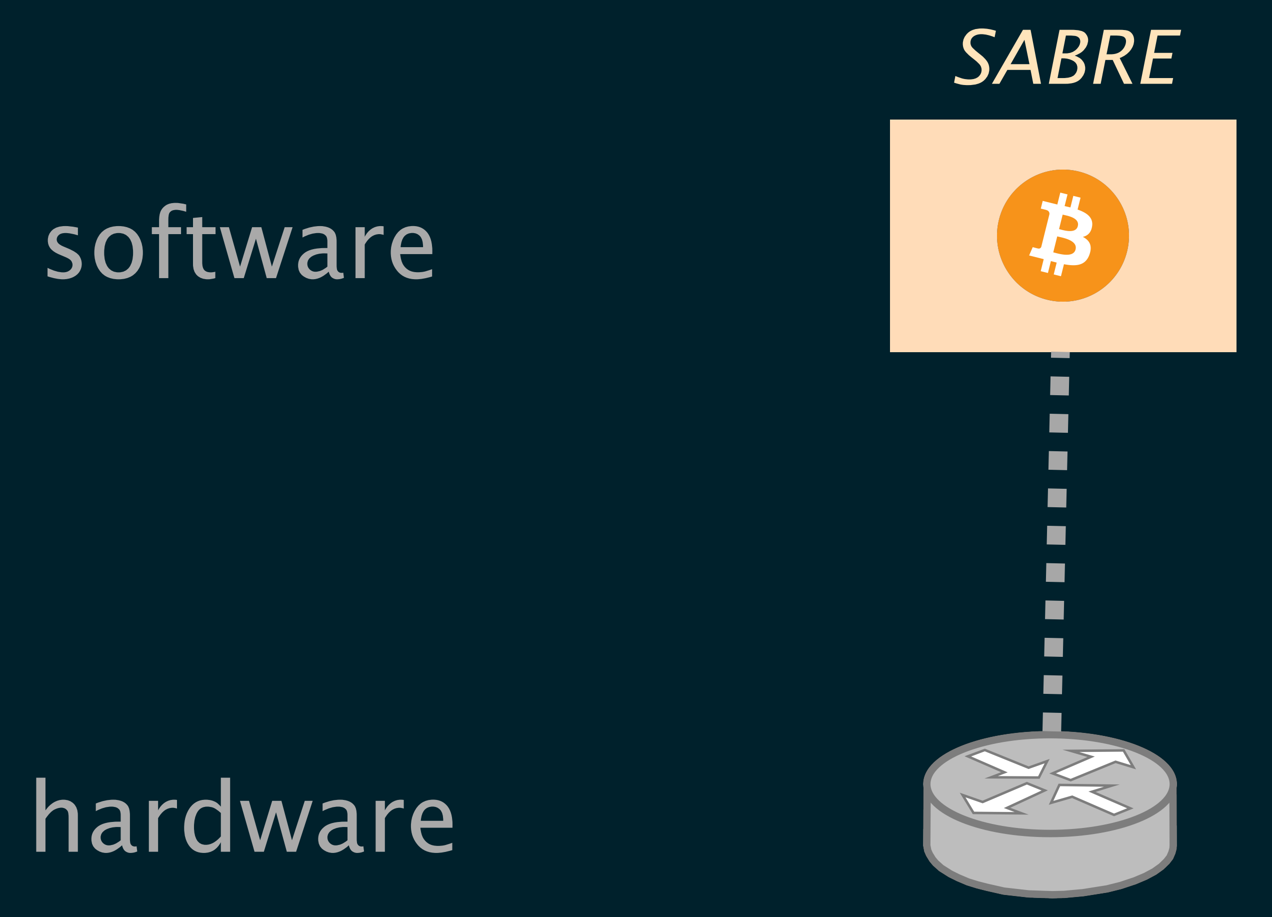
SABRE DP allows relay nodes to deal with high malicious or benign load



Validates new blocks

Updates hardware with new blocks

Not all operations can be done in hardware, a SABRE node combines a software and a hardware part

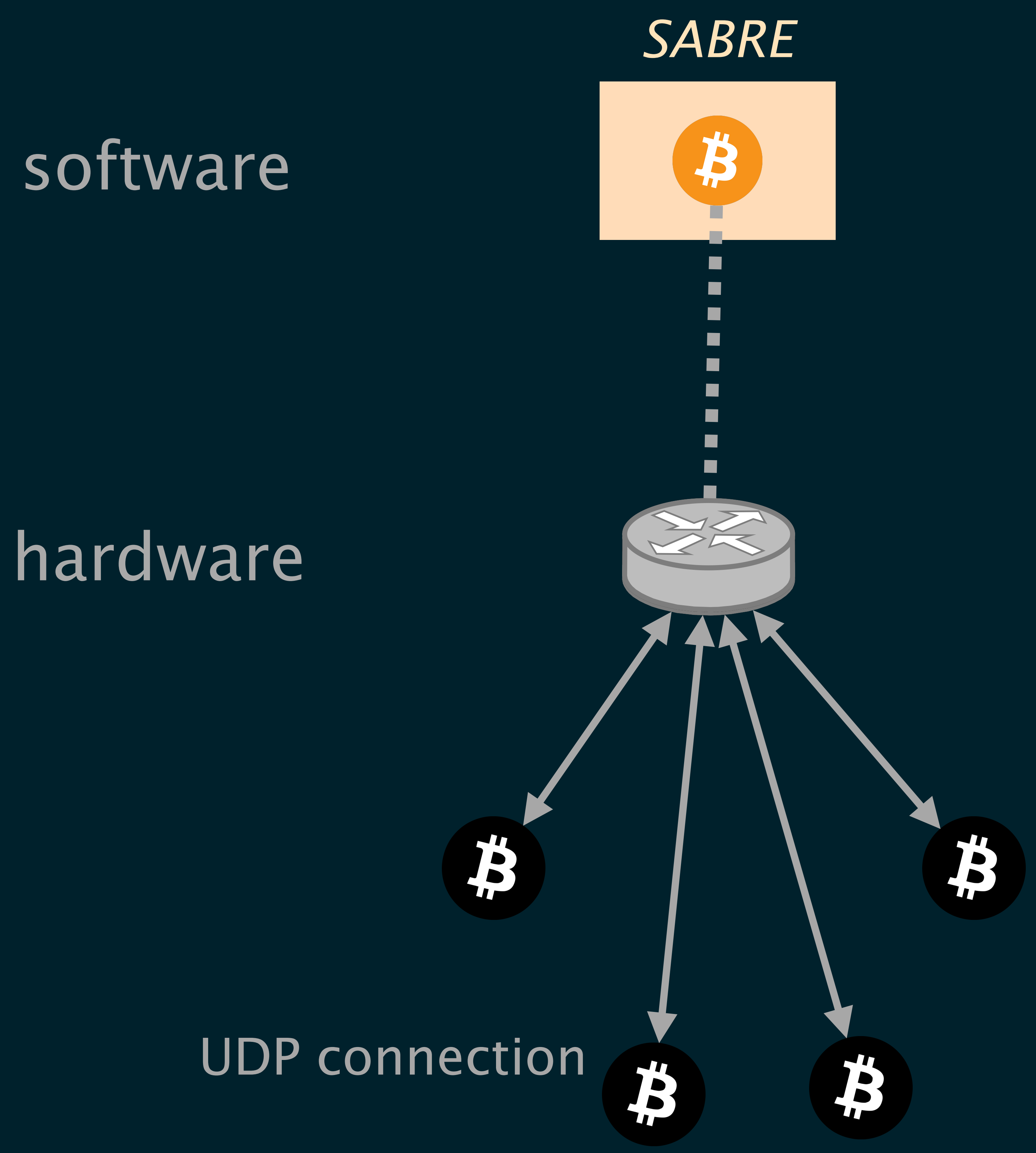


Validates new blocks

Updates hardware with new blocks

Maintains connections to 1000s of clients

Not all operations can be done in hardware, a SABRE node combines a software and a hardware part



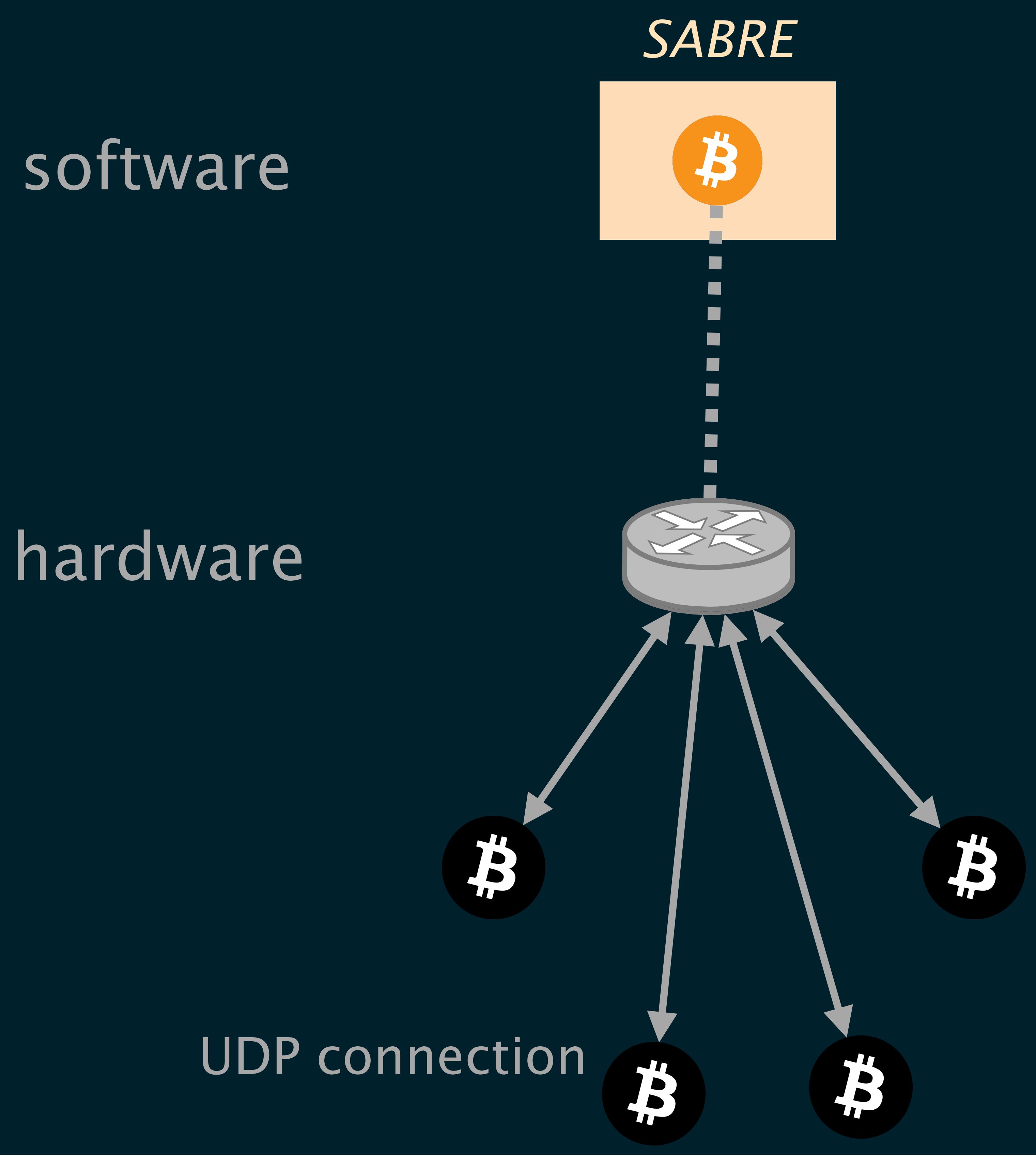
Validates new blocks

Updates hardware with new blocks

Maintains connections to 1000s of clients

Propagates new blocks to connected clients

Not all operations can be done in hardware, a SABRE node combines a software and a hardware part



Validates new blocks

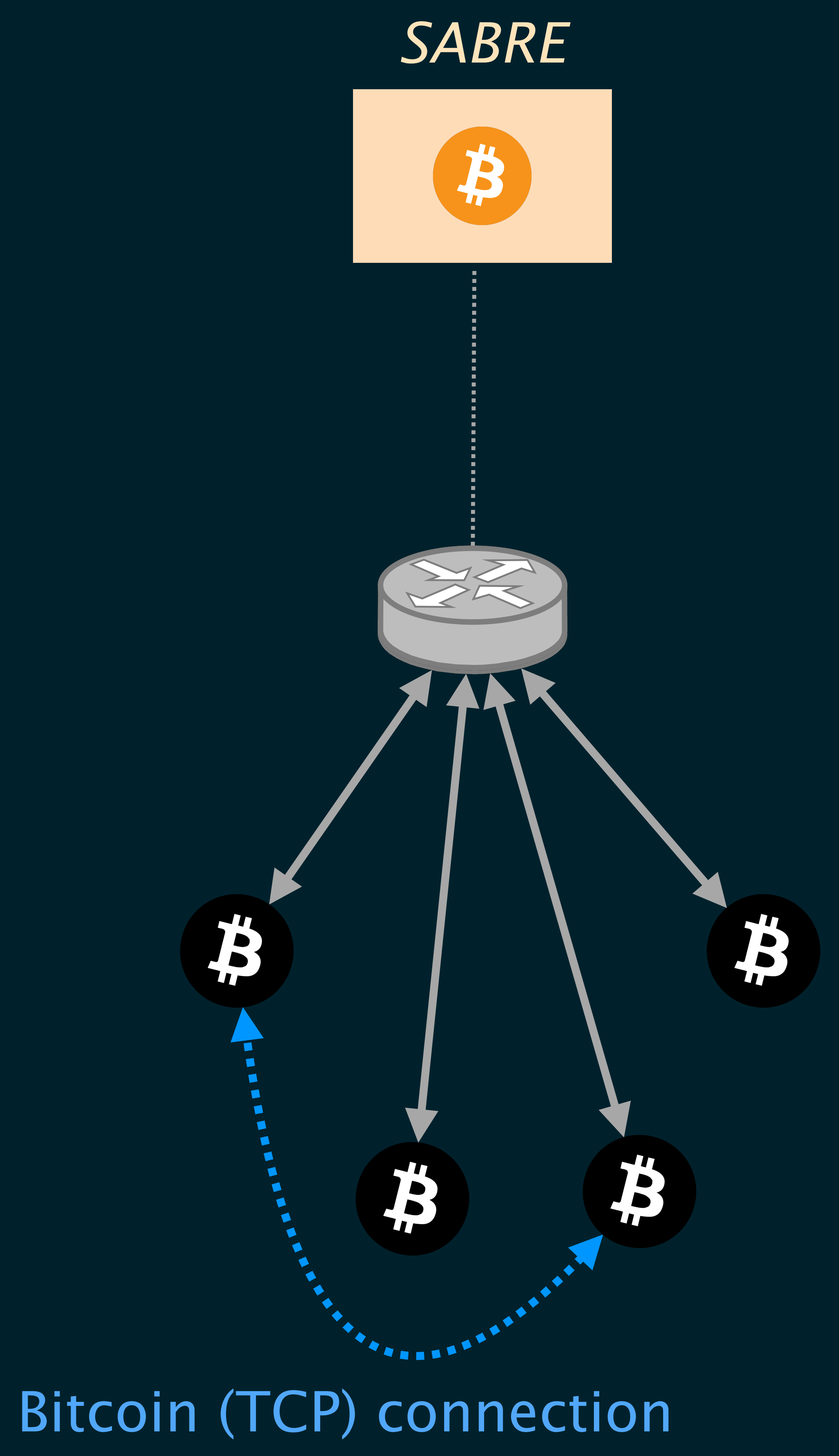
Updates hardware with new blocks

Maintains connections to 1000s of clients

Propagates new blocks to connected clients

Protects the software from malicious clients

Not all operations can be done in hardware, a SABRE node combines a software and a hardware part



Validates new blocks

Updates hardware with new blocks

Maintains connections to 1000s of clients

Propagates new blocks to connected clients

Protects the software from malicious clients

What can you do with a couple of programmable points in the Internet?

Tango: performance-driven routing system

NSDI’24

SABRE: secure overlay for BTC block propagation

NDSS’19

What can you do with a couple of programmable points in the Internet?

<your answer here>

NET Σyn

Maria Apostolaki

netsyn.princeton.edu