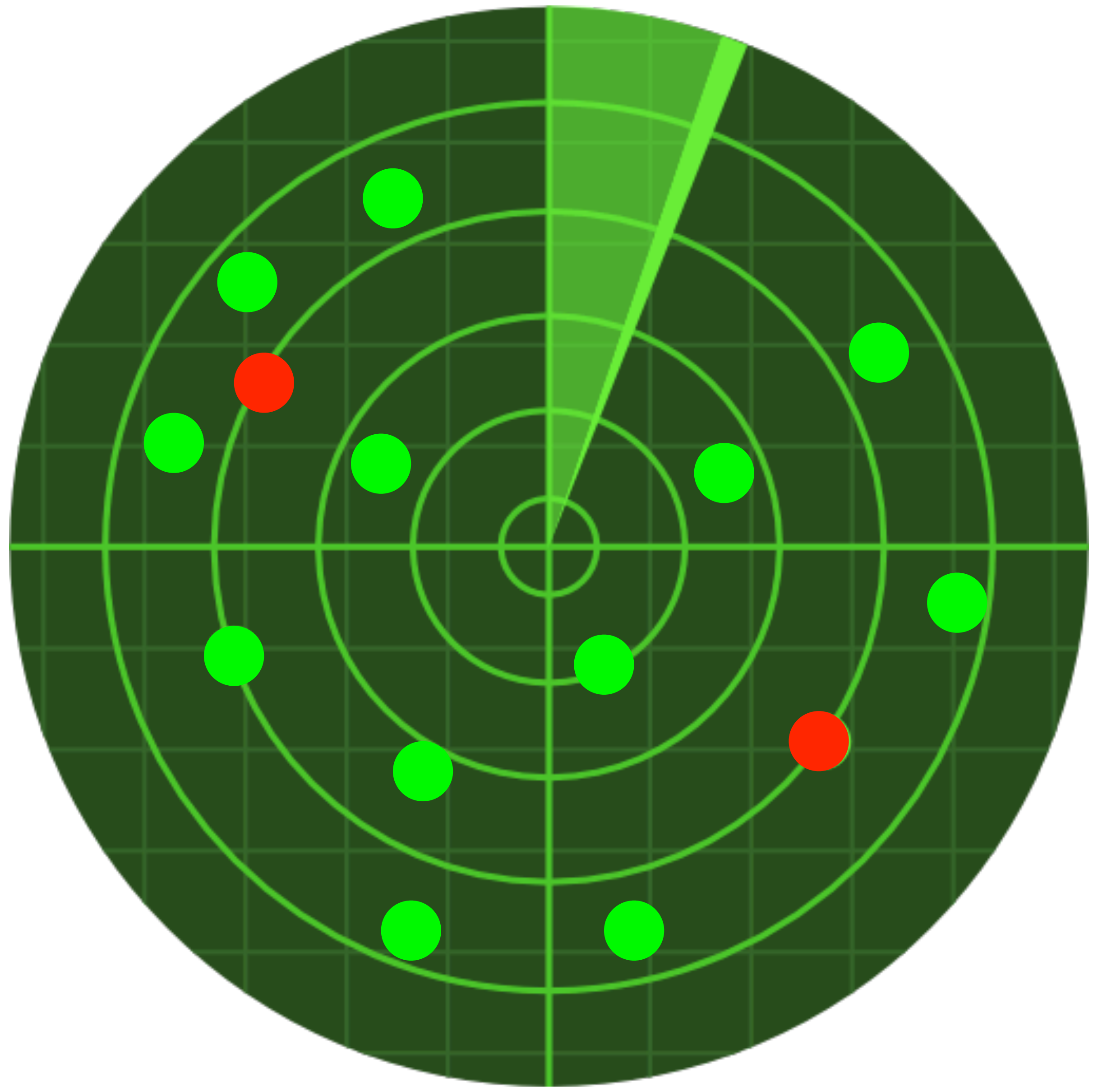
A System to Detect Forged-Origin Hijacks

Thomas Holterbach
University of Strasbourg

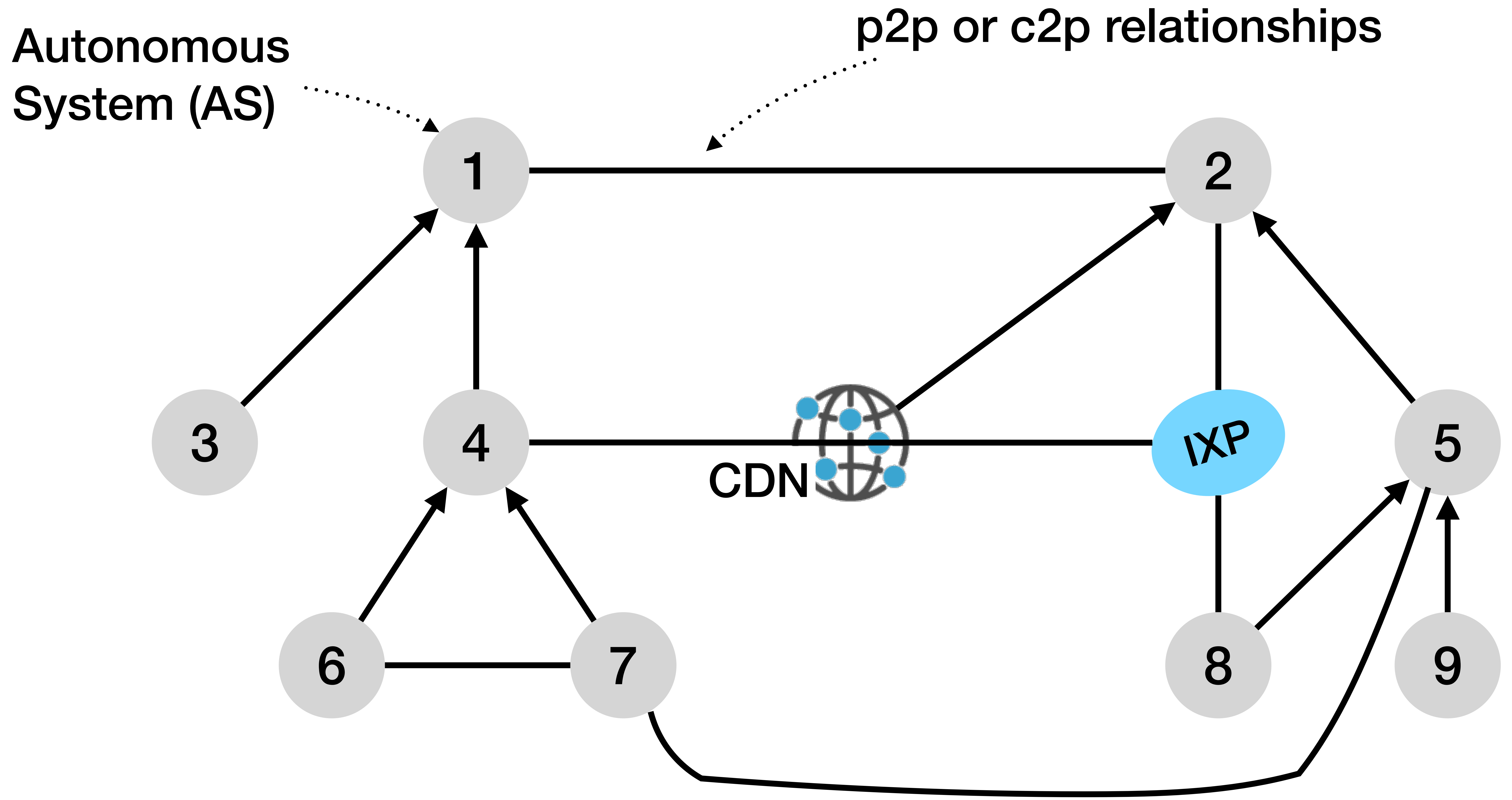
Routing Security Summit
2023



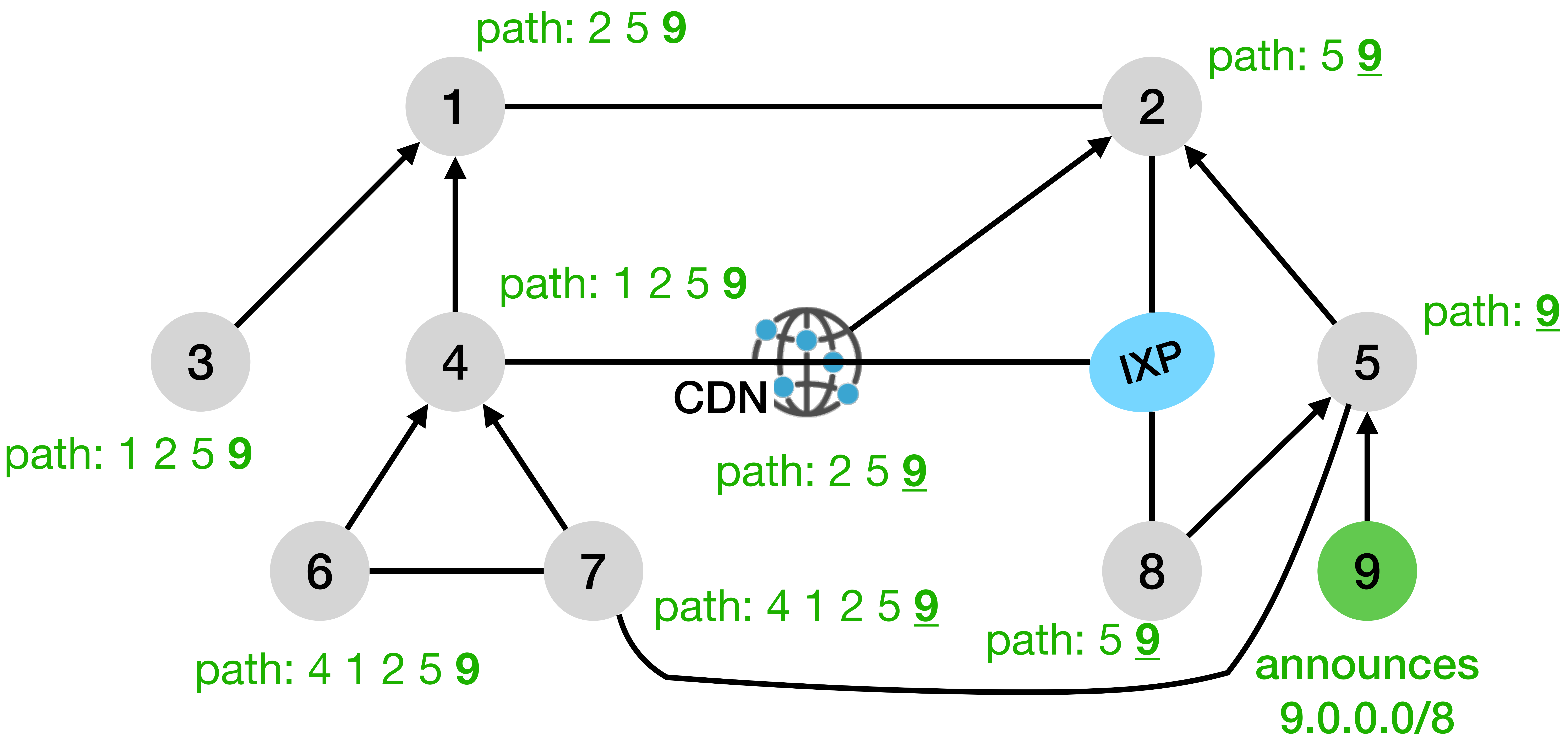
Joint work with:

Thomas Alfroy
Amreesh D. Phokeer
Alberto Dainotti
Cristel Pelsser

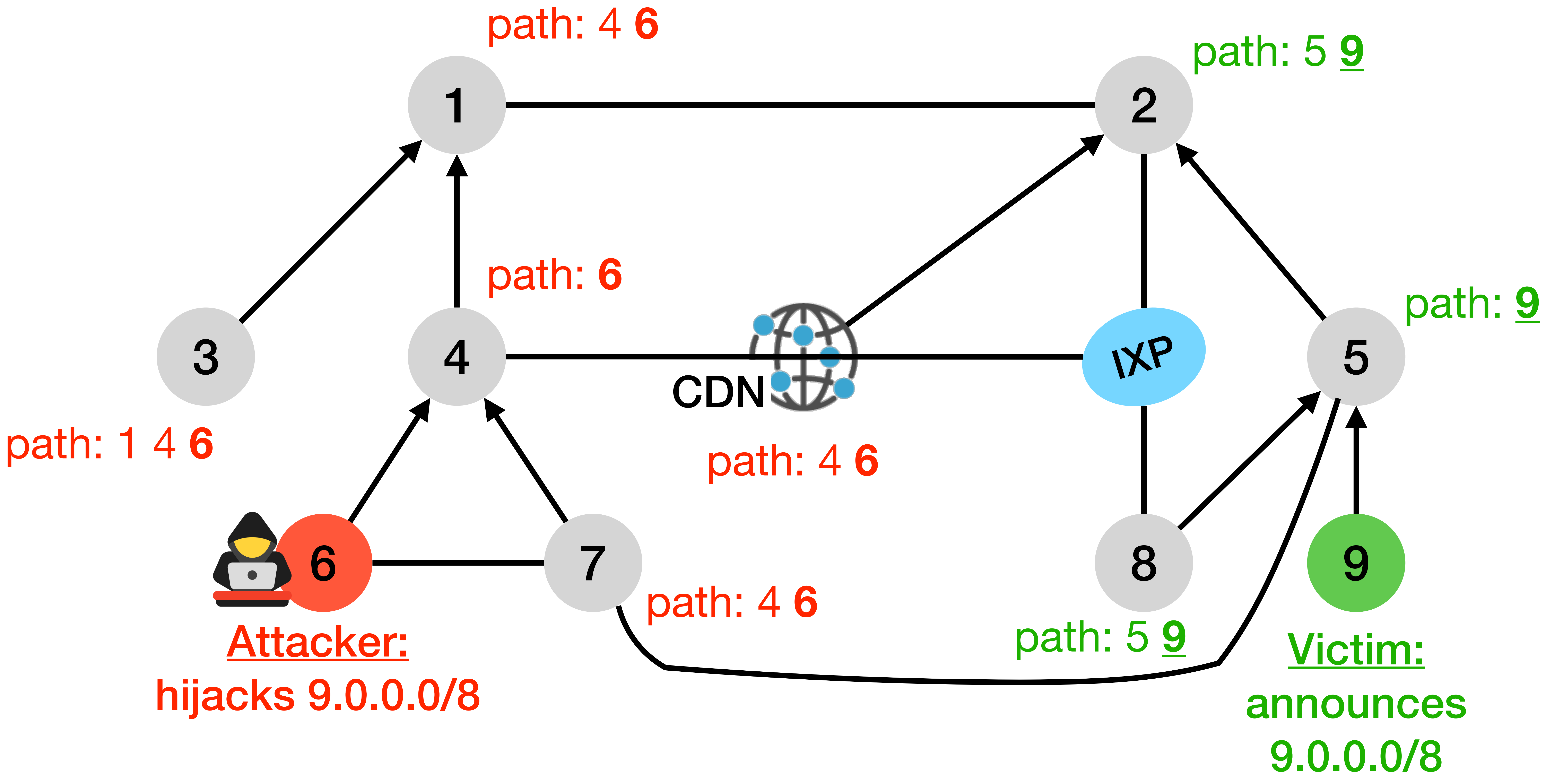
Internet routing (BGP) is vulnerable to traffic hijacking



Internet routing (BGP) is vulnerable to traffic hijacking



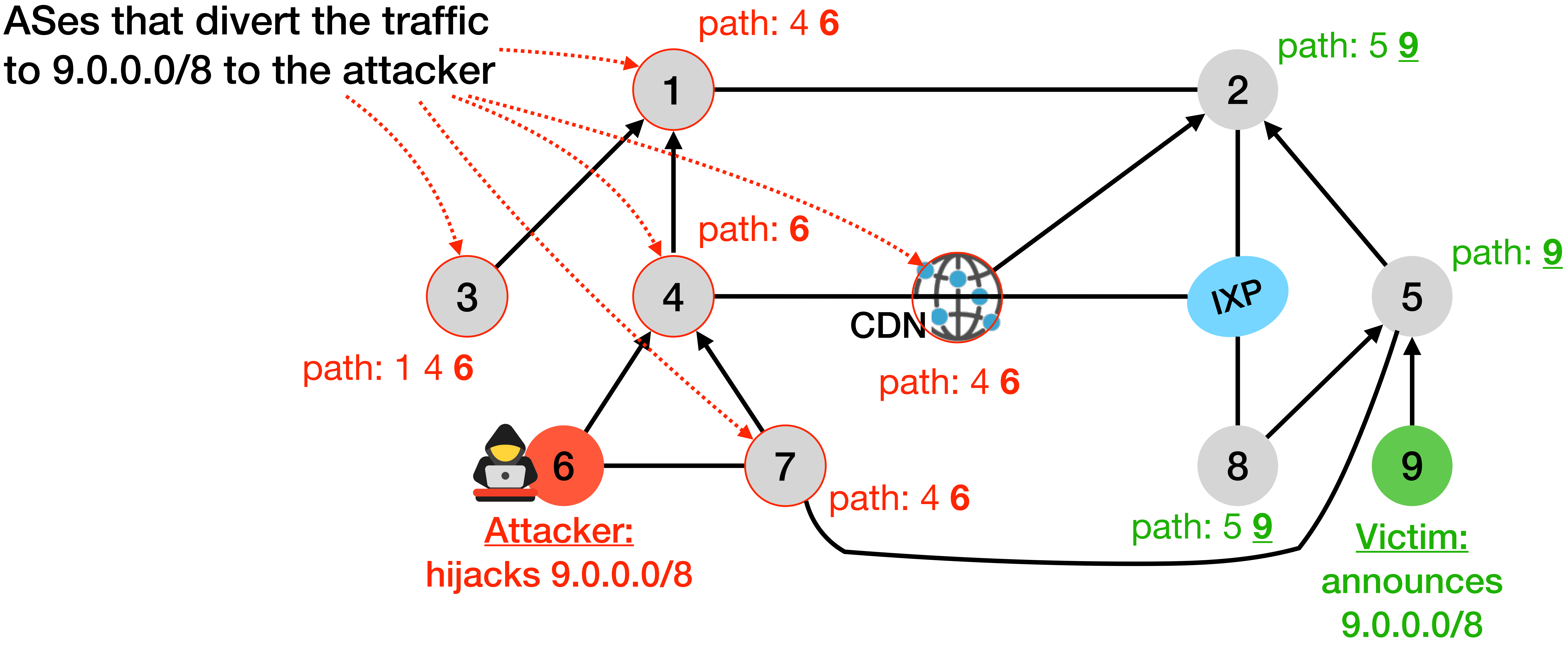
Internet routing (BGP) is vulnerable to traffic hijacking



Attacker:
hijacks 9.0.0.0/8

Victim:
announces
9.0.0.0/8

Internet routing (BGP) is vulnerable to traffic hijacking

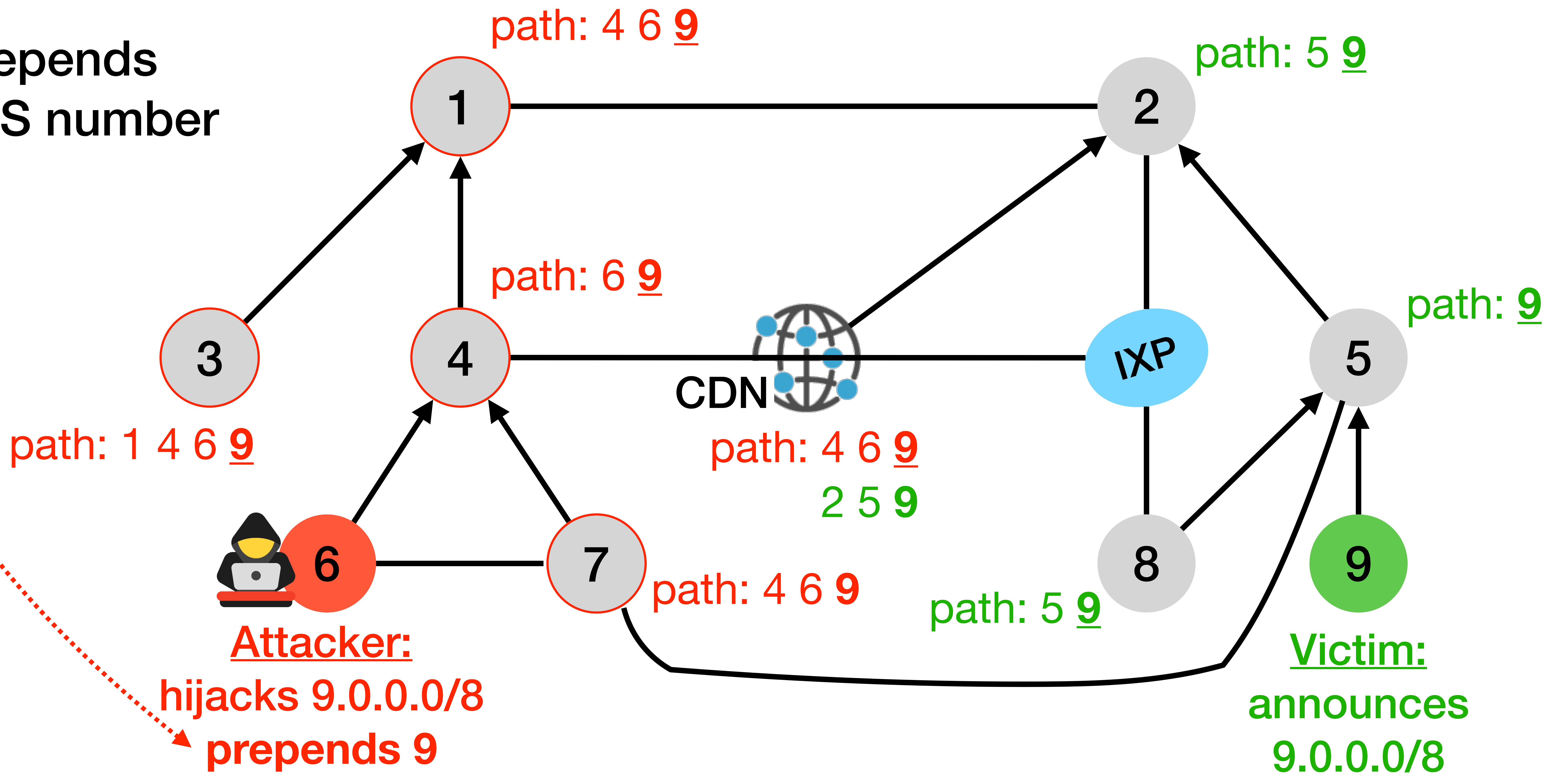


Fortunately, there are defences against BGP hijacking

Protocol extensions	RPKI + ROV BGPSec, ASPA
Configuration guidelines	Route filters
Monitoring platforms	ARTEMIS BGPAlerter

Despite the efforts, BGP is *still* vulnerable to forged-origin hijacks

The attacker prepends the legitimate AS number to the AS path



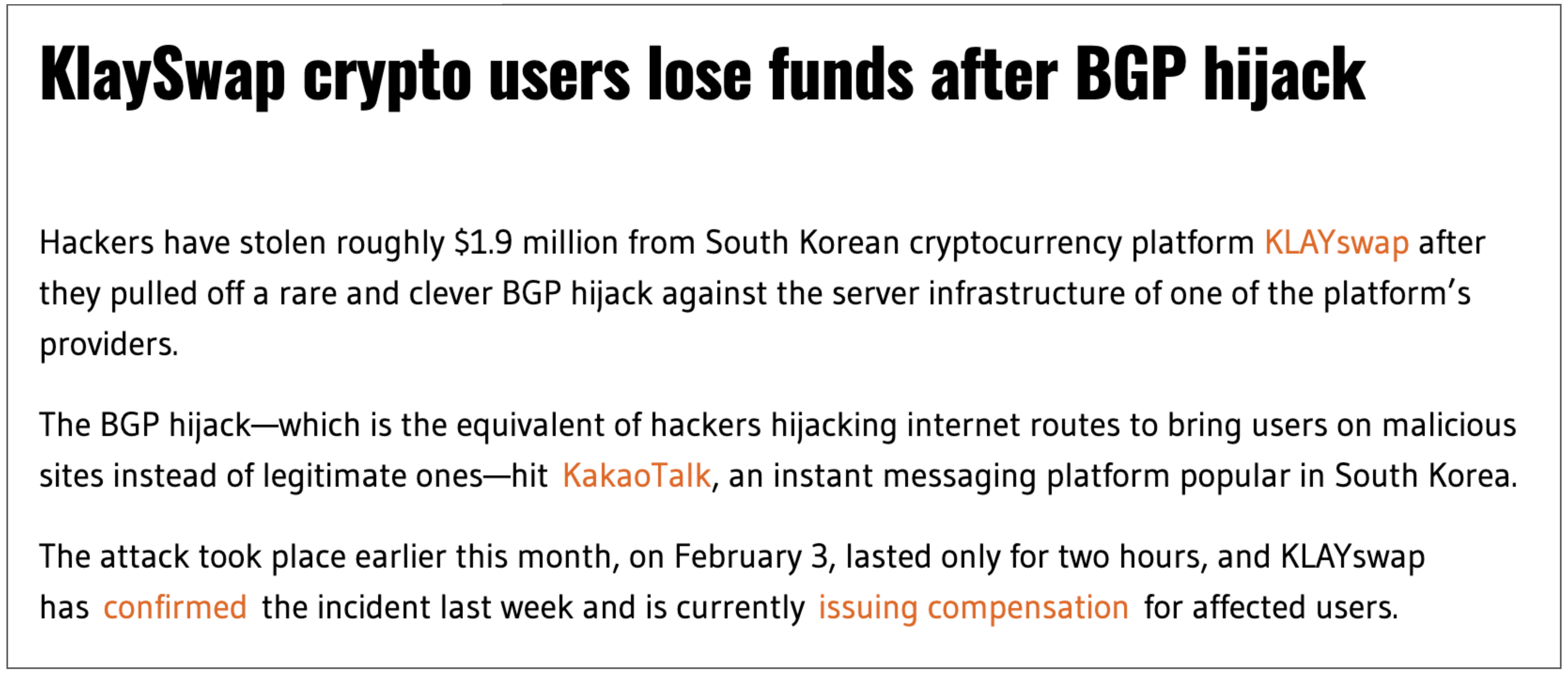
Existing defenses poorly neutralise forged-origin hijacks

Protocol extensions	RPKI + ROV BGPSec, ASPA	RPKI+ROV can’t detect forged-origin hijacks ASPA will take years to be deployed
Configuration guidelines	Route filters	Often missing and inaccurate as they are constructed based on the IRR
Monitoring platforms	ARTEMIS BGPAlerter	Narrowly focused as they detect hijacks that only pertain to the AS deploying it

Forged-origin hijacks are actively used by attackers

The Record.
Recorded Future® News

February 3, 2022



KlaySwap crypto users lose funds after BGP hijack

Hackers have stolen roughly $1.9 million from South Korean cryptocurrency platform KLAYswap after they pulled off a rare and clever BGP hijack against the server infrastructure of one of the platform's providers.

The BGP hijack—which is the equivalent of hackers hijacking internet routes to bring users on malicious sites instead of legitimate ones—hit KakaoTalk, an instant messaging platform popular in South Korea.

The attack took place earlier this month, on February 3, lasted only for two hours, and KLAYswap has confirmed the incident last week and is currently issuing compensation for affected users.

August 17, 2022

CelerNetwork
@CelerNetwork · **Follow**

📢📢📢We are seeing reports that reflects potential DNS hijacking of cbridge frontend. We are investigating at the moment and please do not use the frontend for bridging at the moment.

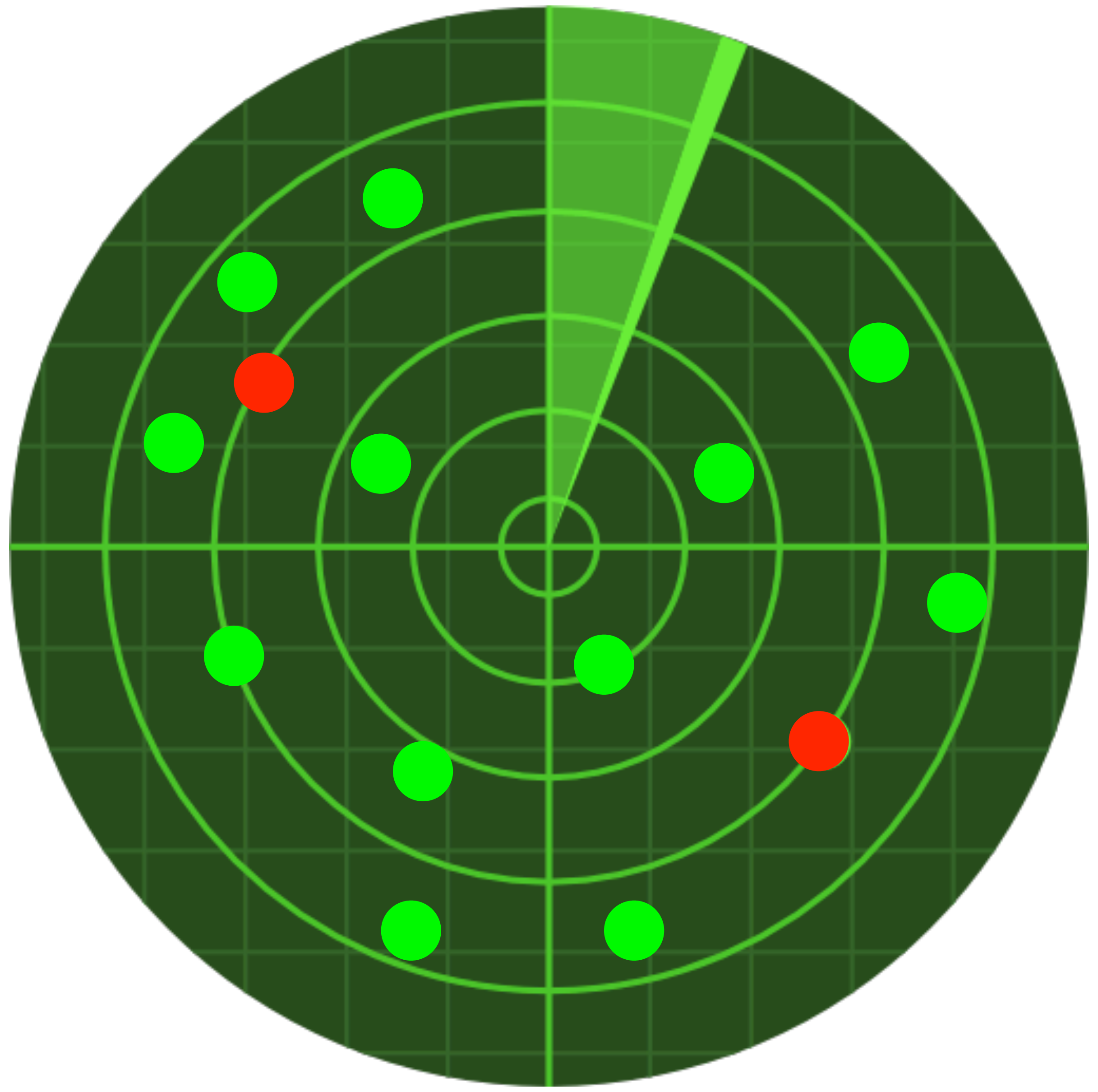
11:56 PM · Aug 17, 2022

Both attacks are the result of a forged-origin hijack

DFOH: A System to Detect Forged-Origin Hijacks on the Whole Internet

Thomas Holterbach
University of Strasbourg

Routing Security Summit
2023



Joint work with:

Thomas Alfroy
Amreesh D. Phokeer
Alberto Dainotti
Cristel Pelsser

Outline

1. ***DFOH***’s main challenge
is to detect fake AS links

2. ***DFOH’s*** key ingredients are
carefully selected features and a balanced sampling

3. ***DFOH*** is accurate
and practical for users

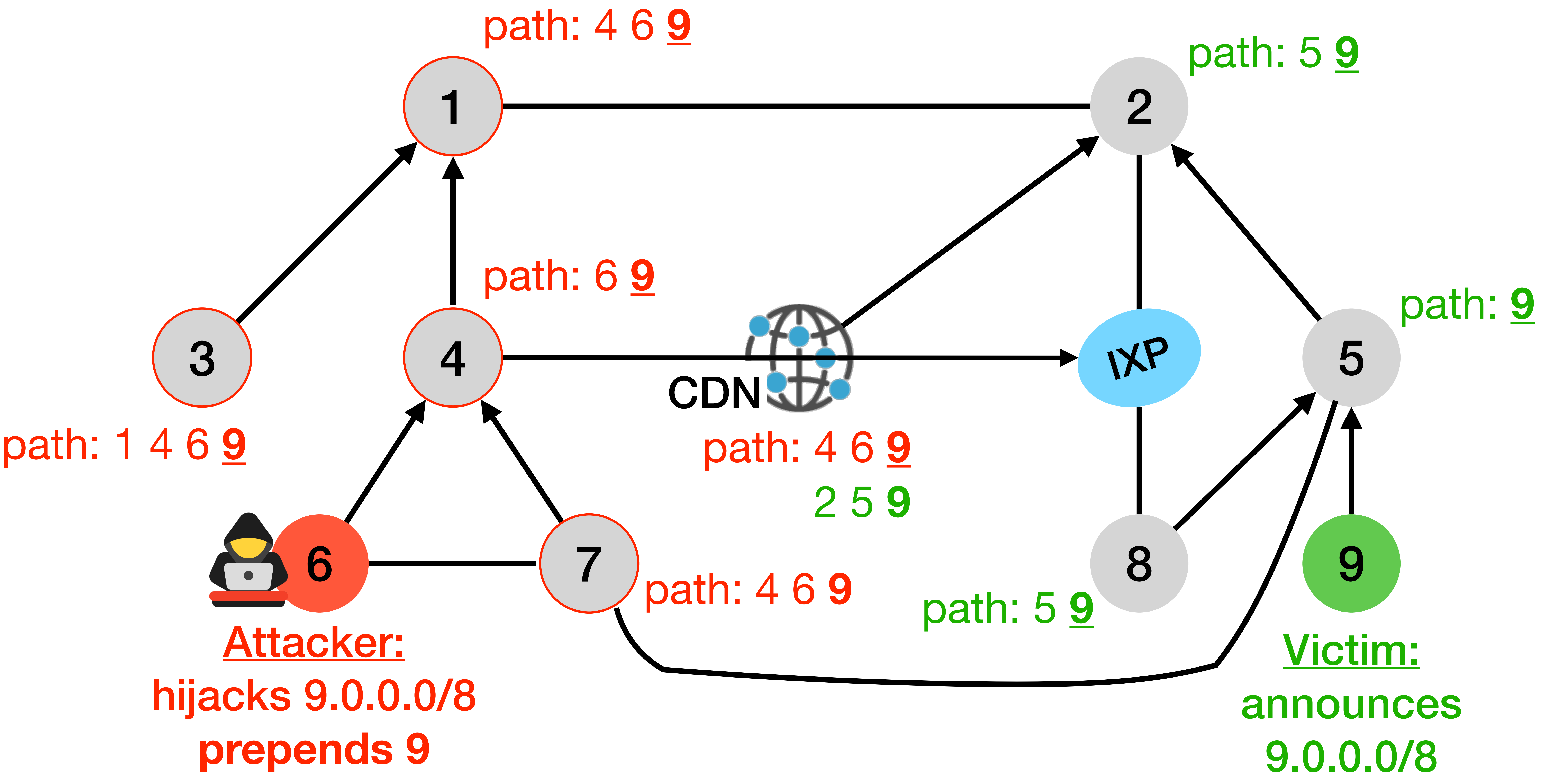
Outline

1. ***DFOH***'s main challenge
is to detect fake AS links

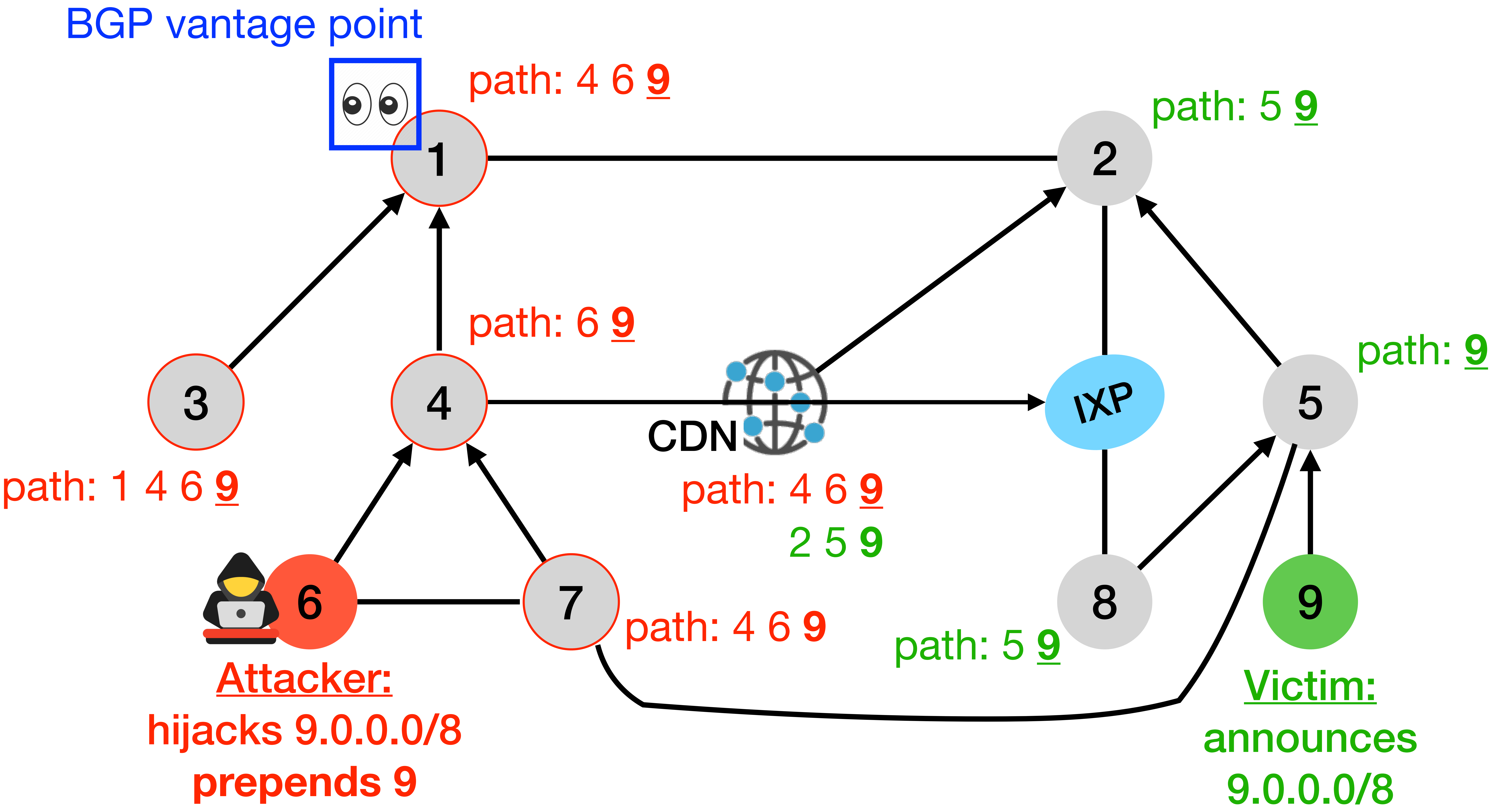
2. ***DFOH's*** key ingredients are
carefully selected features and a balanced sampling

3. ***DFOH*** is accurate
and practical for users

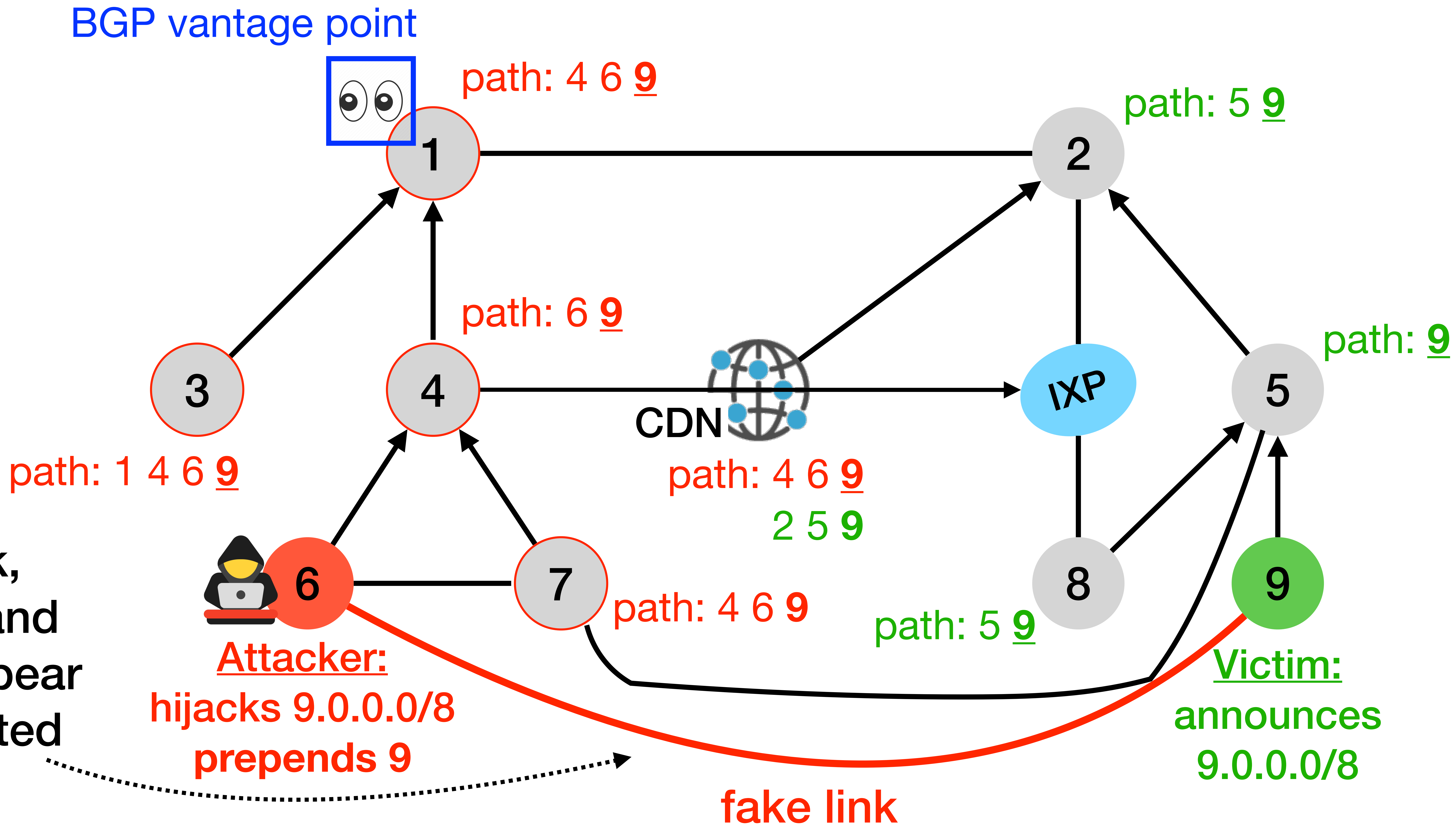
DFOH aims to detect the fake AS links induced by forged-origin hijacks



DFOH aims to detect the fake AS links induced by forged-origin hijacks



DFOH aims to detect the fake AS links induced by forged-origin hijacks

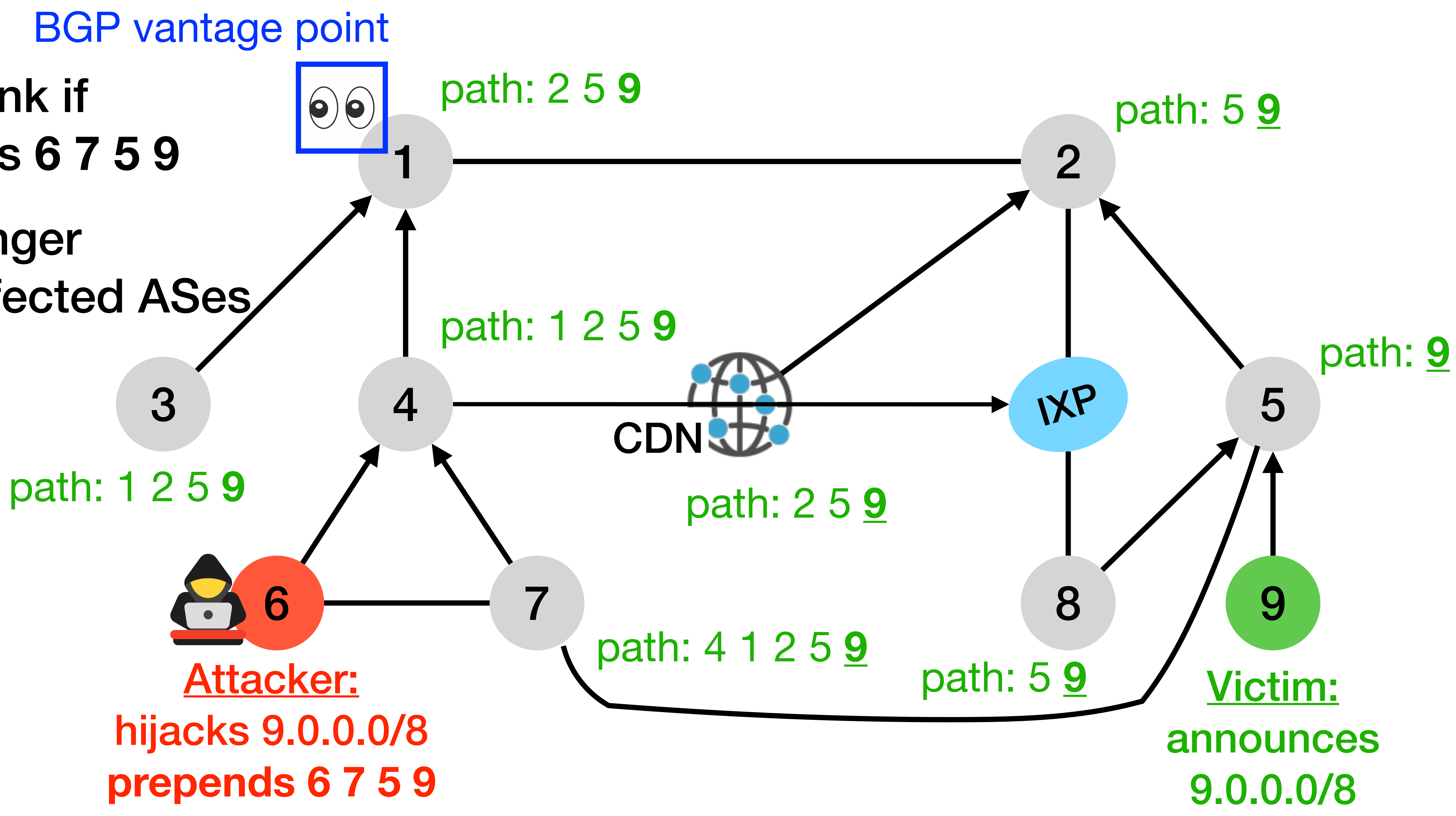


Upon the attack, AS6 (*attacker*) and AS9 (*victim*) appear directly connected

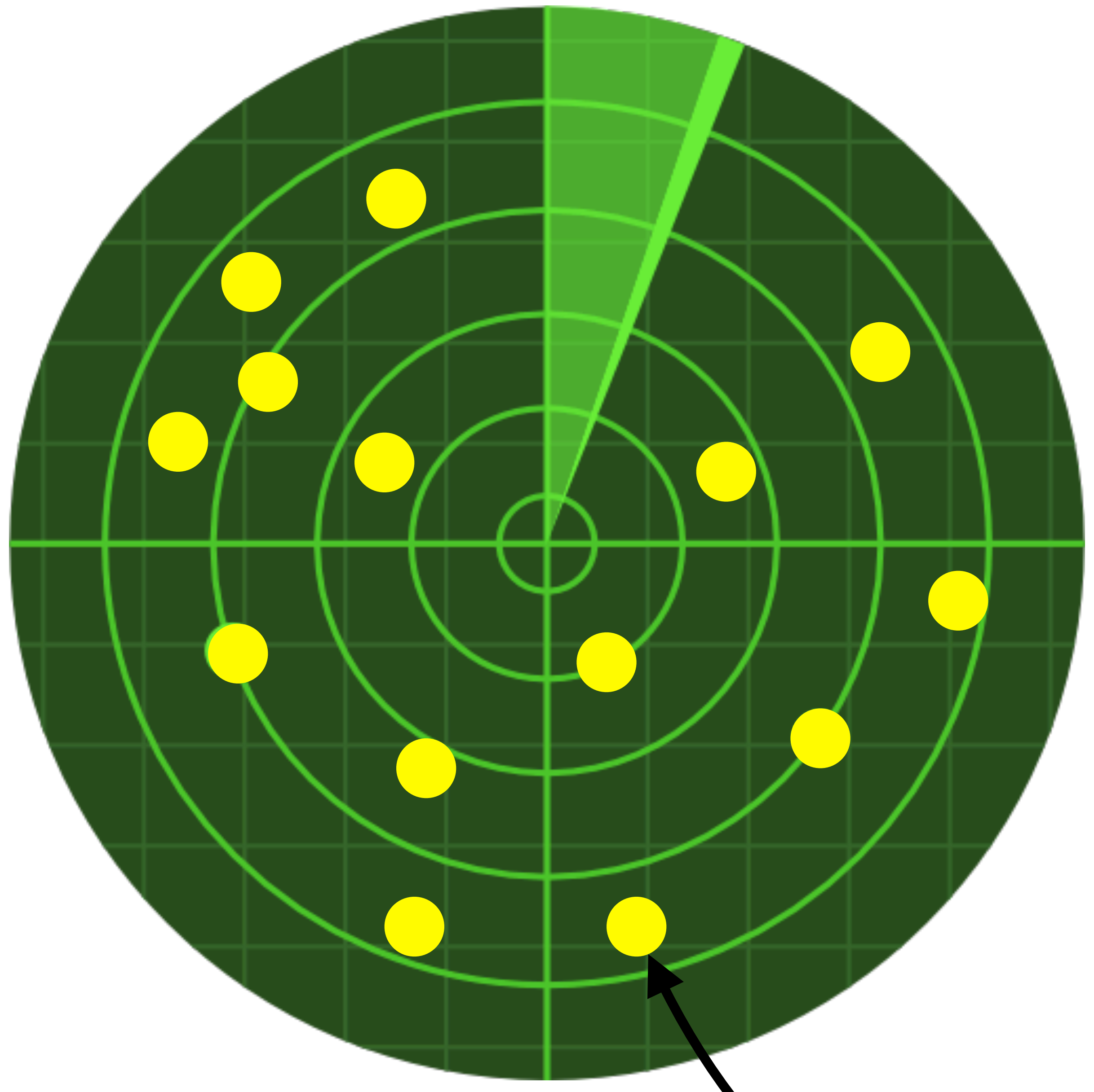
An attacker cannot escape from creating a new AS link without hampering the effectiveness of its attack

There is no new AS link if the attacker prepends **6 7 5 9**

but the AS path is longer making much less infected ASes



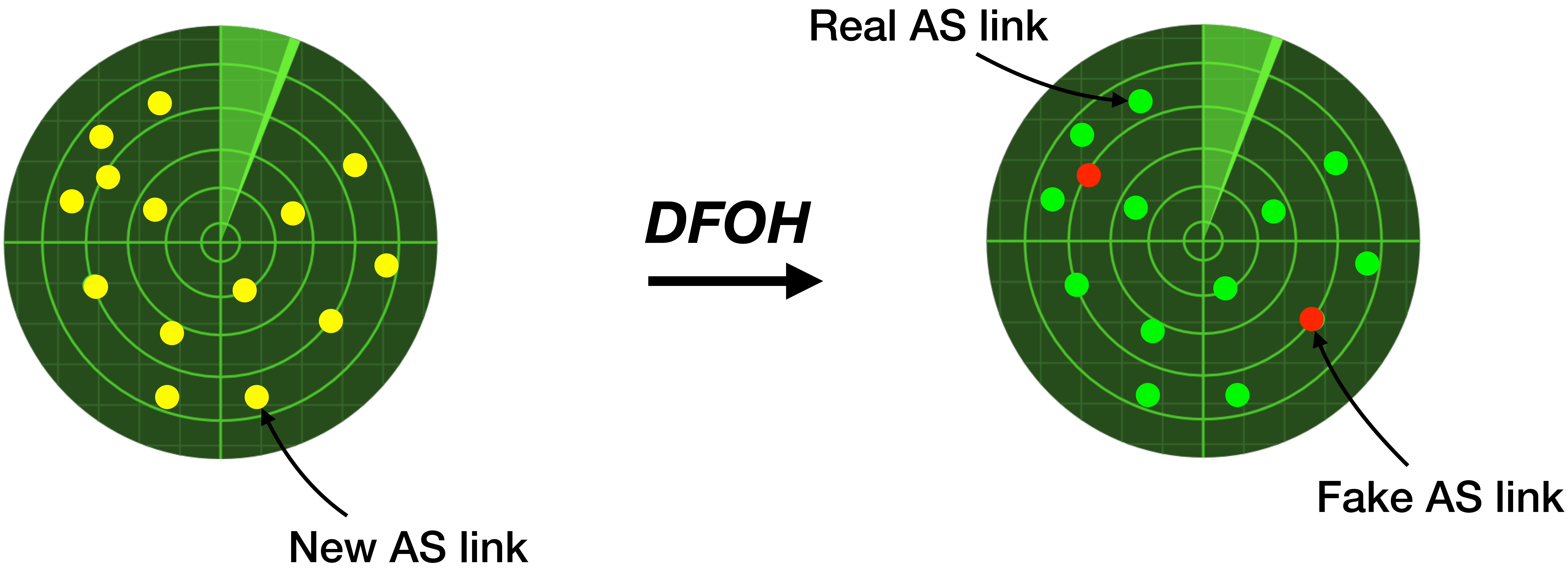
Problem: There are many new AS links every day but no simple property that tells whether they are real or fake


New AS link

We find 166 new AS links every day (median)

Using the BGP data from 200 RIS and RouteViews peers and collected during ten months in 2022

Problem: There are many new AS links every day but no simple property that tells whether they are real or fake



New AS link

DFOH

Real AS link

Fake AS link

Outline

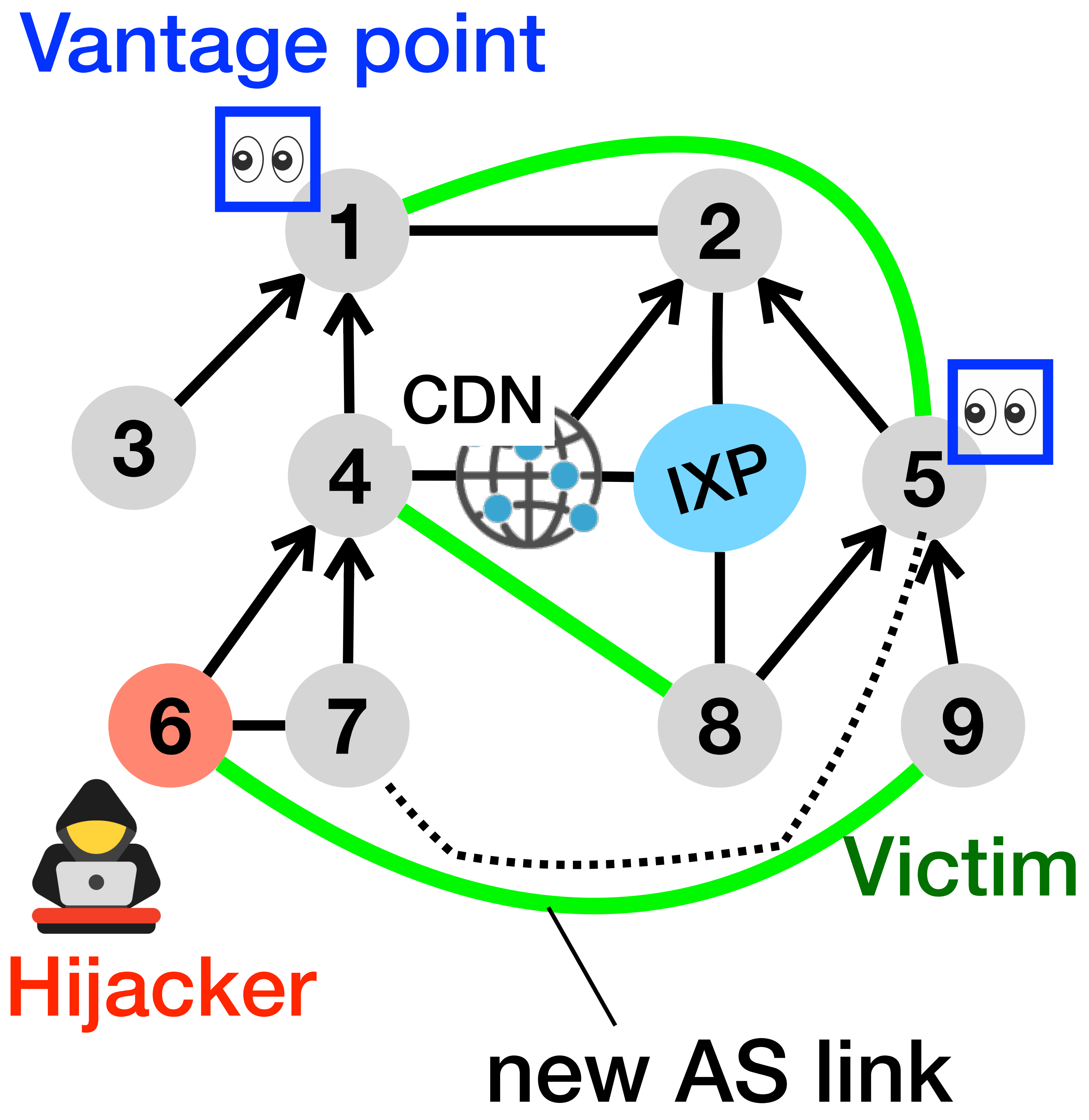
1. ***DFOH***'s main challenge
is to detect fake AS links

2. ***DFOH's*** **key ingredients are**
carefully selected features and a balanced sampling

3. ***DFOH*** is accurate
and practical for users

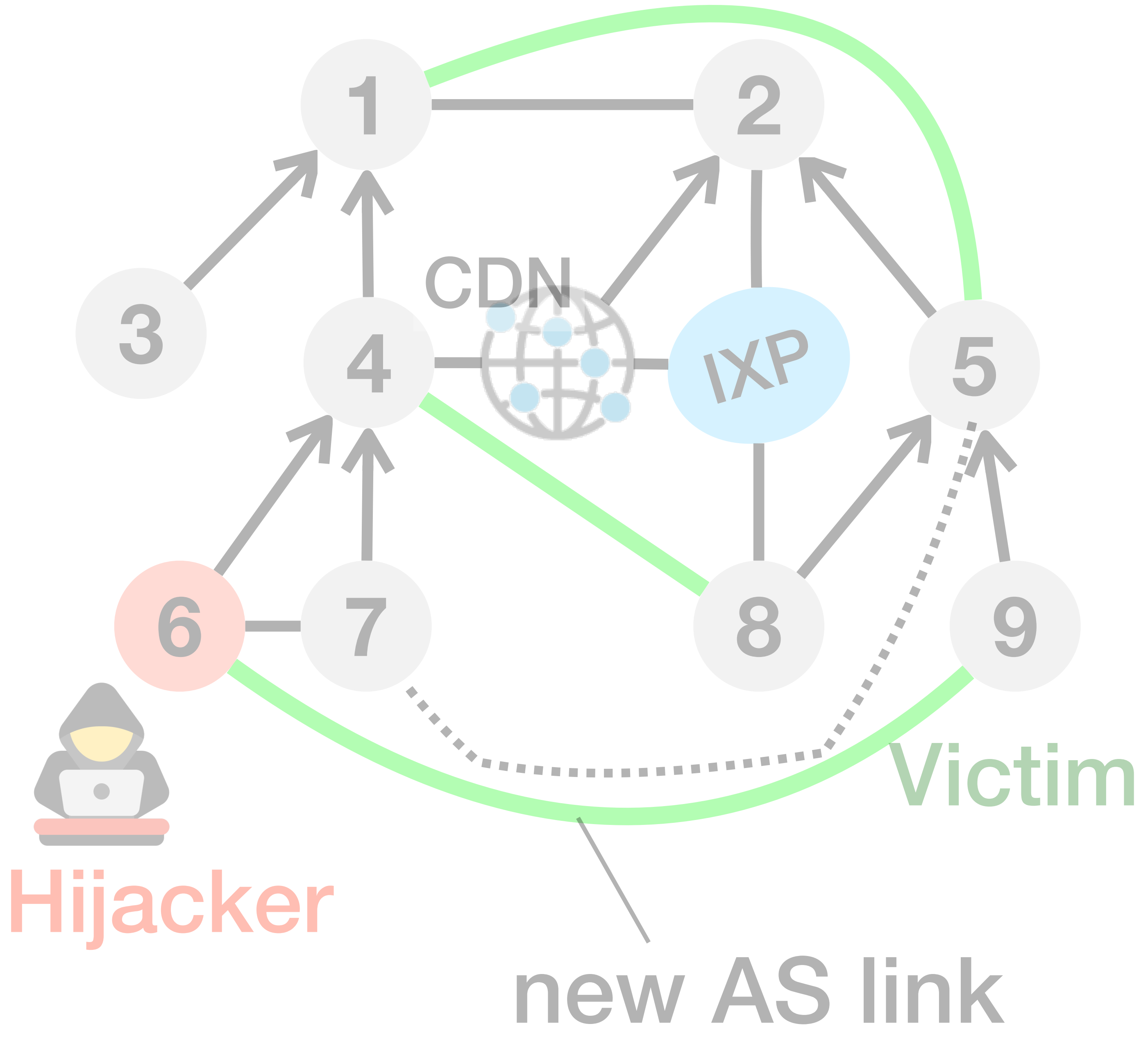
DFOH's fake AS link inference algorithm comprises three steps

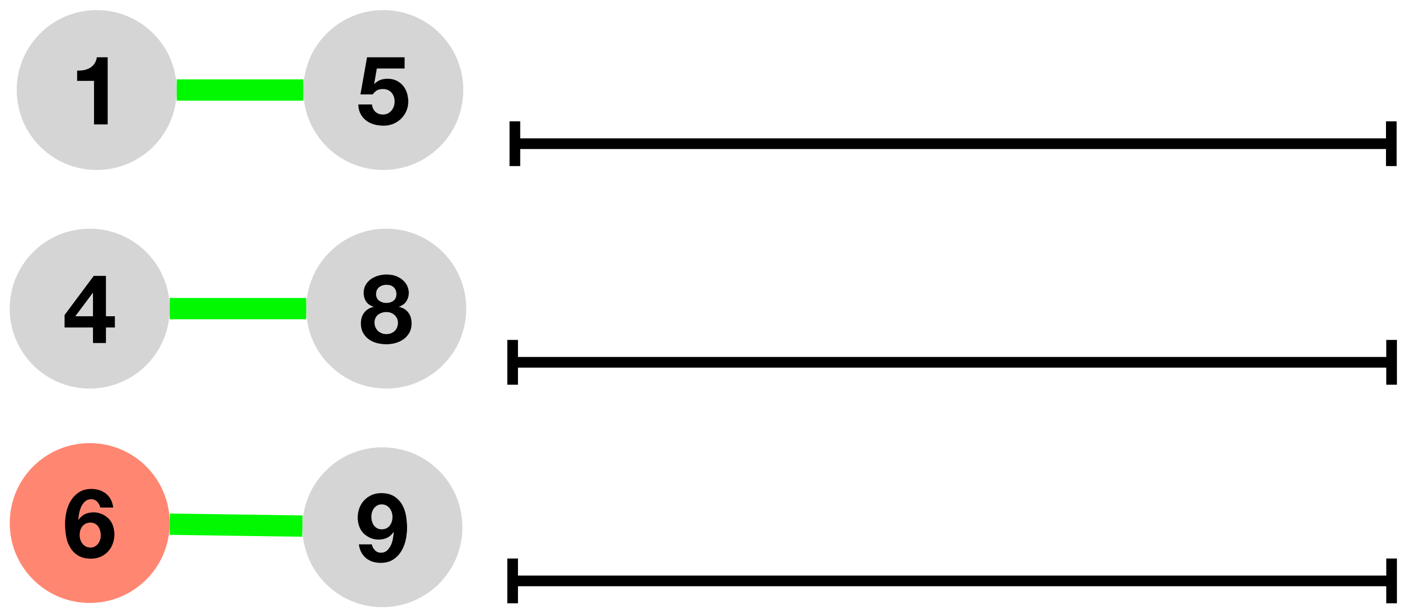
Finding New Links → Computing Features → Inferring Hijacks



DFOH's fake AS links inference algorithm comprises three steps

Finding New Links → Computing Features → Inferring Hijacks

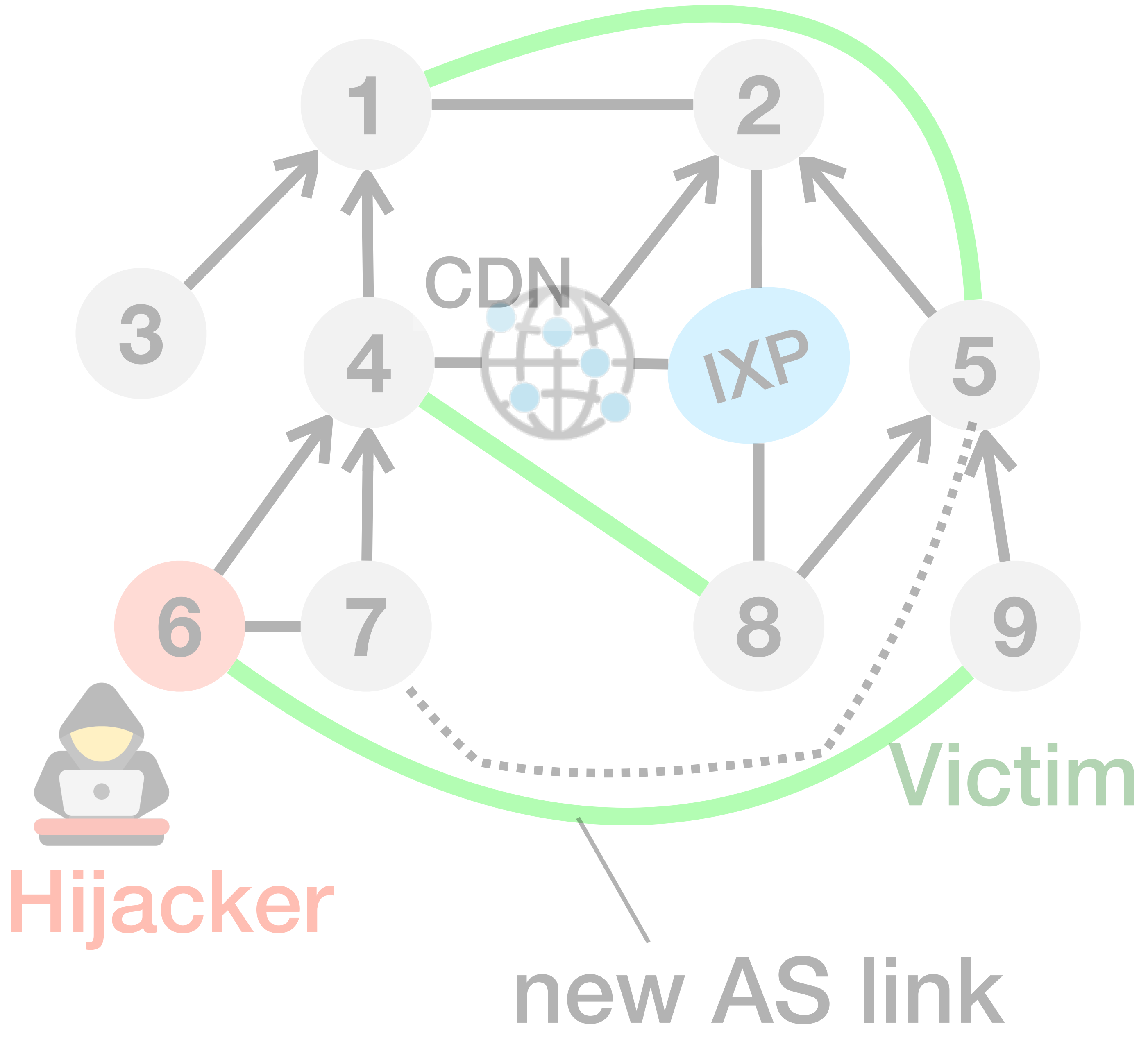




Feature vectors

DFOH’s fake AS links inference algorithm comprises three steps

Finding New Links → Computing Features → Inferring Hijacks



Feature categories:

Feature vectors

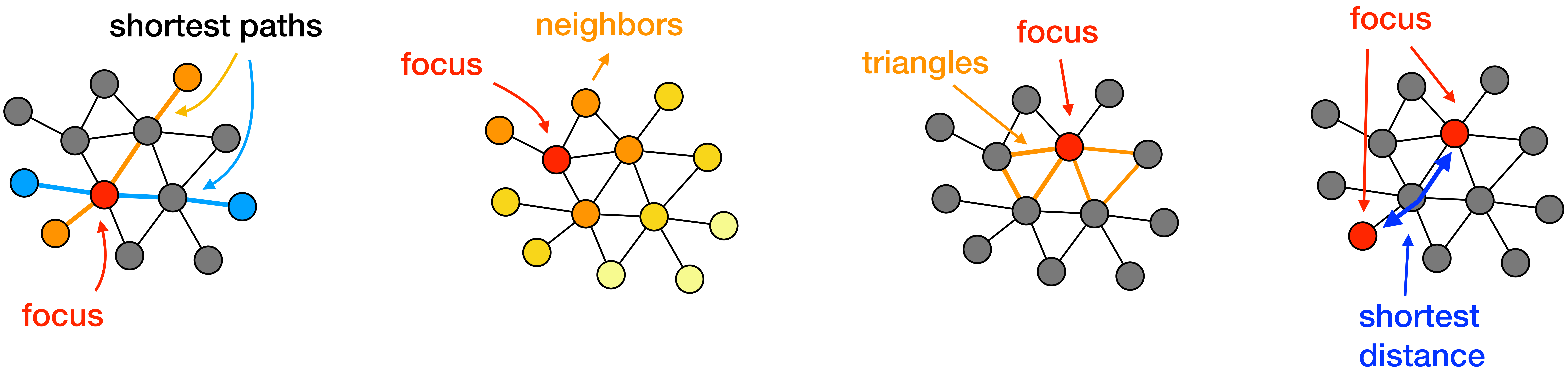
DFOH uses a total of 11 topological features that can be divided into four categories

Node centrality

Neighborhood richness

Topological patterns

Closeness

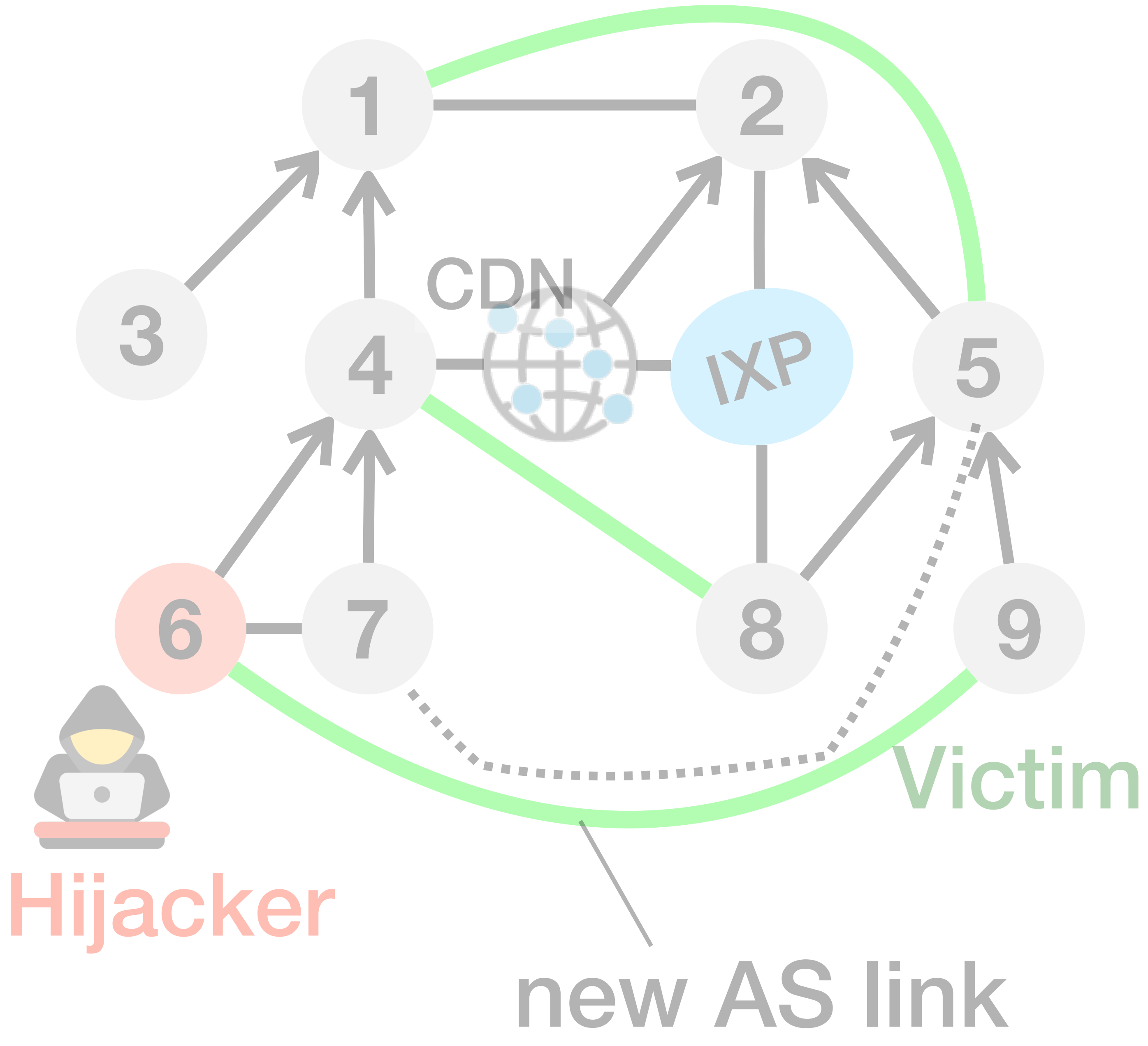


DFOH's fake AS links inference algorithm comprises three steps

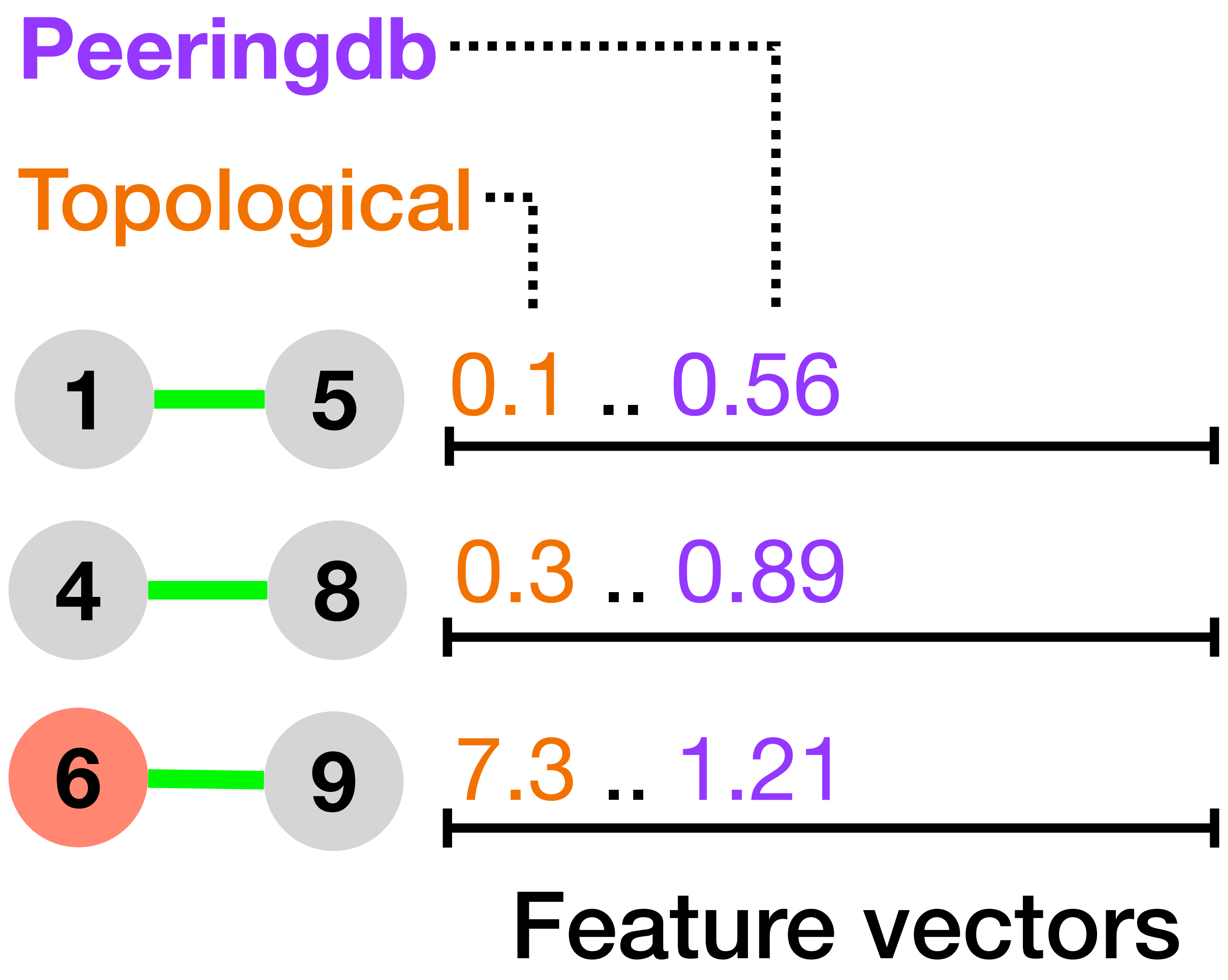
Finding New Links

Computing Features

Inferring Hijacks



Feature categories:



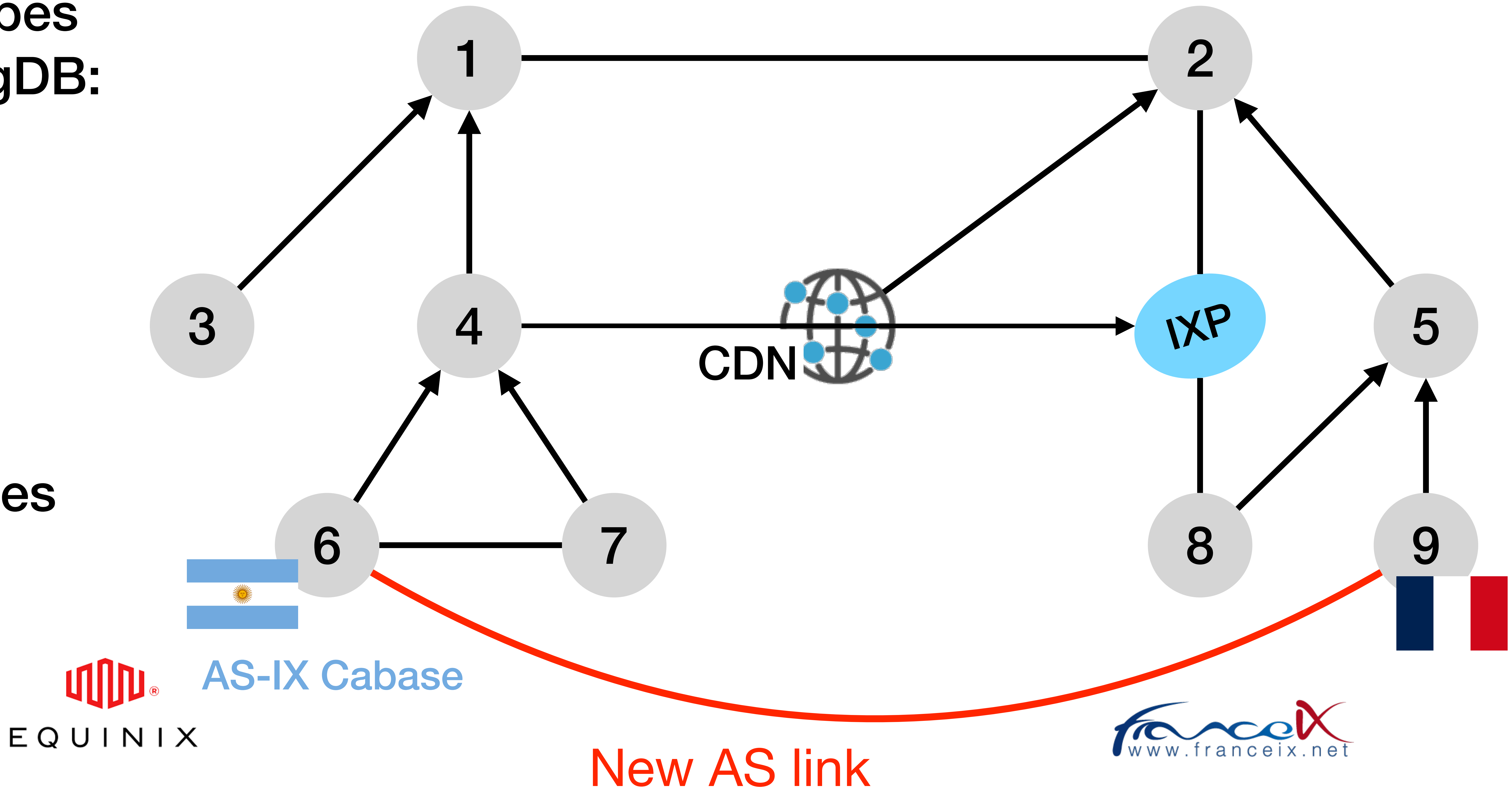
DFOH leverages correlations in the public peering information

DFOH looks for three types of information in PeeringDB:

1. Country

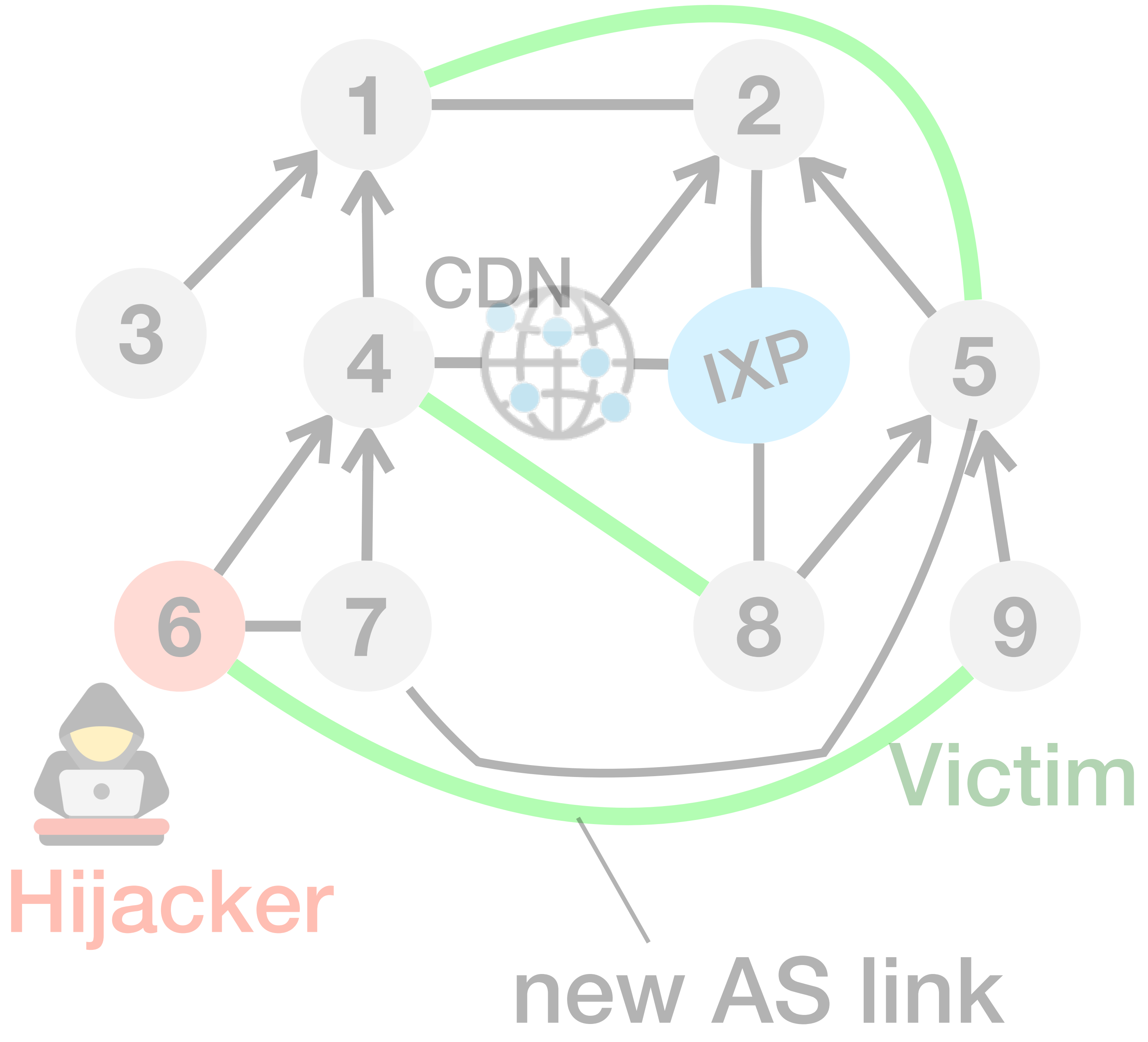
2. Public peering exchange points

3. Private peering facilities

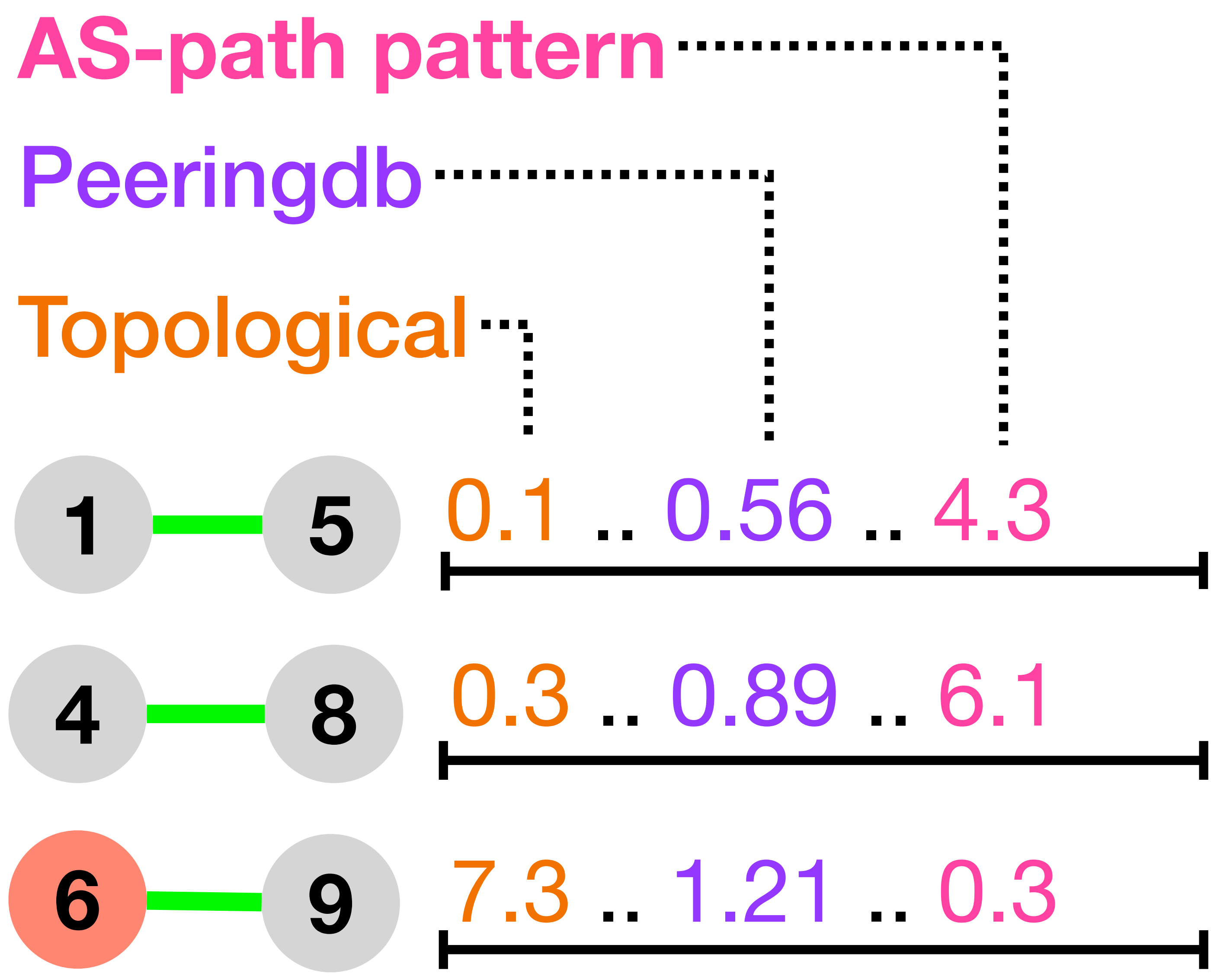


DFOH’s fake AS links inference algorithm comprises three steps

Finding New Links → Computing Features → Inferring Hijacks

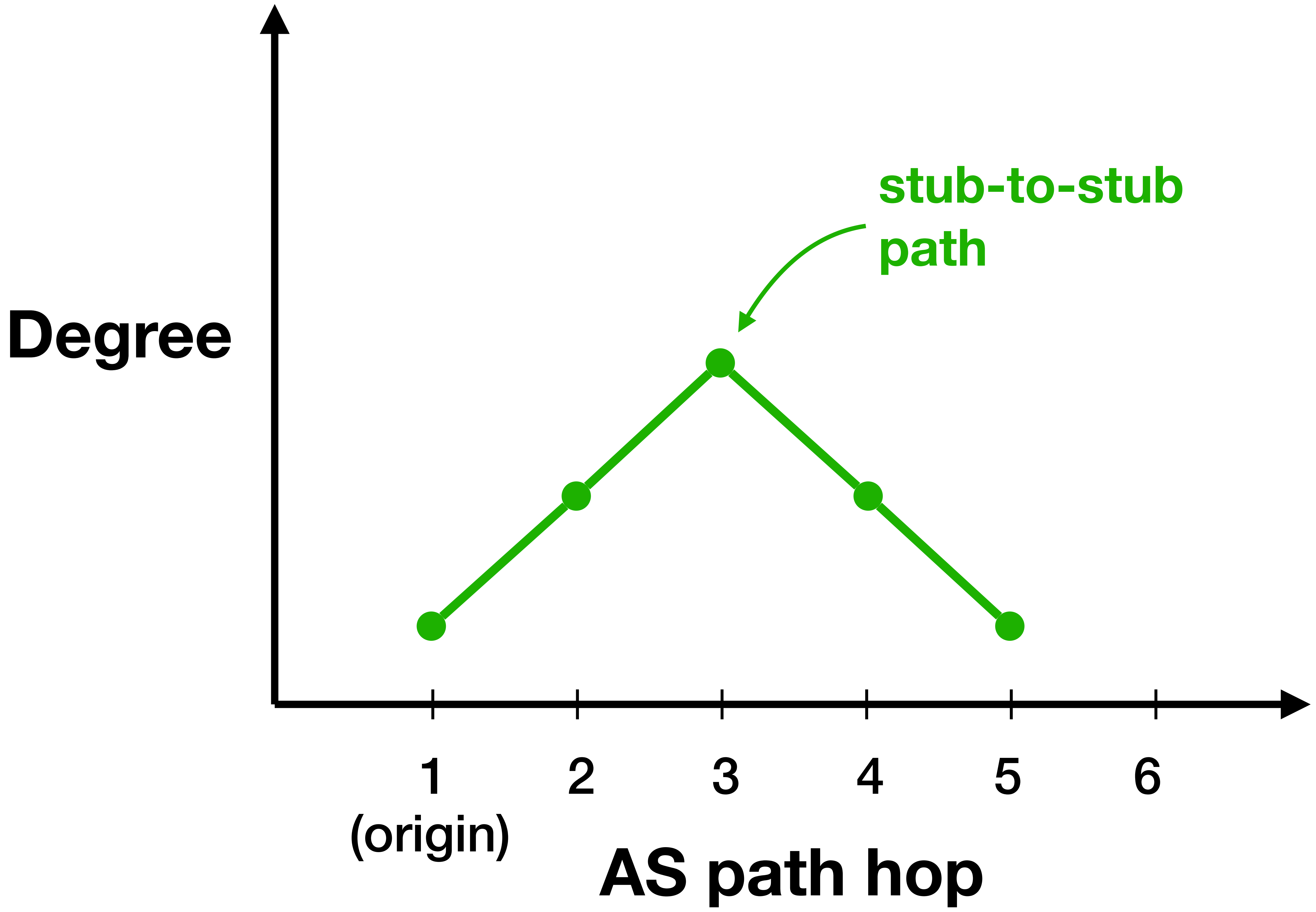


Feature categories:

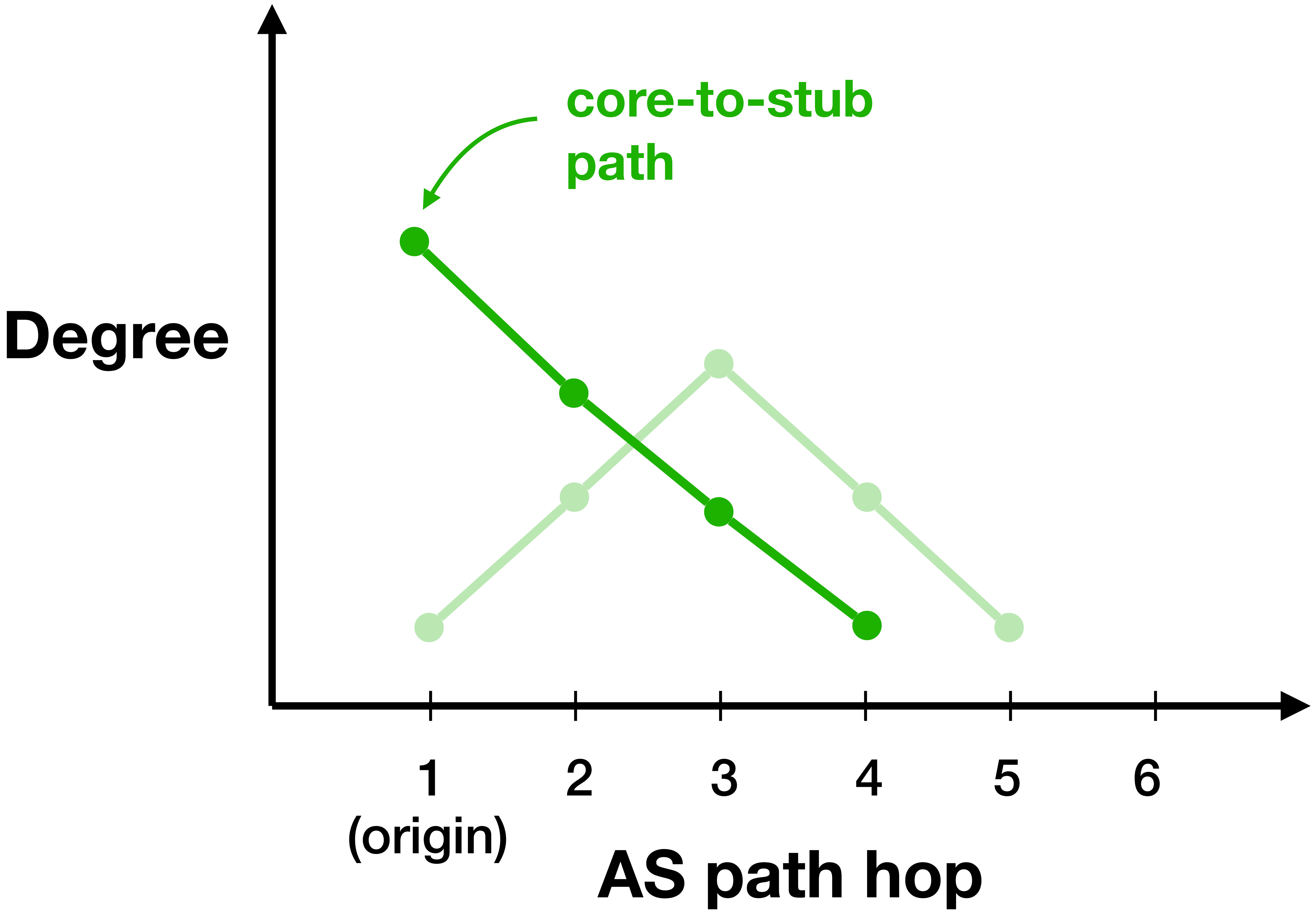


Feature vectors

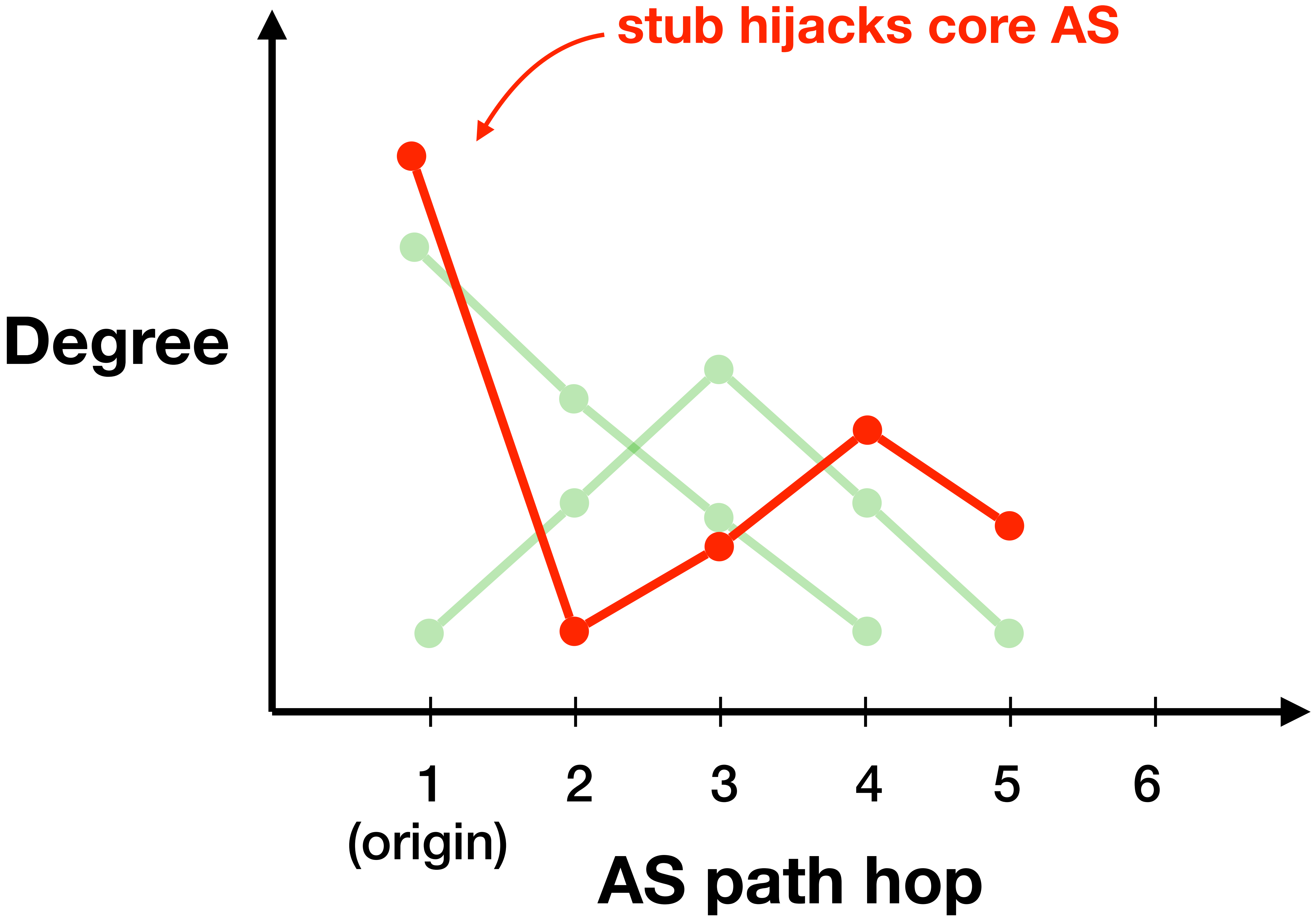
DFOH detects fake AS paths as they often violate patterns induced by business relationships



DFOH detects fake AS paths as they often violate patterns induced by business relationships

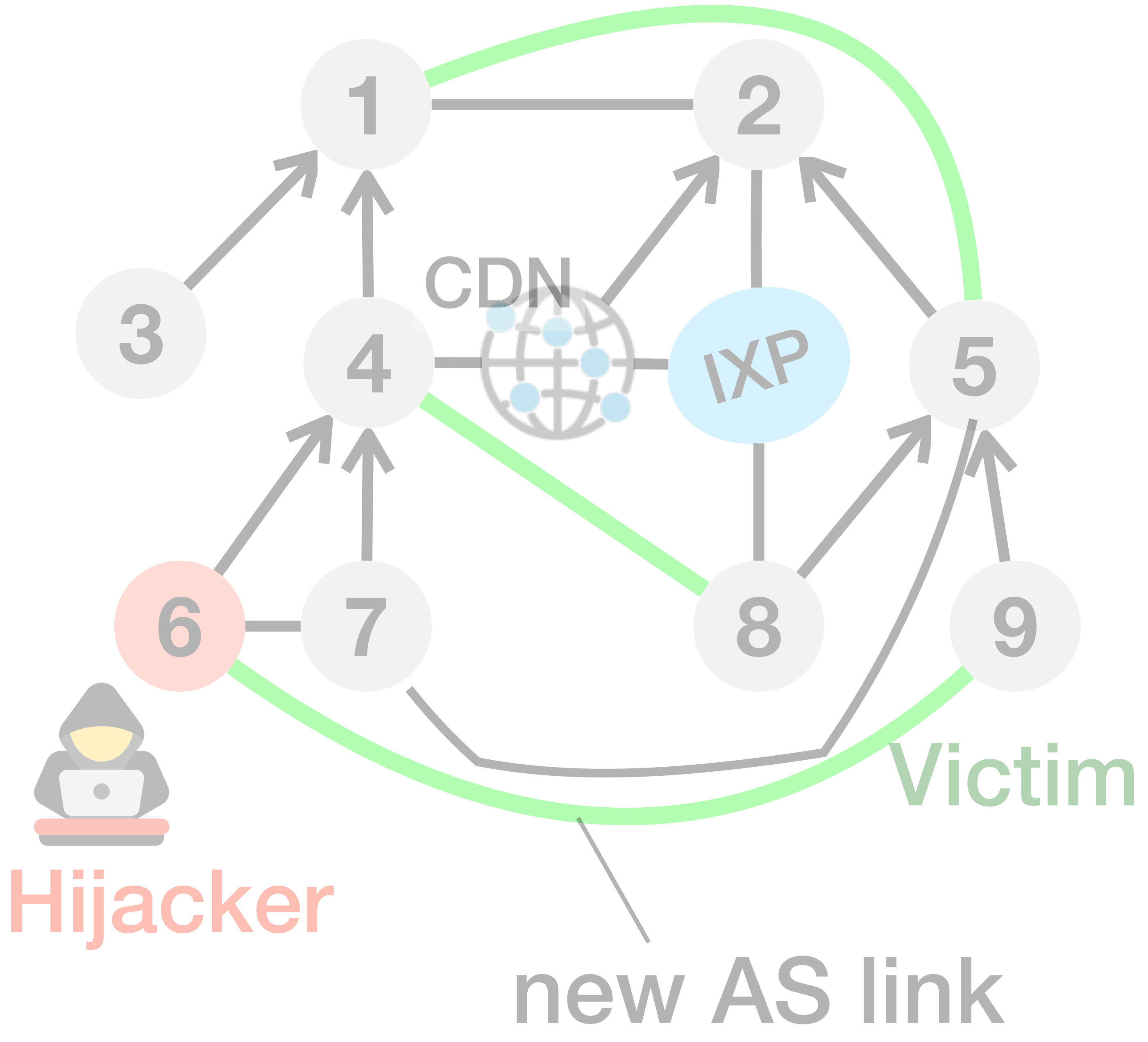


DFOH detects fake AS paths as they often violate patterns induced by business relationships



DFOH's fake AS links inference algorithm comprises three steps

Finding New Links → Computing Features → Inferring Hijacks



Feature categories:

Bidirectionality

AS-path pattern

Peeringdb

Topological

1 — 5: 0.1 .. 0.56 .. 4.3 .. 6

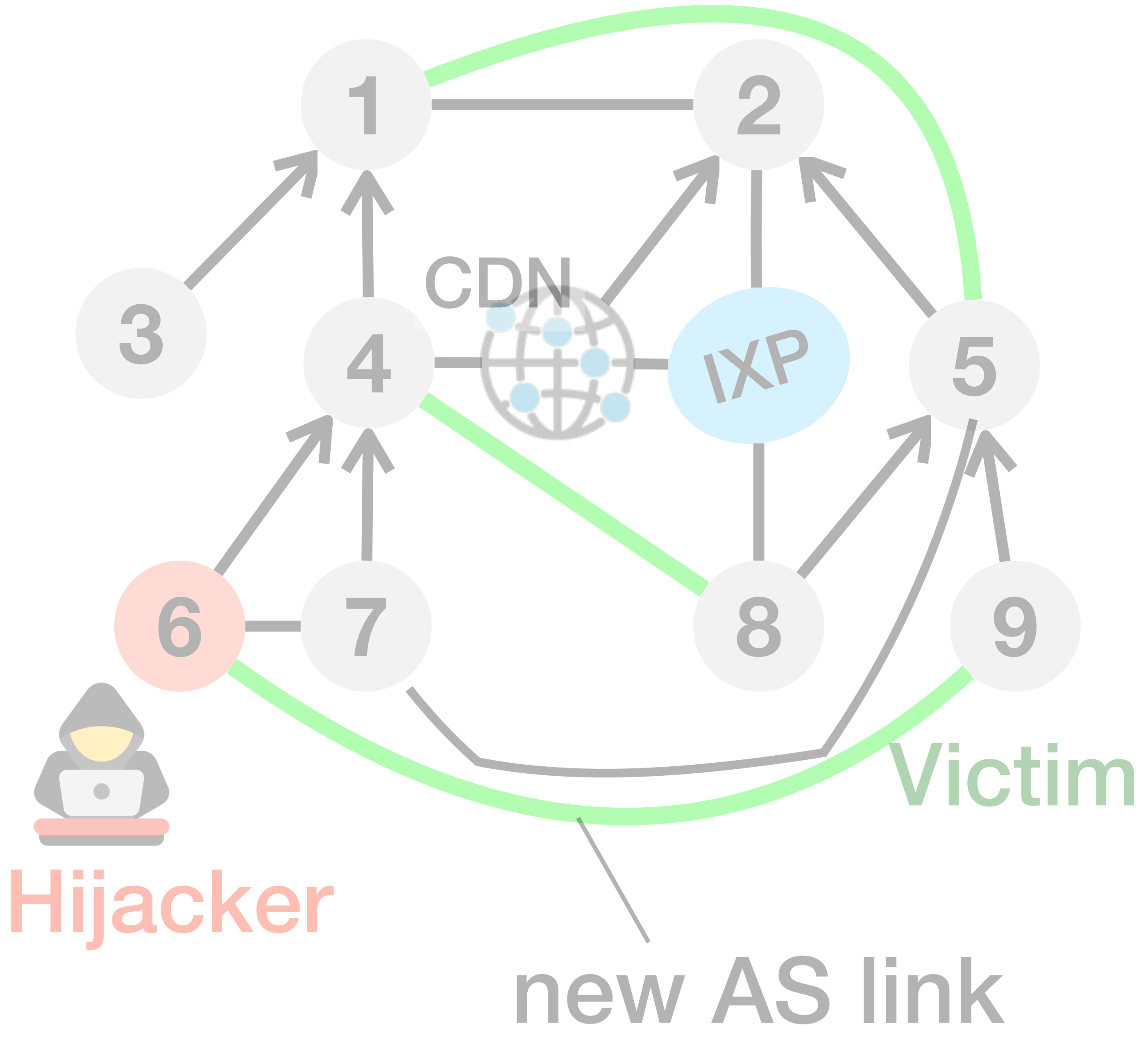
4 — 8: 0.3 .. 0.89 .. 6.1 .. 0

6 — 9: 7.3 .. 1.21 .. 0.3 .. 8

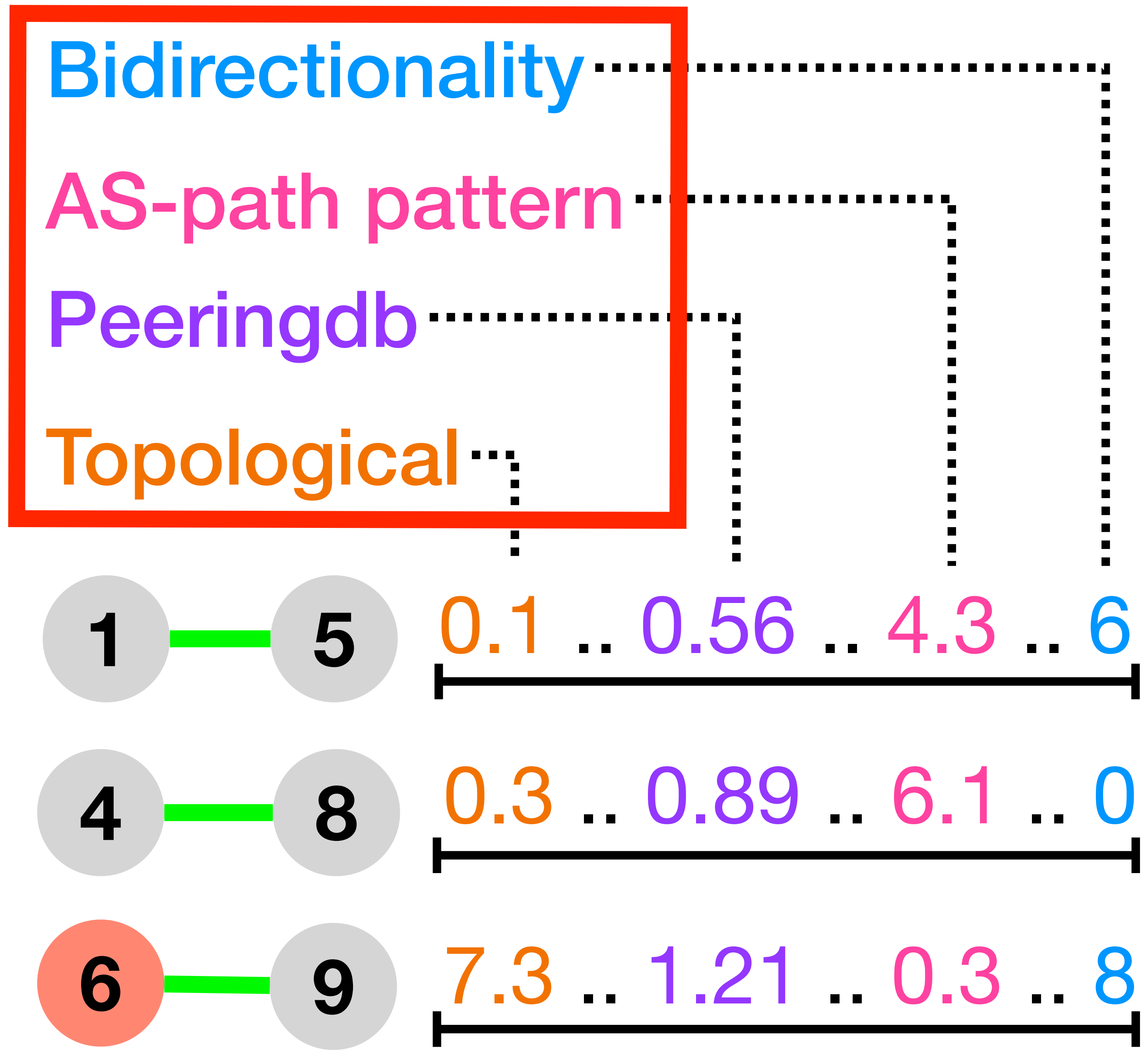
Feature vectors

DFOH's fake AS links inference algorithm comprises three steps

→ Finding New Links → Computing Features → Inferring Hijacks



Key ingredient #1



1 — 5: 0.1 .. 0.56 .. 4.3 .. 6

4 — 8: 0.3 .. 0.89 .. 6.1 .. 0

6 — 9: 7.3 .. 1.21 .. 0.3 .. 8

Feature vectors

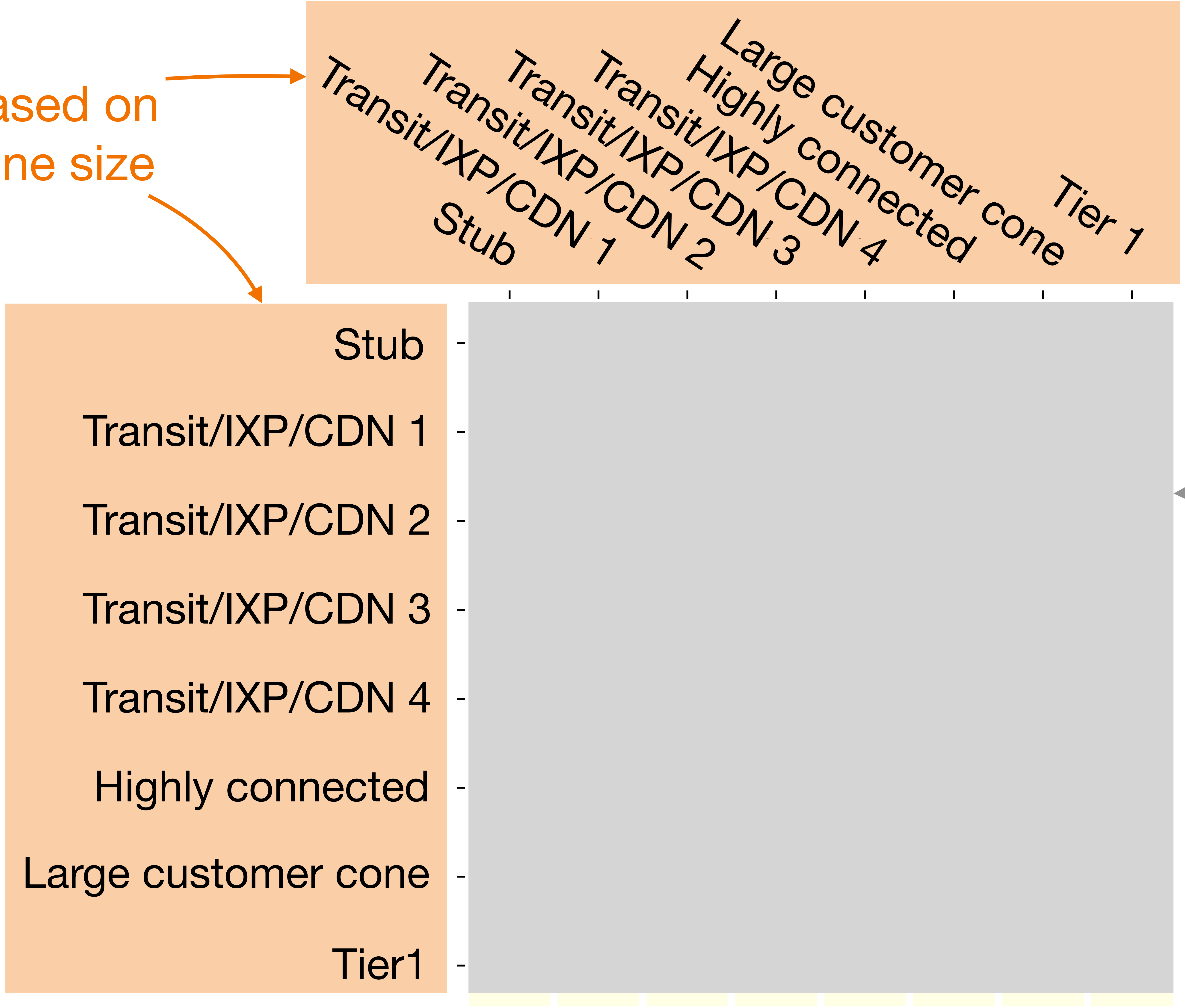
DFOH's fake AS links inference algorithm comprises three steps

Finding New Links → Computing Features → Inferring Hijacks

Problem: randomly sampling **nonexistent** links makes DFOH skewed towards stub-to-stub links as they are overrepresented

<u>Problem:</u> randomly sampling **nonexistent** links makes DFOH skewed towards stub-to-stub links as they are overrepresented

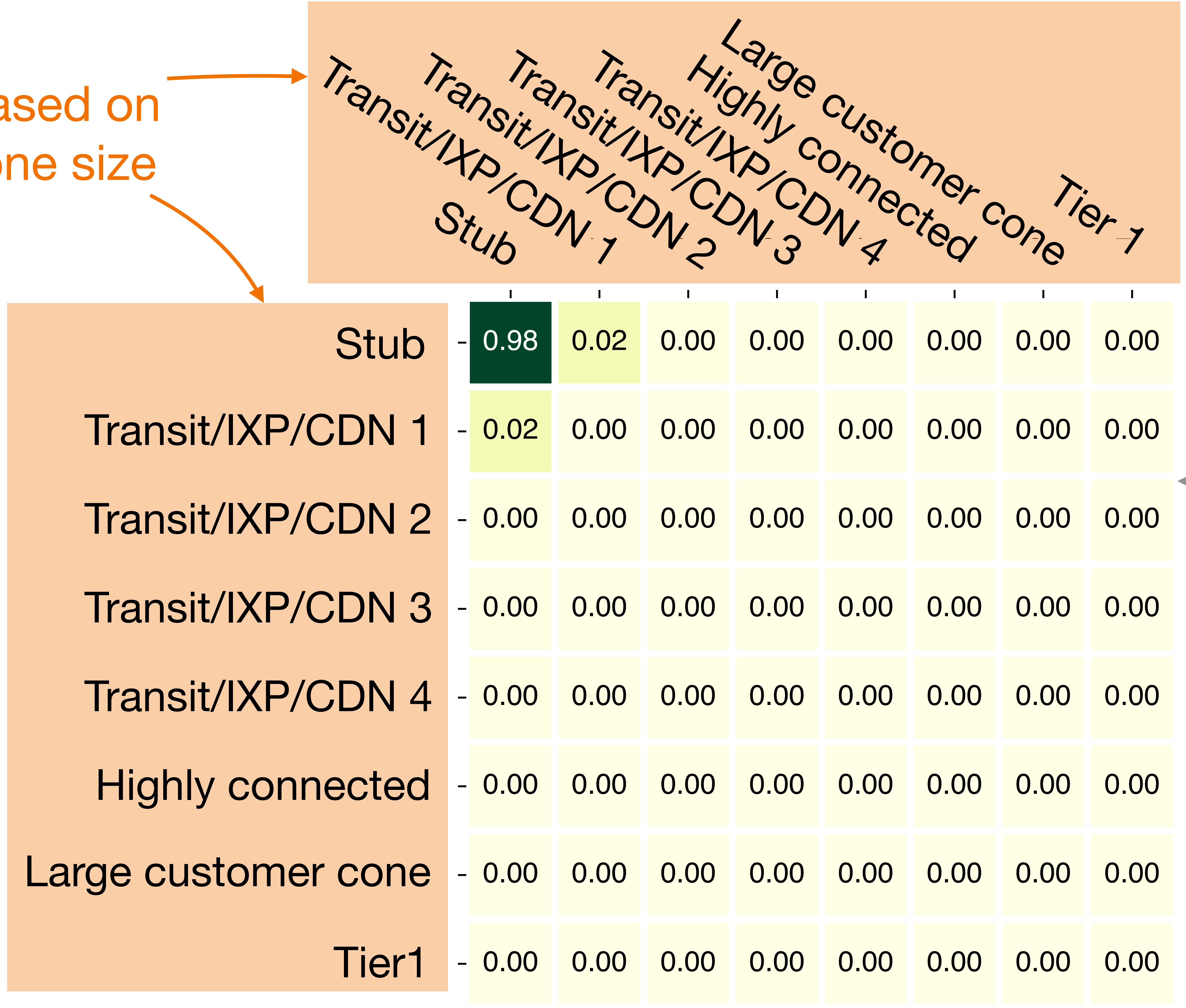
Clusters of ASes based on their degree and cone size



Proportion of sampled **nonexistent** AS links *(random sampling)*

Problem: randomly sampling **nonexistent** links makes DFOH skewed towards stub-to-stub links as they are overrepresented

Clusters of ASes based on their degree and cone size



	Stub	Transit/IXP/CDN 1	Transit/IXP/CDN 2	Transit/IXP/CDN 3	Transit/IXP/CDN 4	Highly connected	Large customer cone	Tier 1
Stub	0.98	0.02	0.00	0.00	0.00	0.00	0.00	0.00
Transit/IXP/CDN 1	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Transit/IXP/CDN 2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Transit/IXP/CDN 3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Transit/IXP/CDN 4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Highly connected	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Large customer cone	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Tier1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Proportion of sampled **nonexistent** AS links *(random sampling)*

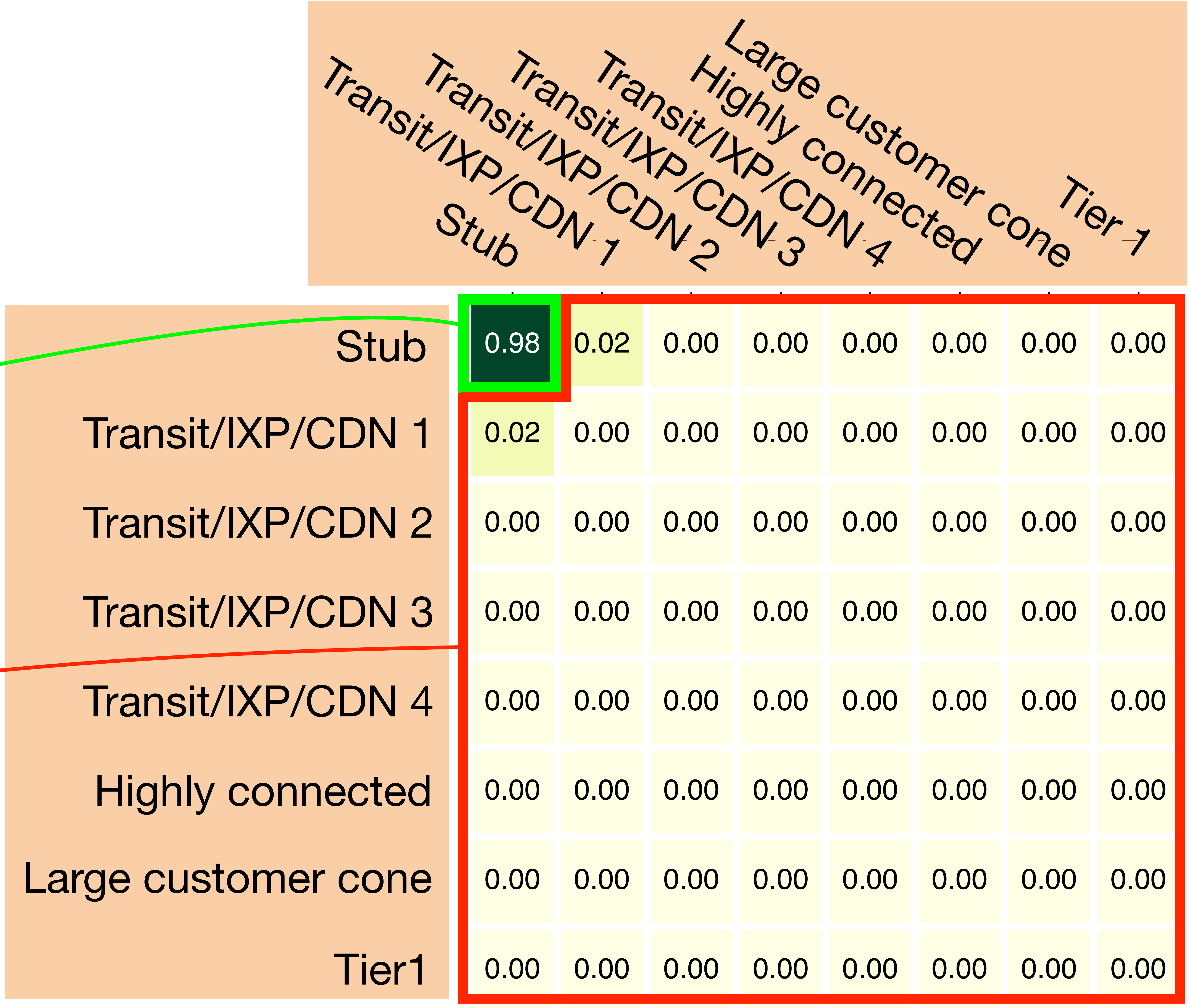
Problem: randomly sampling **nonexistent** links makes DFOH skewed towards stub-to-stub links as they are overrepresented

DFOH would perform well on scenarios involving two stubs

	Stub	Transit/IXP/CDN 1	Transit/IXP/CDN 2	Transit/IXP/CDN 3	Transit/IXP/CDN 4	Highly connected	Large customer cone	Tier 1
Stub	0.98	0.02	0.00	0.00	0.00	0.00	0.00	0.00
Transit/IXP/CDN 1	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Transit/IXP/CDN 2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Transit/IXP/CDN 3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Transit/IXP/CDN 4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Highly connected	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Large customer cone	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Tier1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Proportion of sampled **nonexistent** AS links *(random sampling)*

Problem: randomly sampling **nonexistent** links makes DFOH skewed towards stub-to-stub links as they are overrepresented



	Stub	Transit/IXP/CDN 1	Transit/IXP/CDN 2	Transit/IXP/CDN 3	Transit/IXP/CDN 4	Highly connected	Large customer cone	Tier 1
Stub	0.98	0.02	0.00	0.00	0.00	0.00	0.00	0.00
Transit/IXP/CDN 1	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Transit/IXP/CDN 2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Transit/IXP/CDN 3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Transit/IXP/CDN 4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Highly connected	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Large customer cone	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Tier1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

DFOH would perform well on scenarios involving two stubs

But not on the other scenarios

Proportion of sampled **nonexistent** AS links *(random sampling)*

DFOH's fake AS links inference algorithm comprises three steps

Outline

1. ***DFOH***’s main challenge
is to detect fake AS links

2. ***DFOH’s*** key ingredients are
carefully selected features and a balanced sampling

3. ***DFOH*** is accurate
and practical for users

We evaluate ***DFOH*** on artificially created forged-origin hijacks and measure its accuracy upon every attack scenario

Methodology:

Step #1: We take existing AS paths
and prepend a new origin to create a new link

Step #2: We consider 9k cases where the new link exists (*legitimate cases*)
and 9k cases where the new link does not existent (*malicious cases*)

We evaluate ***DFOH*** on artificially created forged-origin hijacks and measure its accuracy upon every attack scenario

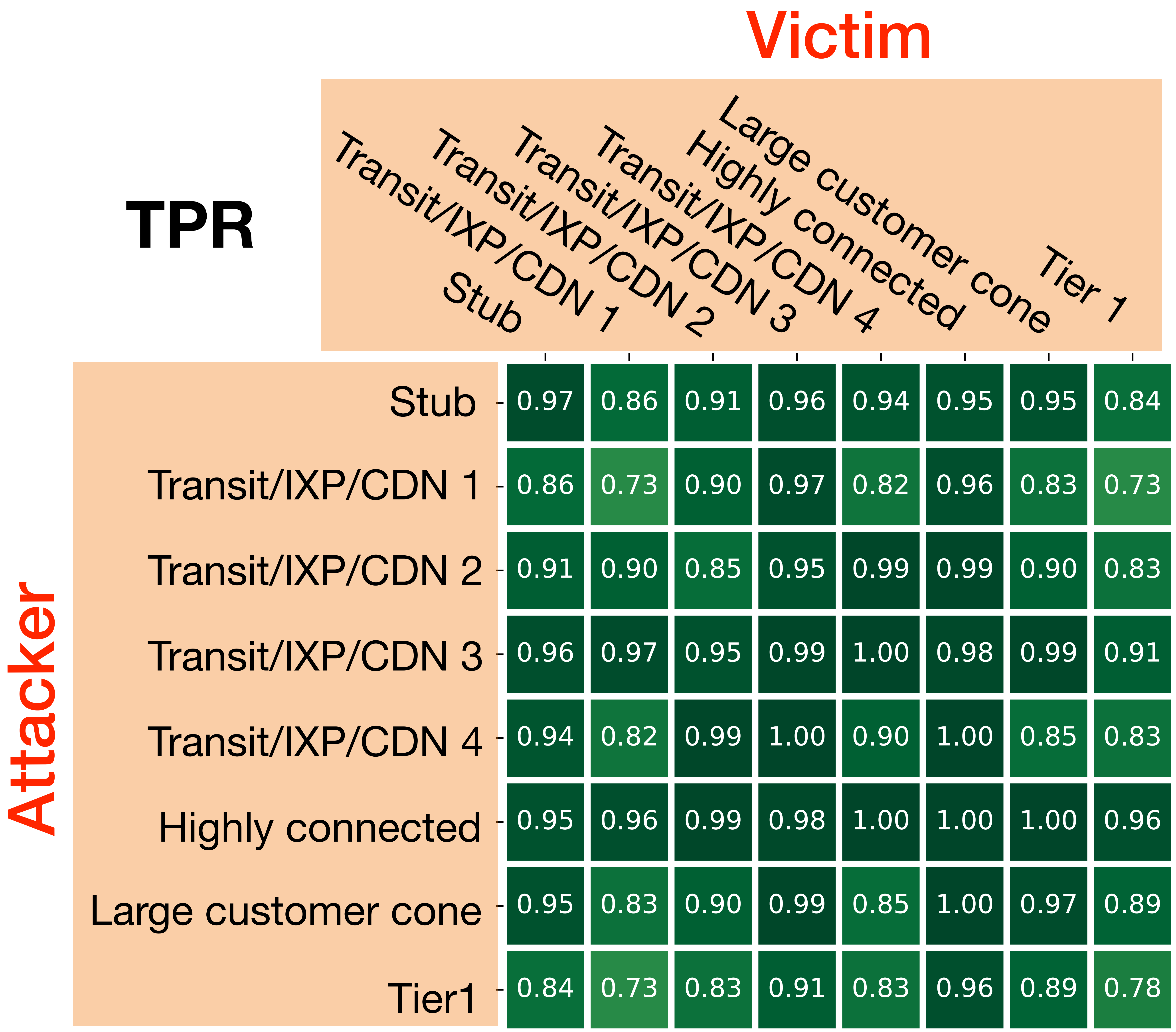
Methodology:

Step #1: We take existing AS paths
and prepend a new origin to create a new link

Step #2: We consider 9k cases where the new link exists (*legitimate cases*)
and 9k cases where the new link does not existent (*malicious cases*)

We focus on the **True Positive Rate** (TPR)
and the **False Positive Rate** (FPR)

DFOH is accurate upon every attack scenario

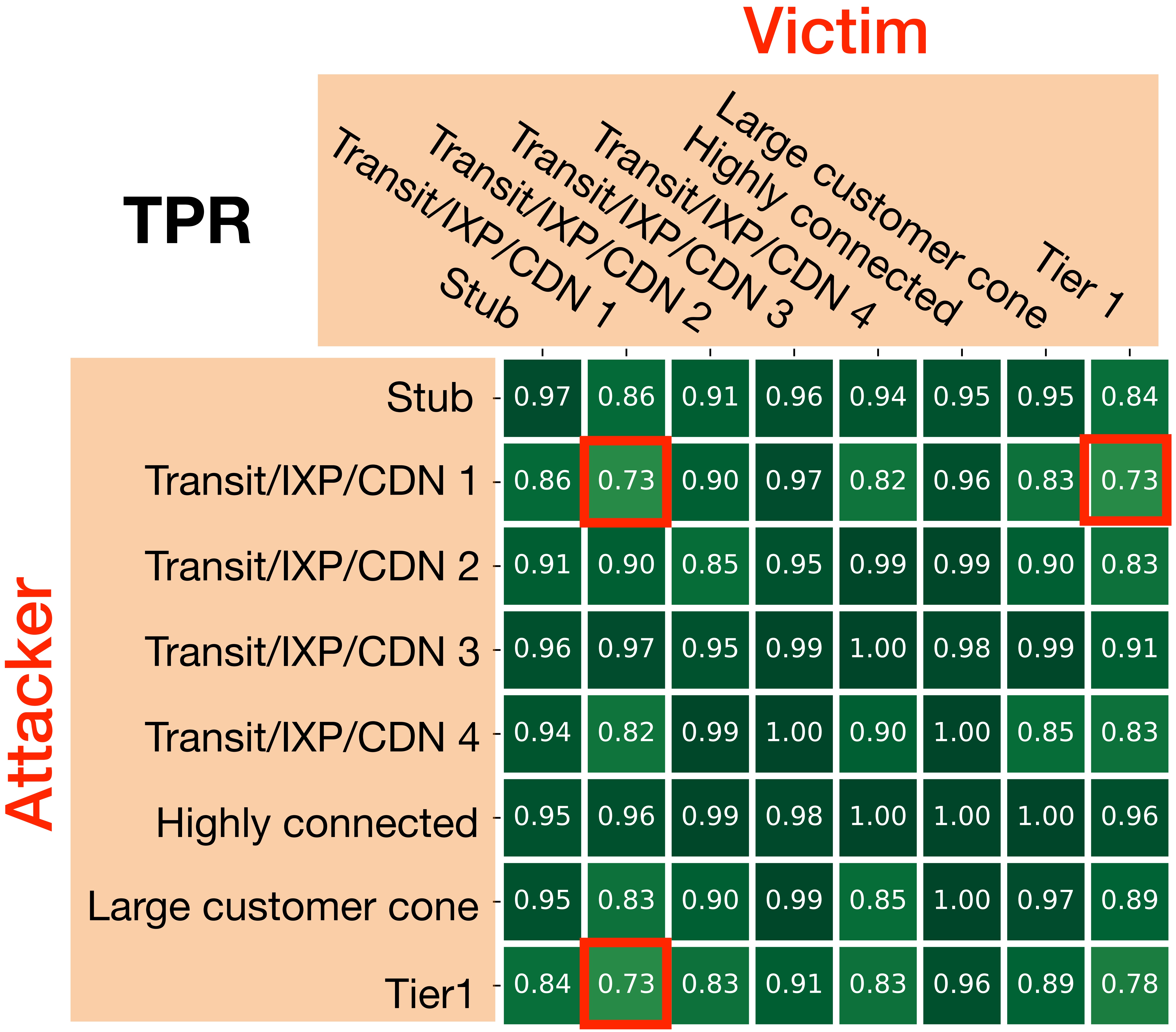


Victim

Attacker

TPR	Stub	Transit/IXP/CDN 1	Transit/IXP/CDN 2	Transit/IXP/CDN 3	Transit/IXP/CDN 4	Highly connected	Large customer cone	Tier 1
Stub	0.97	0.86	0.91	0.96	0.94	0.95	0.95	0.84
Transit/IXP/CDN 1	0.86	0.73	0.90	0.97	0.82	0.96	0.83	0.73
Transit/IXP/CDN 2	0.91	0.90	0.85	0.95	0.99	0.99	0.90	0.83
Transit/IXP/CDN 3	0.96	0.97	0.95	0.99	1.00	0.98	0.99	0.91
Transit/IXP/CDN 4	0.94	0.82	0.99	1.00	0.90	1.00	0.85	0.83
Highly connected	0.95	0.96	0.99	0.98	1.00	1.00	1.00	0.96
Large customer cone	0.95	0.83	0.90	0.99	0.85	1.00	0.97	0.89
Tier1	0.84	0.73	0.83	0.91	0.83	0.96	0.89	0.78

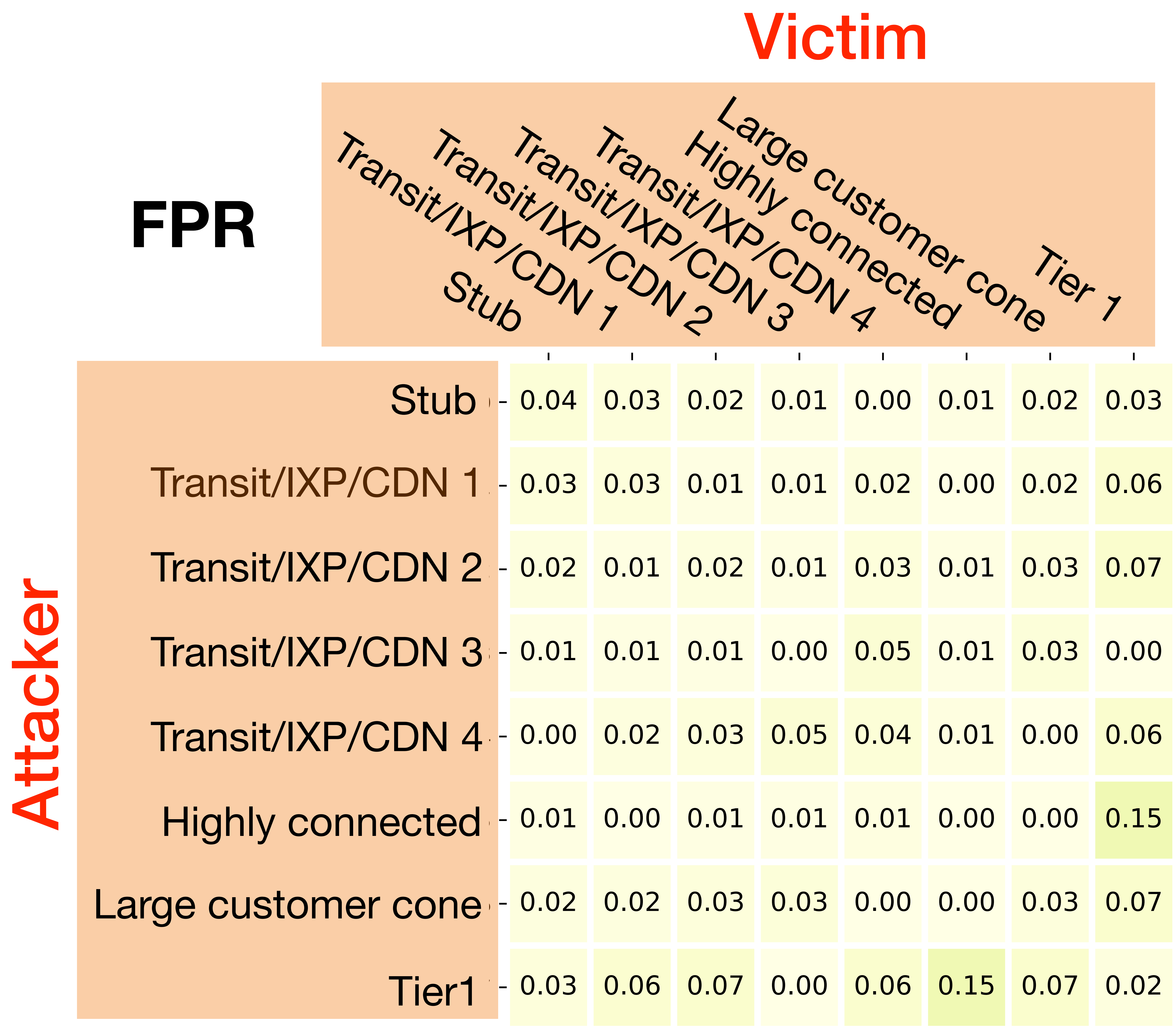
DFOH is accurate upon every attack scenario



TPR	Stub	Transit/IXP/CDN 1	Transit/IXP/CDN 2	Transit/IXP/CDN 3	Transit/IXP/CDN 4	Highly connected	Large customer cone	Tier 1
Stub	0.97	0.86	0.91	0.96	0.94	0.95	0.95	0.84
Transit/IXP/CDN 1	0.86	0.73	0.90	0.97	0.82	0.96	0.83	0.73
Transit/IXP/CDN 2	0.91	0.90	0.85	0.95	0.99	0.99	0.90	0.83
Transit/IXP/CDN 3	0.96	0.97	0.95	0.99	1.00	0.98	0.99	0.91
Transit/IXP/CDN 4	0.94	0.82	0.99	1.00	0.90	1.00	0.85	0.83
Highly connected	0.95	0.96	0.99	0.98	1.00	1.00	1.00	0.96
Large customer cone	0.95	0.83	0.90	0.99	0.85	1.00	0.97	0.89
Tier1	0.84	0.73	0.83	0.91	0.83	0.96	0.89	0.78

The minimum TPR is 0.73

DFOH is accurate upon every attack scenario



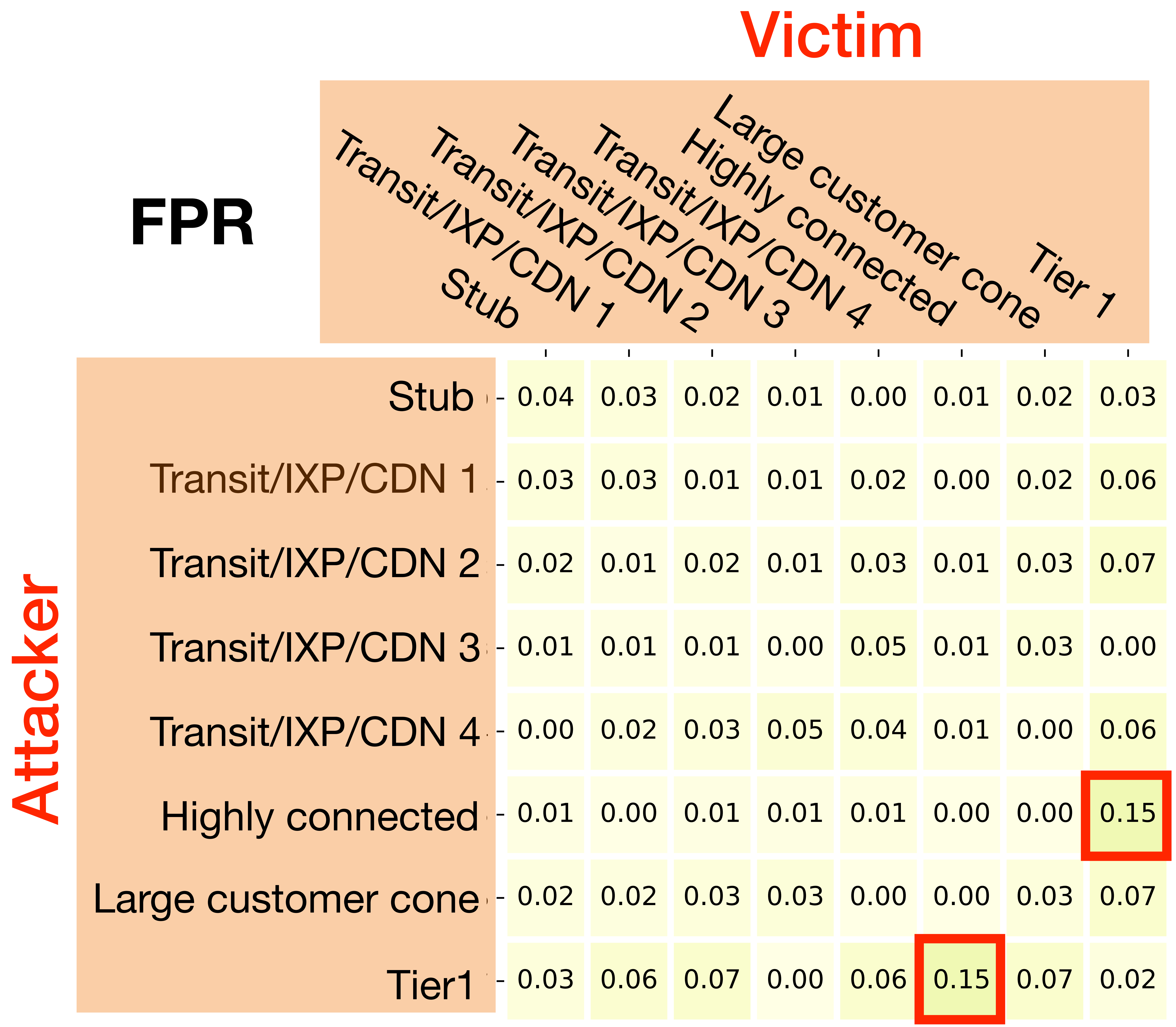
FPR

Victim

Attacker

	Stub	Transit/IXP/CDN 1	Transit/IXP/CDN 2	Transit/IXP/CDN 3	Transit/IXP/CDN 4	Highly connected	Large customer cone	Tier 1
Stub	0.04	0.03	0.02	0.01	0.00	0.01	0.02	0.03
Transit/IXP/CDN 1	0.03	0.03	0.01	0.01	0.02	0.00	0.02	0.06
Transit/IXP/CDN 2	0.02	0.01	0.02	0.01	0.03	0.01	0.03	0.07
Transit/IXP/CDN 3	0.01	0.01	0.01	0.00	0.05	0.01	0.03	0.00
Transit/IXP/CDN 4	0.00	0.02	0.03	0.05	0.04	0.01	0.00	0.06
Highly connected	0.01	0.00	0.01	0.01	0.01	0.00	0.00	0.15
Large customer cone	0.02	0.02	0.03	0.03	0.00	0.00	0.03	0.07
Tier1	0.03	0.06	0.07	0.00	0.06	0.15	0.07	0.02

DFOH is accurate upon every attack scenario

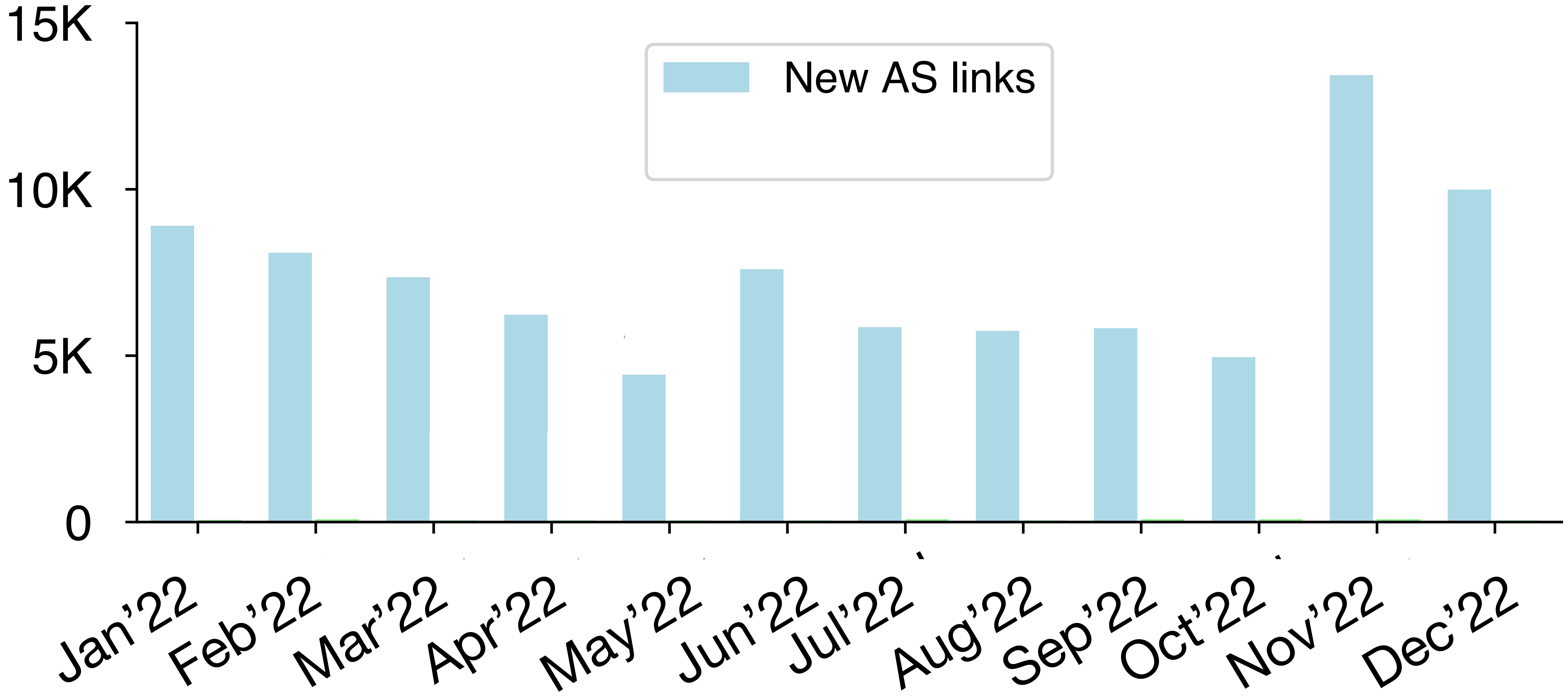


	Stub	Transit/IXP/CDN 1	Transit/IXP/CDN 2	Transit/IXP/CDN 3	Transit/IXP/CDN 4	Highly connected	Large customer cone	Tier 1
Stub	0.04	0.03	0.02	0.01	0.00	0.01	0.02	0.03
Transit/IXP/CDN 1	0.03	0.03	0.01	0.01	0.02	0.00	0.02	0.06
Transit/IXP/CDN 2	0.02	0.01	0.02	0.01	0.03	0.01	0.03	0.07
Transit/IXP/CDN 3	0.01	0.01	0.01	0.00	0.05	0.01	0.03	0.00
Transit/IXP/CDN 4	0.00	0.02	0.03	0.05	0.04	0.01	0.00	0.06
Highly connected	0.01	0.00	0.01	0.01	0.01	0.00	0.00	0.15
Large customer cone	0.02	0.02	0.03	0.03	0.00	0.00	0.03	0.07
Tier1	0.03	0.06	0.07	0.00	0.06	0.15	0.07	0.02

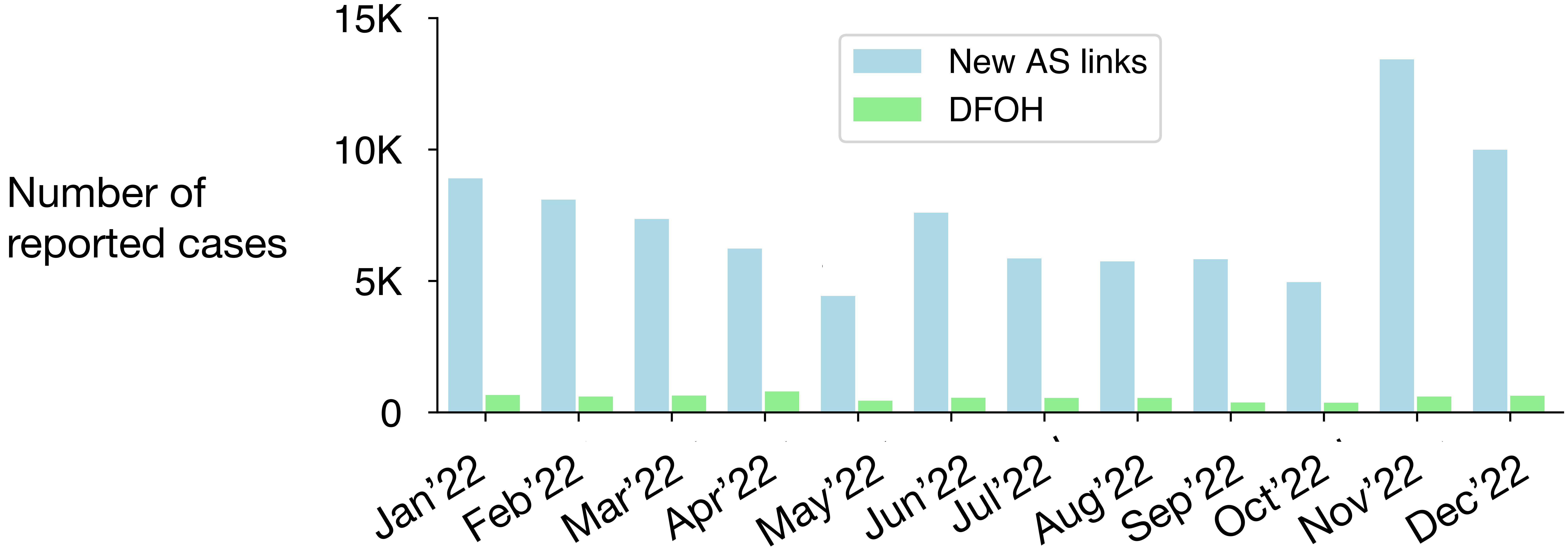
The maximum FPR is 0.15

DFOH makes the detection of forged-origin hijacks practical for operators

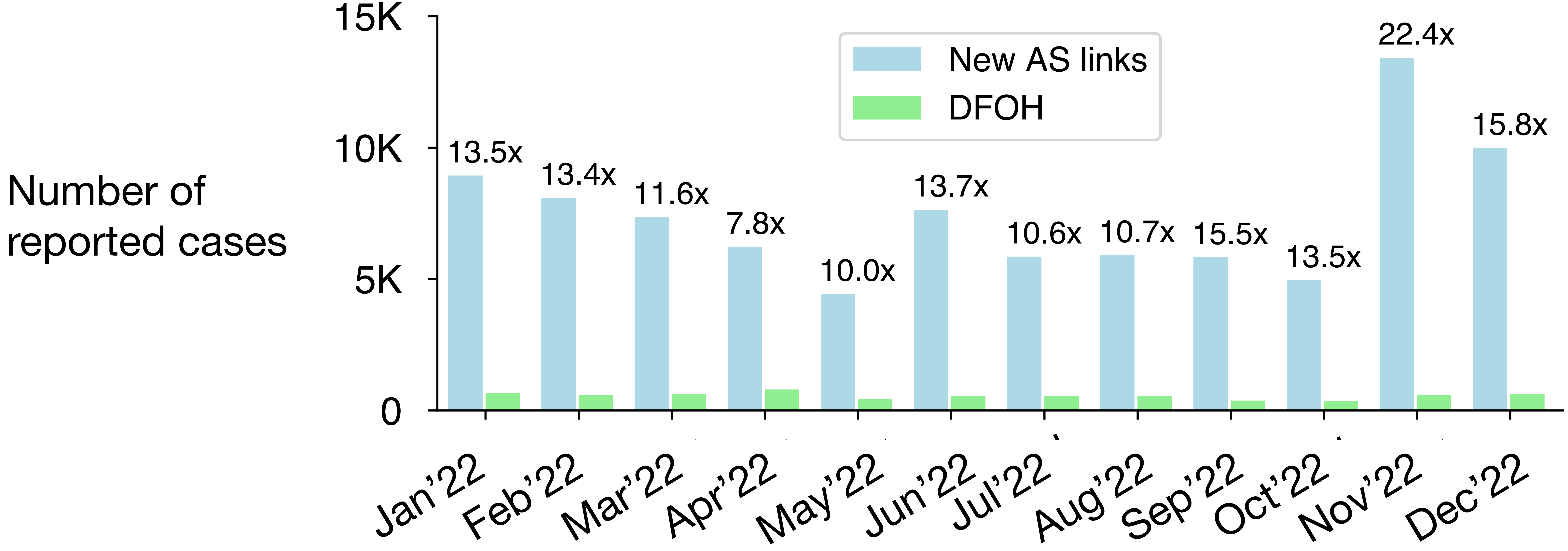
Number of reported cases



DFOH makes the detection of forged-origin hijacks practical for operators

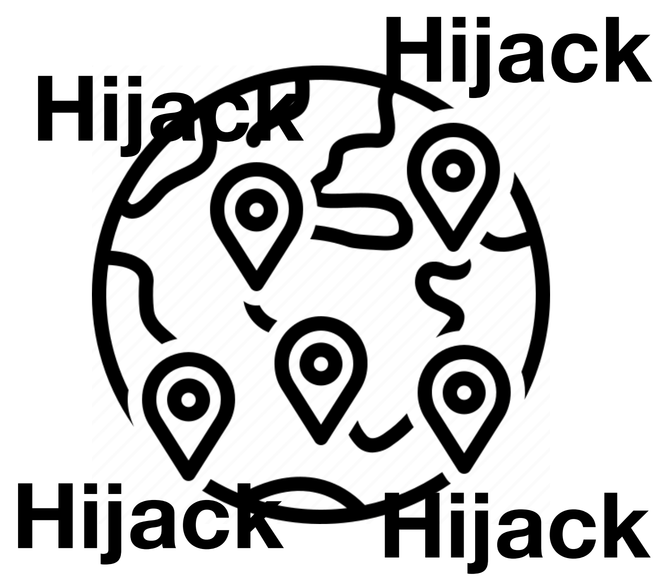


DFOH makes the detection of forged-origin hijacks practical for operators

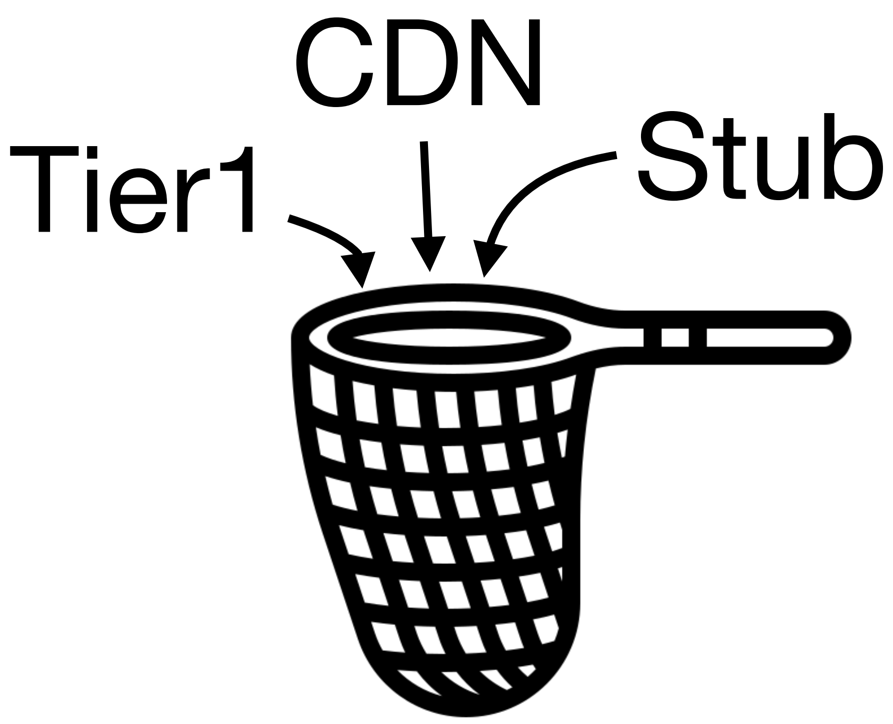


DFOH: A System to Detect Forged-Origin Hijacks

DFOH runs in a commodity server



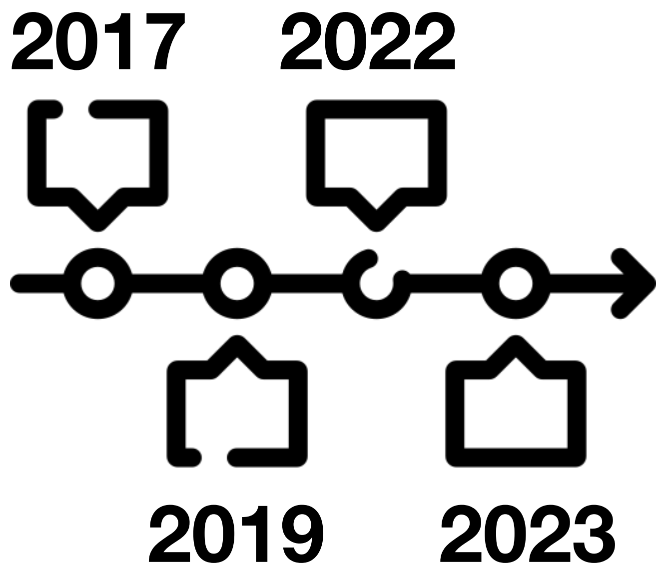
DFOH detects hijacks on the whole Internet



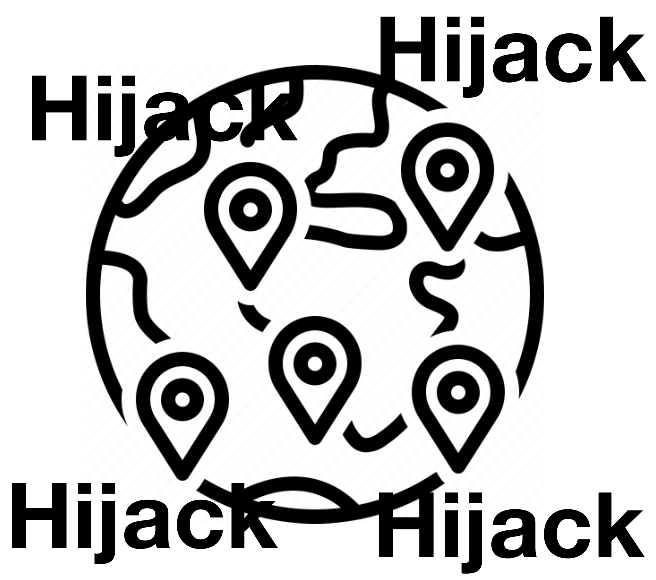
DFOH is accurate in every attack scenario

DFOH: A System to Detect Forged-Origin Hijacks

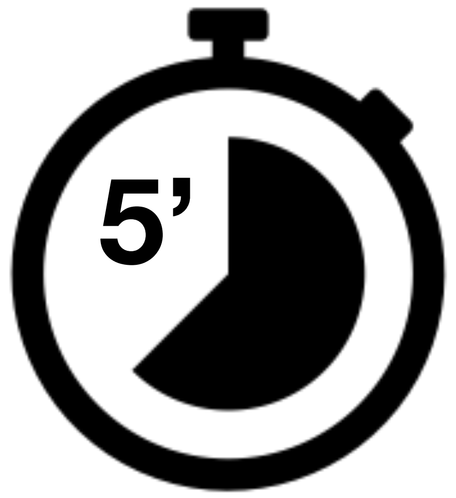
DFOH runs in a commodity server



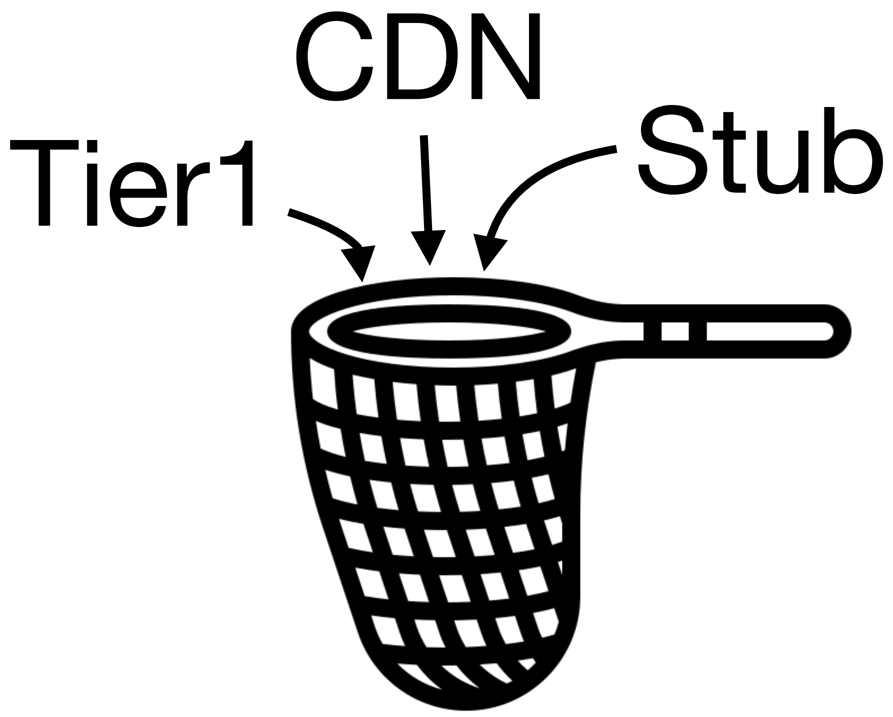
DFOH detects past hijacks



DFOH detects hijacks on the whole Internet



DFOH provides near-real-time detection



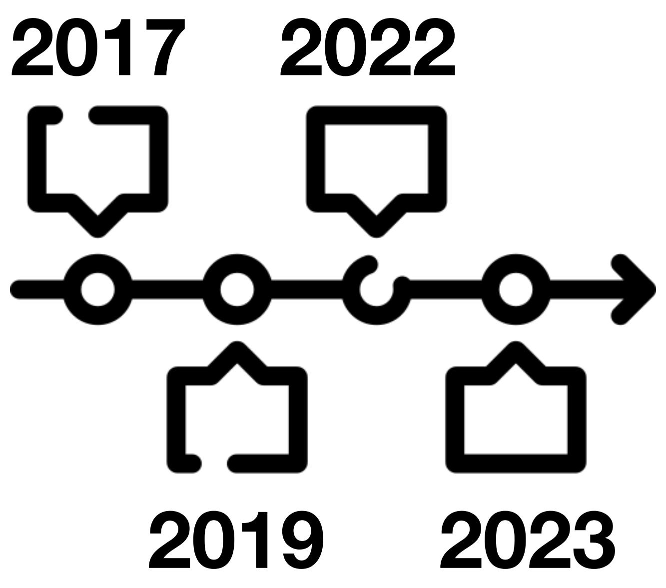
DFOH is accurate in every attack scenario

DFOH is robust against adversarial inputs

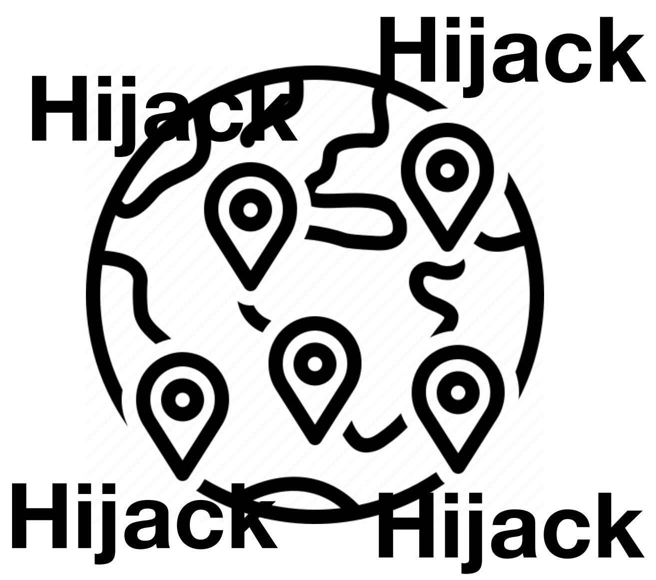
DFOH: A System to Detect Forged-Origin Hijacks

dfoh.info.ucl.ac.be

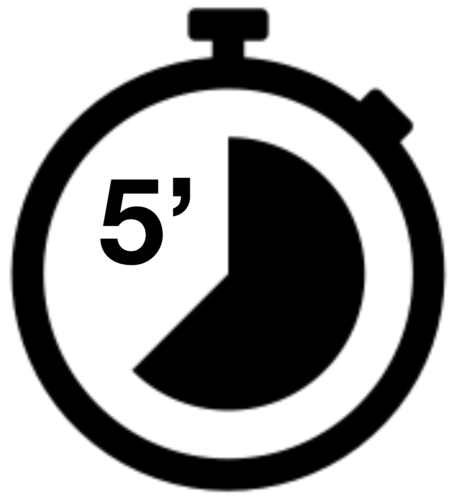
DFOH **runs in a commodity server**



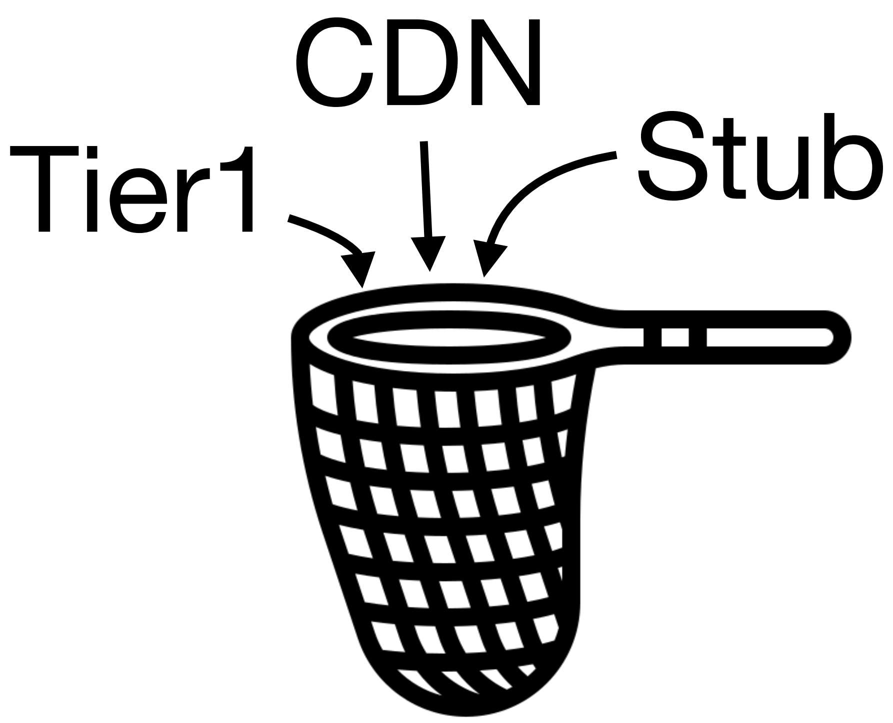
DFOH **detects past hijacks**



DFOH **detects hijacks on the whole Internet**



DFOH **provides near-real-time detection**

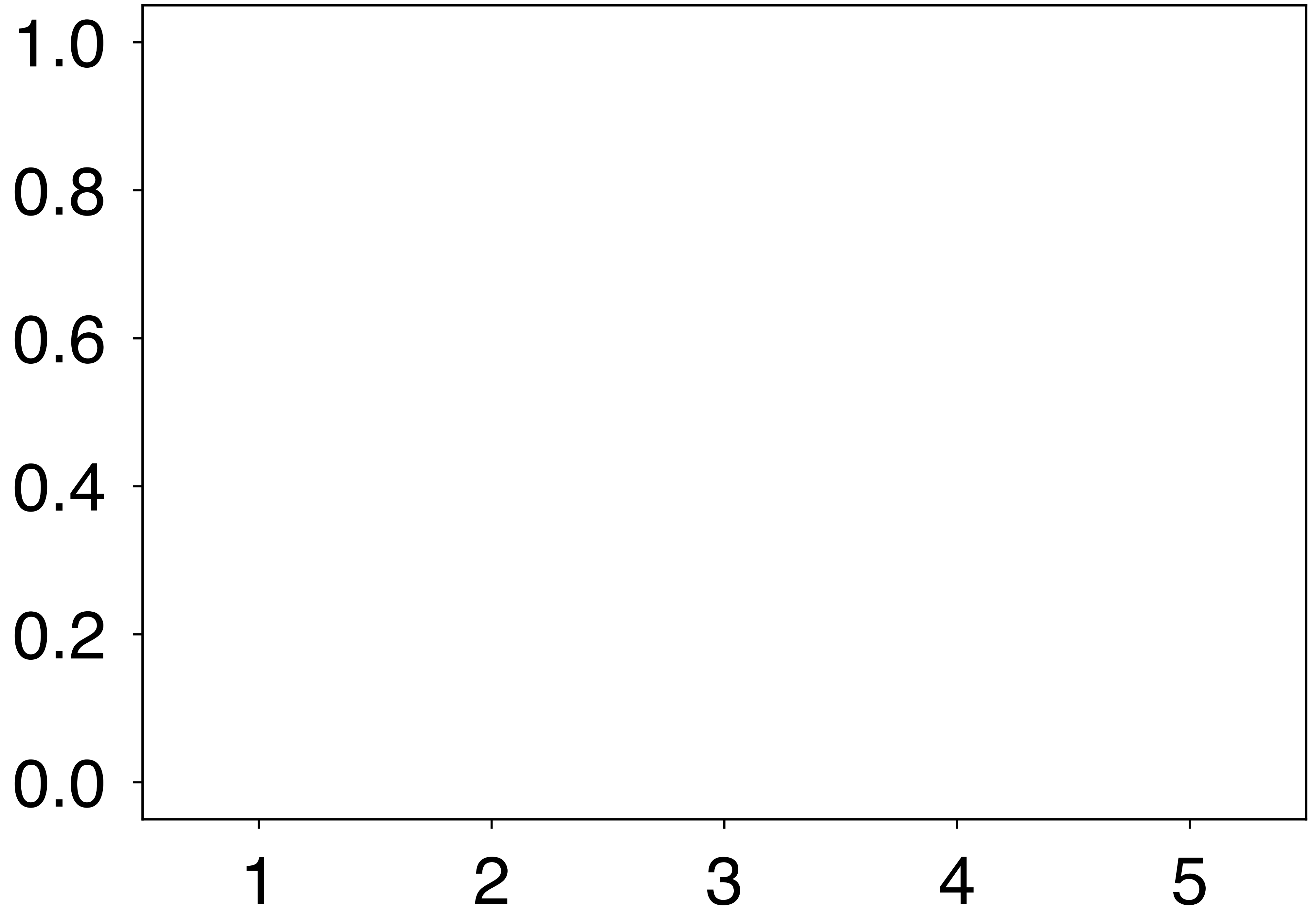


DFOH **is accurate in every attack scenario**

DFOH **is robust against adversarial inputs**

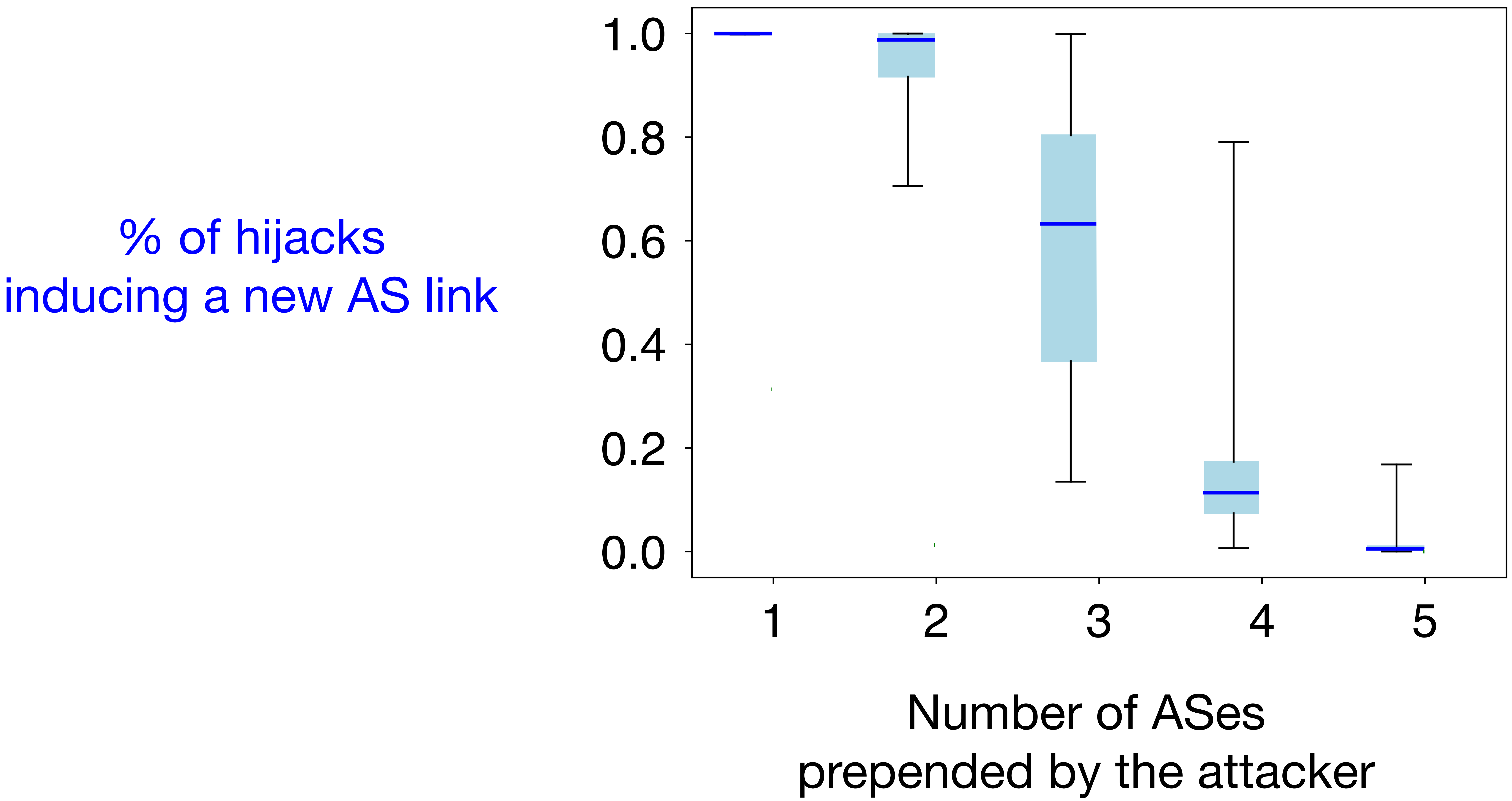
An attacker cannot escape from creating a new AS link without hampering the effectiveness of its attack

% of hijacks
inducing a new AS link

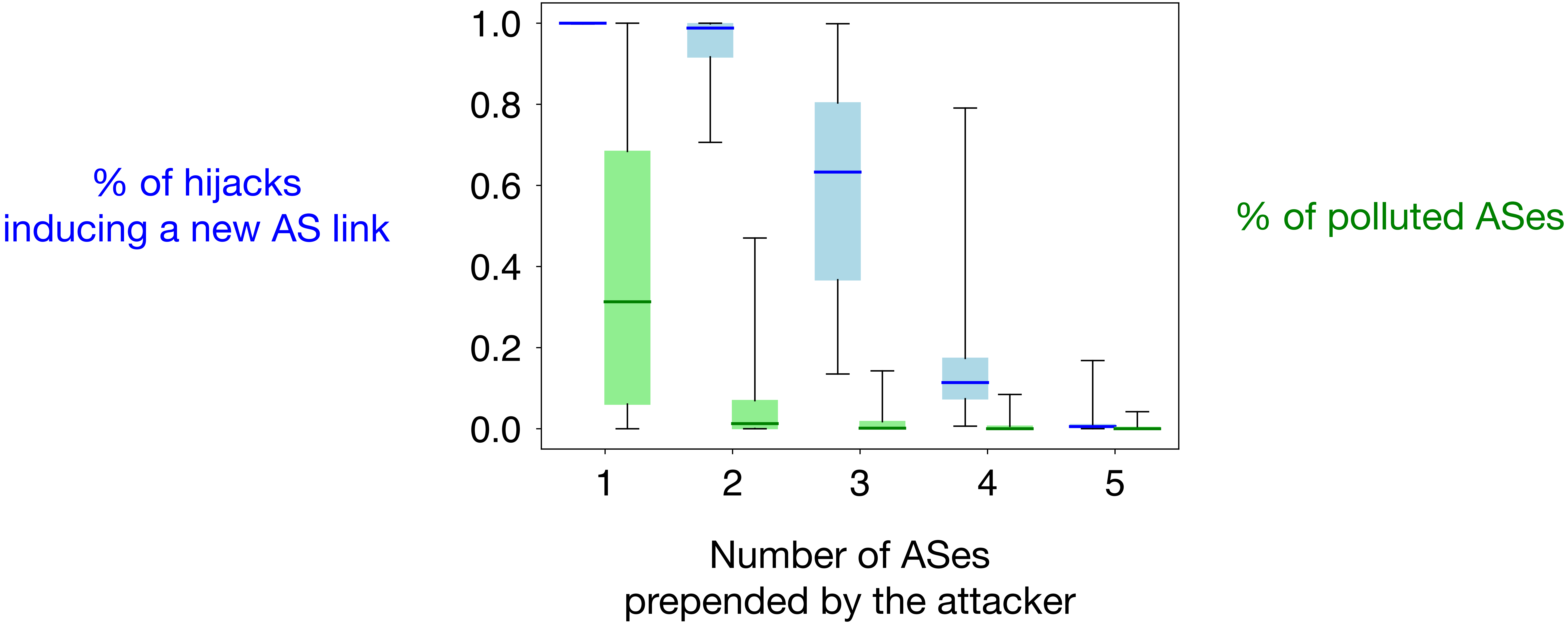


Number of ASes
prepended by the attacker

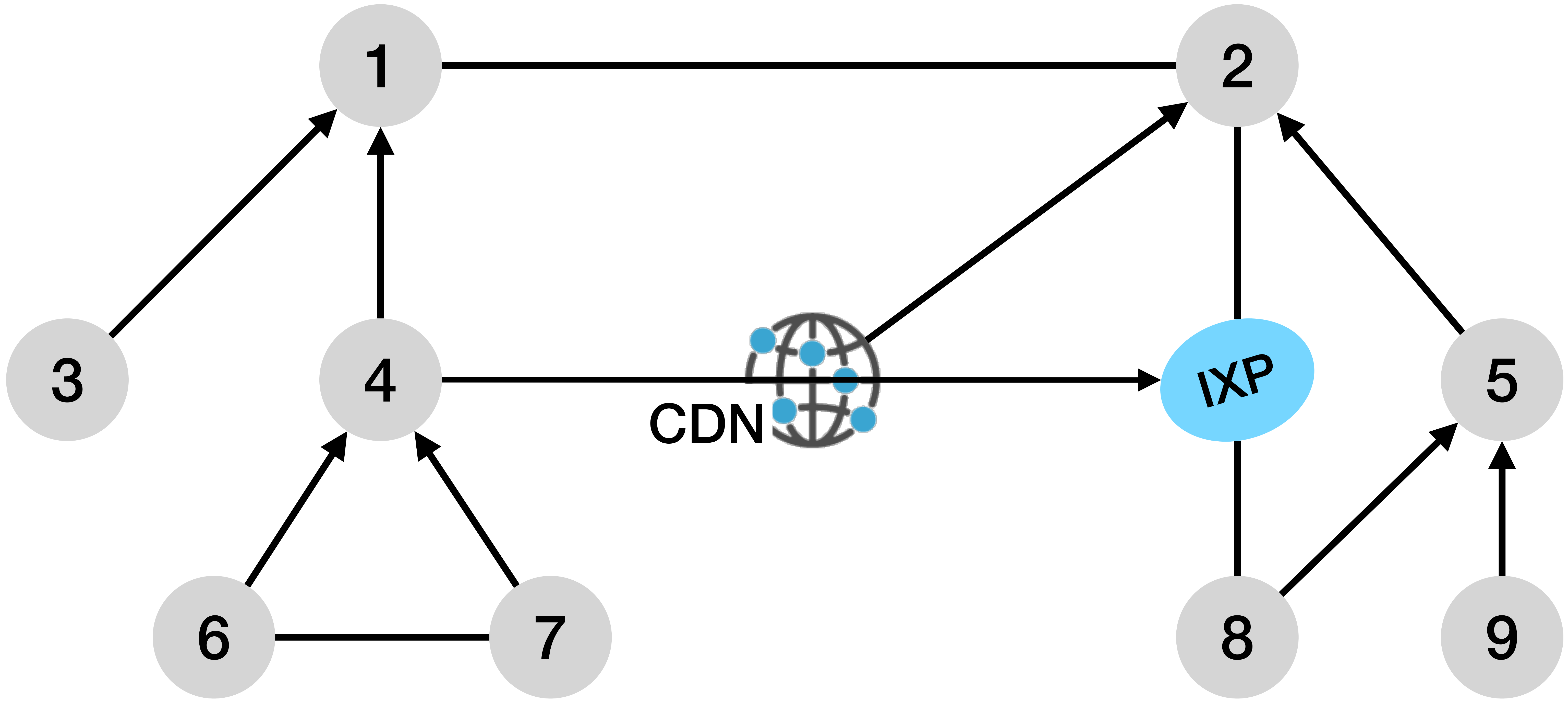
An attacker cannot escape from creating a new AS link without hampering the effectiveness of its attack



An attacker cannot escape from creating a new AS link without hampering the effectiveness of its attack

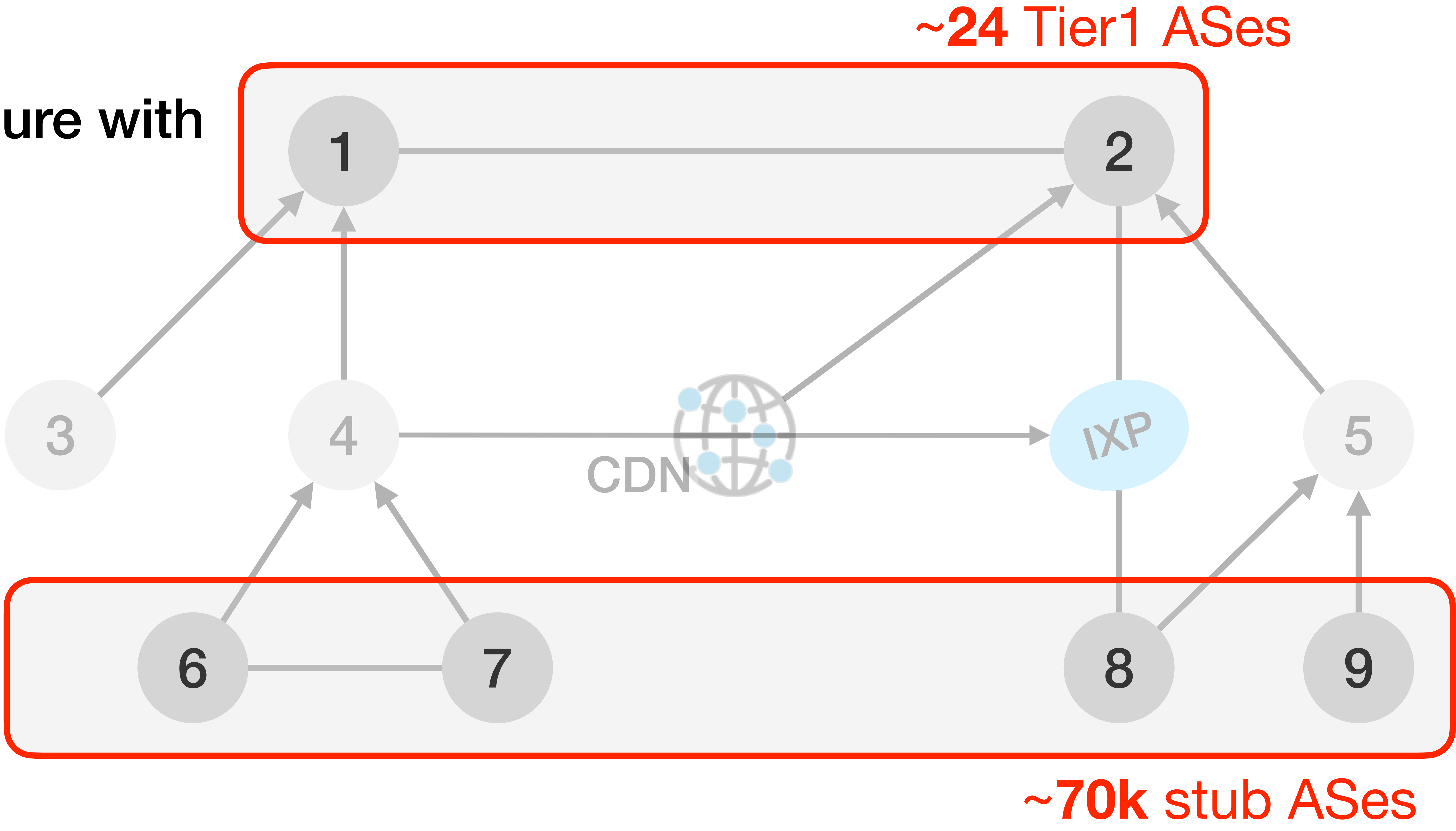


Generic algorithms for link prediction fails to reveal fake AS links because the AS topology exhibits particular patterns



Generic algorithms for link prediction fails to reveal fake AS links because the AS topology exhibits particular patterns

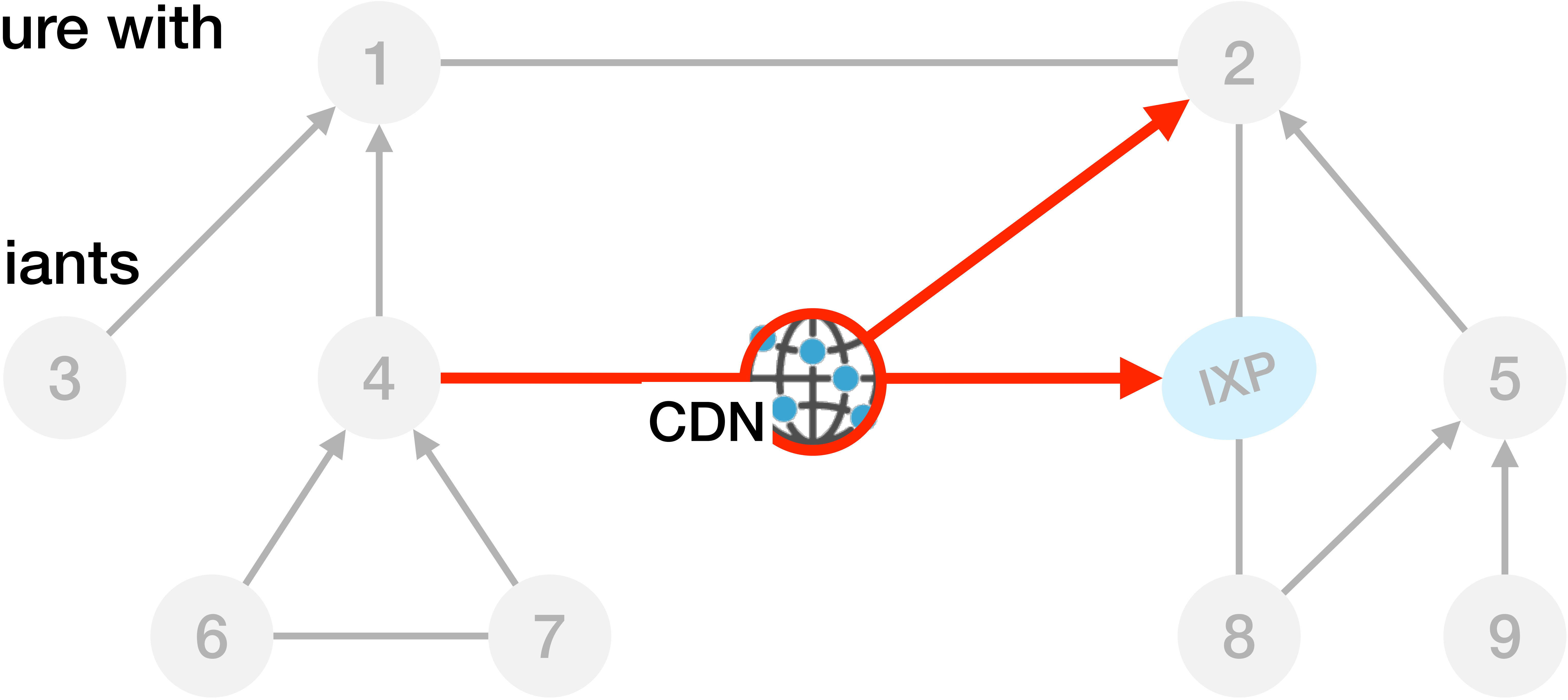
<u>Pattern #1:</u> Hierarchical structure with a few Tier1s and many stubs



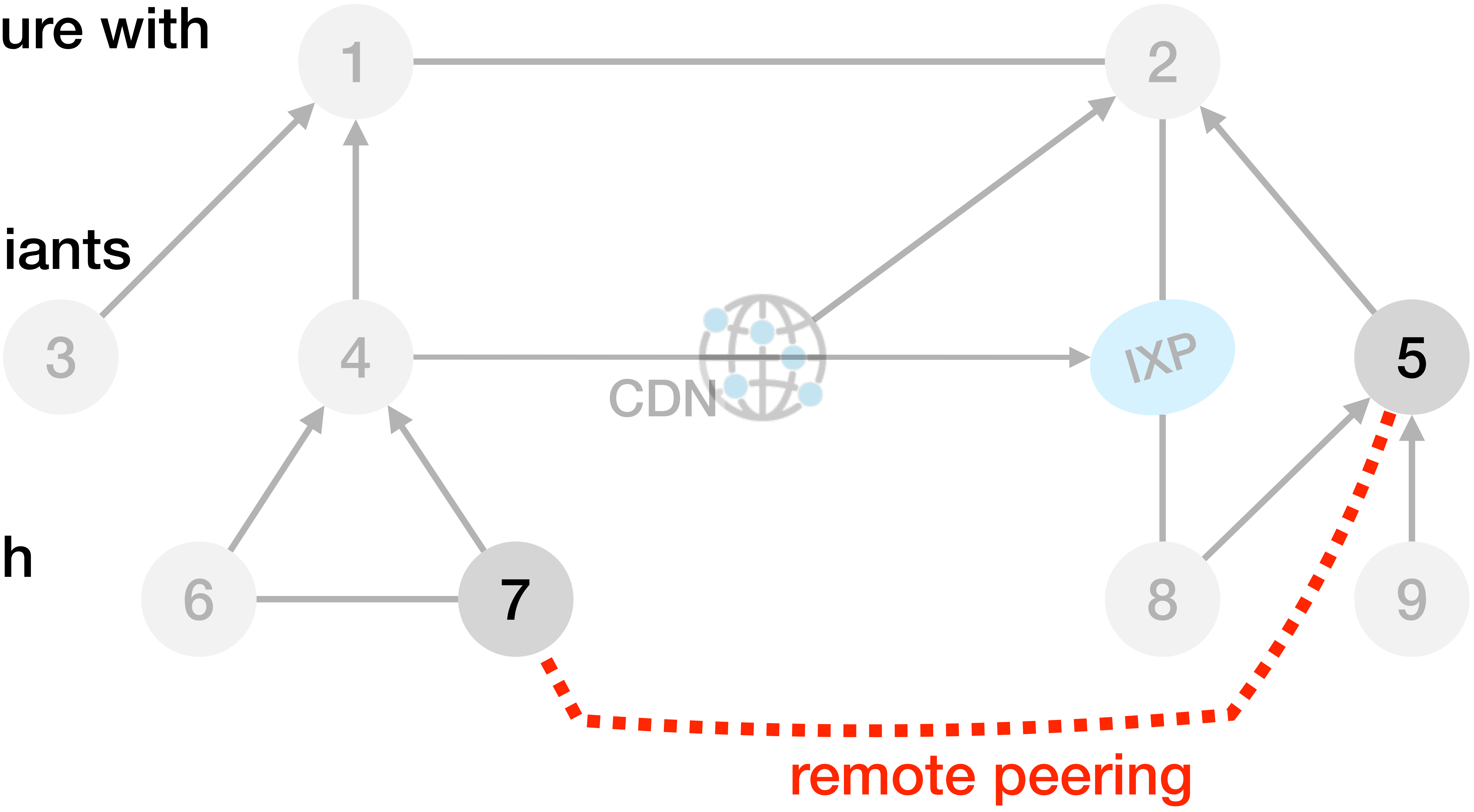
Generic algorithms for link prediction fails to reveal fake AS links because the AS topology exhibits particular patterns

Pattern #1: Hierarchical structure with a few Tier1s and many stubs

Pattern #2: CDNs and HyperGiants are highly connected



Generic algorithms for link prediction fails to reveal fake AS links because the AS topology exhibits particular patterns



<u>Pattern #1:</u> Hierarchical structure with a few Tier1s and many stubs

<u>Pattern #2:</u> CDNs and HyperGiants are highly connected

<u>Pattern #3:</u> Remote peerings and IP tunnels flatten the graph

<u>Step #1:</u> Finding new links

DFOH **takes all updates and one RIB per month**
from 200 BGP vantage points selected using MVP*

DFOH **builds the AS topology at day *d***
using AS paths in BGP routes collected during the 300 days prior *d*

DFOH **infers that an AS link observed at day *d* is new**
if the link is not in the AS topology constructed at day *d*

*mvp.info.ucl.ac.be

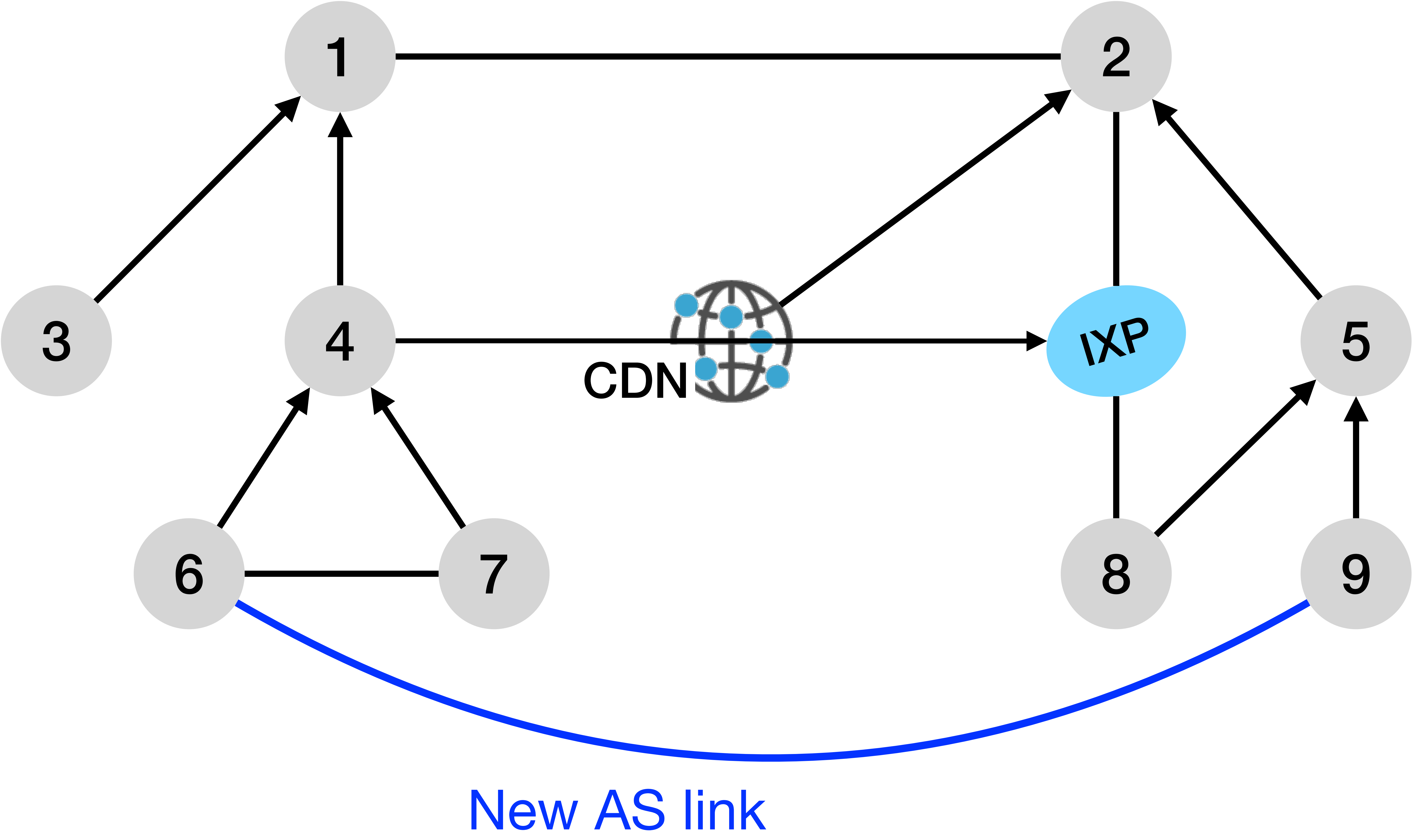
DFOH computes the change induced by the new AS link on topological features

Example with the **shortest distance** feature

Before the new link: shortest distance between 6 and 9 is **5**

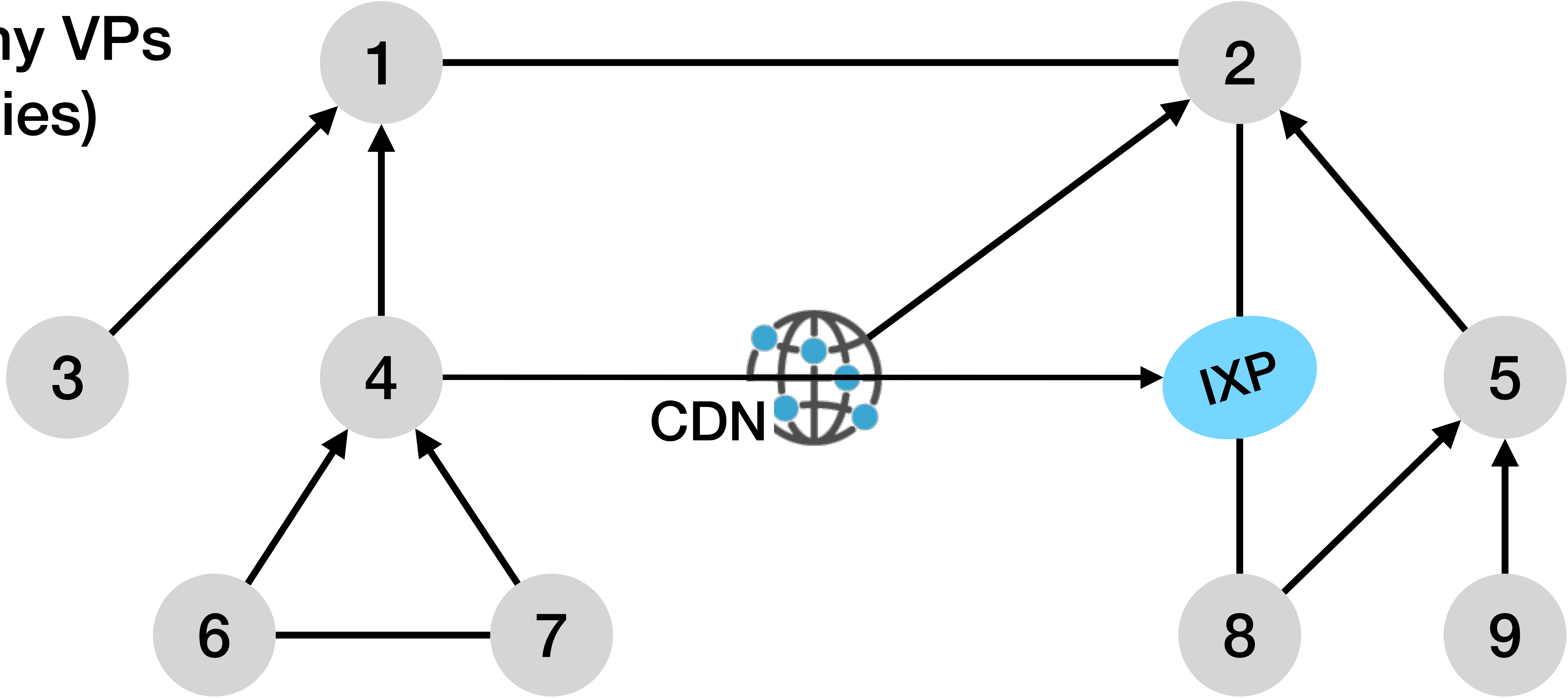
After the new link: shortest distance between 6 and 9 is **1**

Difference is 4



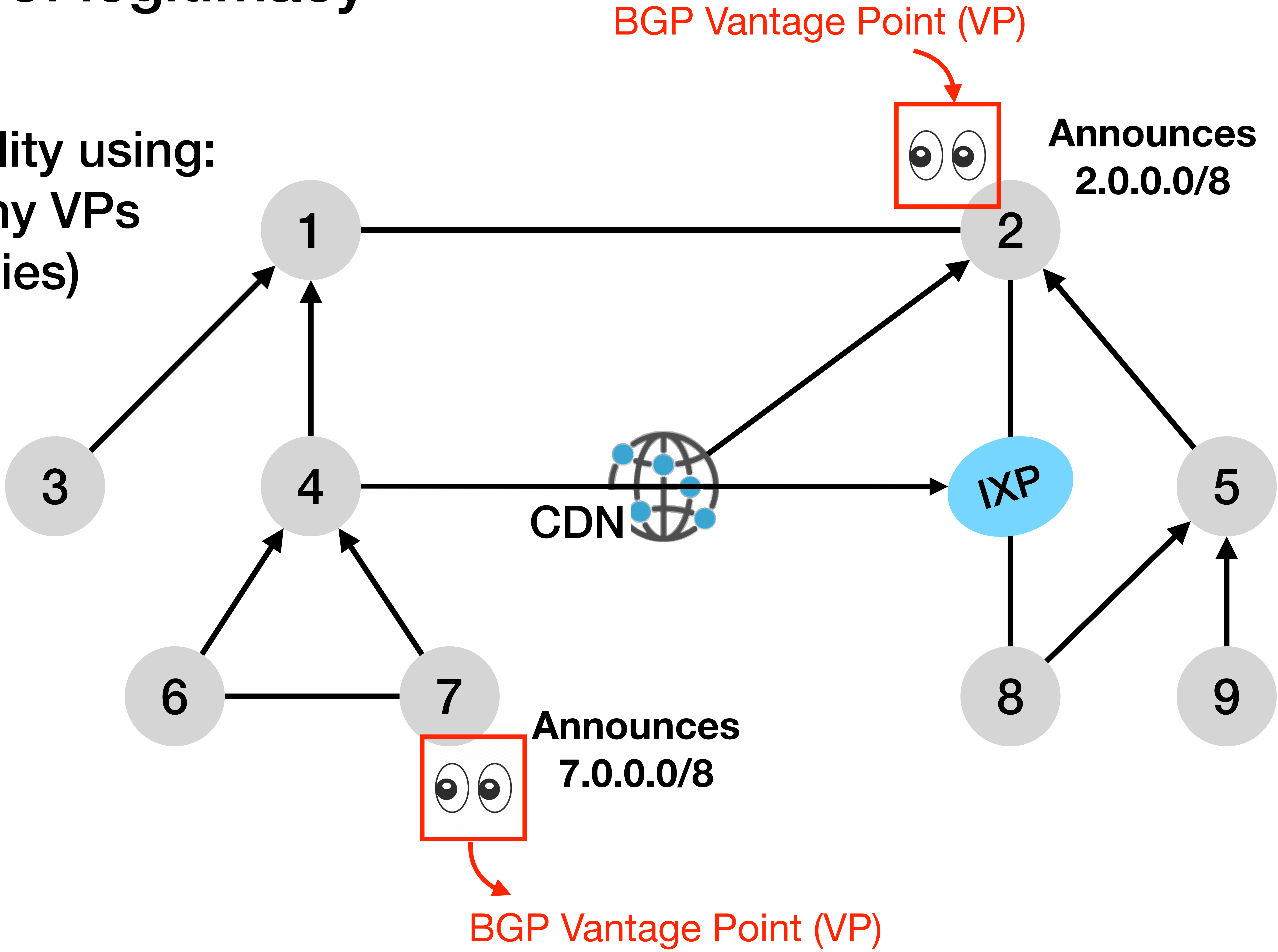
DFOH verifies that a new AS link is observed in both directions as it is a strong indicator of legitimacy

DFOH verifies link bidirectionality using:
- BGP data (AS path) from many VPs
- IRR data (import/export policies)



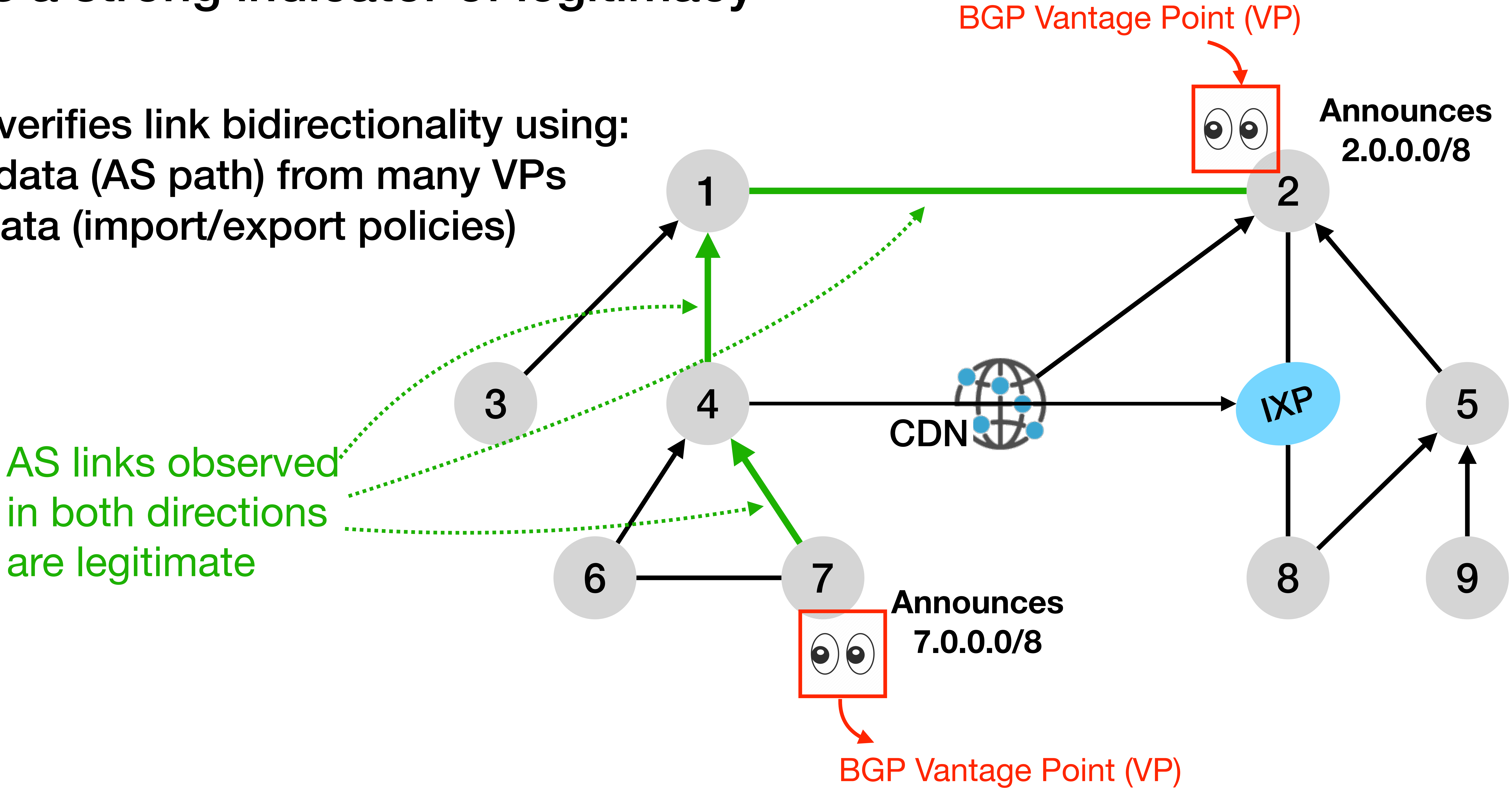
DFOH verifies that a new AS link is observed in both directions as it is a strong indicator of legitimacy

DFOH verifies link bidirectionality using:
- BGP data (AS path) from many VPs
- IRR data (import/export policies)



DFOH verifies that a new AS link is observed in both directions as it is a strong indicator of legitimacy

DFOH verifies link bidirectionality using:
- BGP data (AS path) from many VPs
- IRR data (import/export policies)



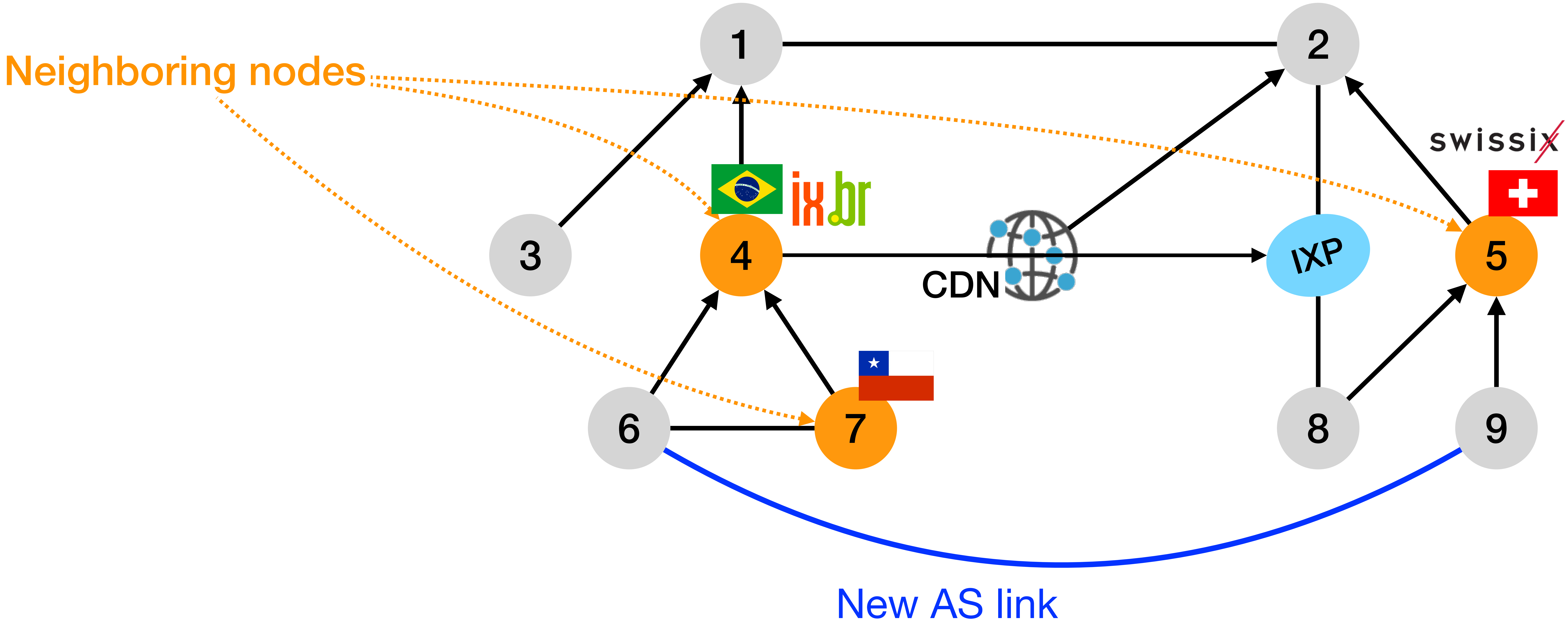
The bidirectionality feature is safe as an attacker cannot intentionally fake both directions of an AS link

Faking both directions in the same AS path
would create a loop

Faking both directions in the IRR is not possible
as the attacker only controls its IRR data

Merging these two datasets is safe
as an attacker can only fake the same direction in BGP and the IRR

DFOH considers the neighbouring nodes to avoid adversarial inputs as the information on Peeringdb is not verified



DFOH learns the pattern of legitimate and malicious AS paths using a supervised training model

DFOH samples X legitimate AS paths
and artificially creates the same number of maliciously-induced AS paths

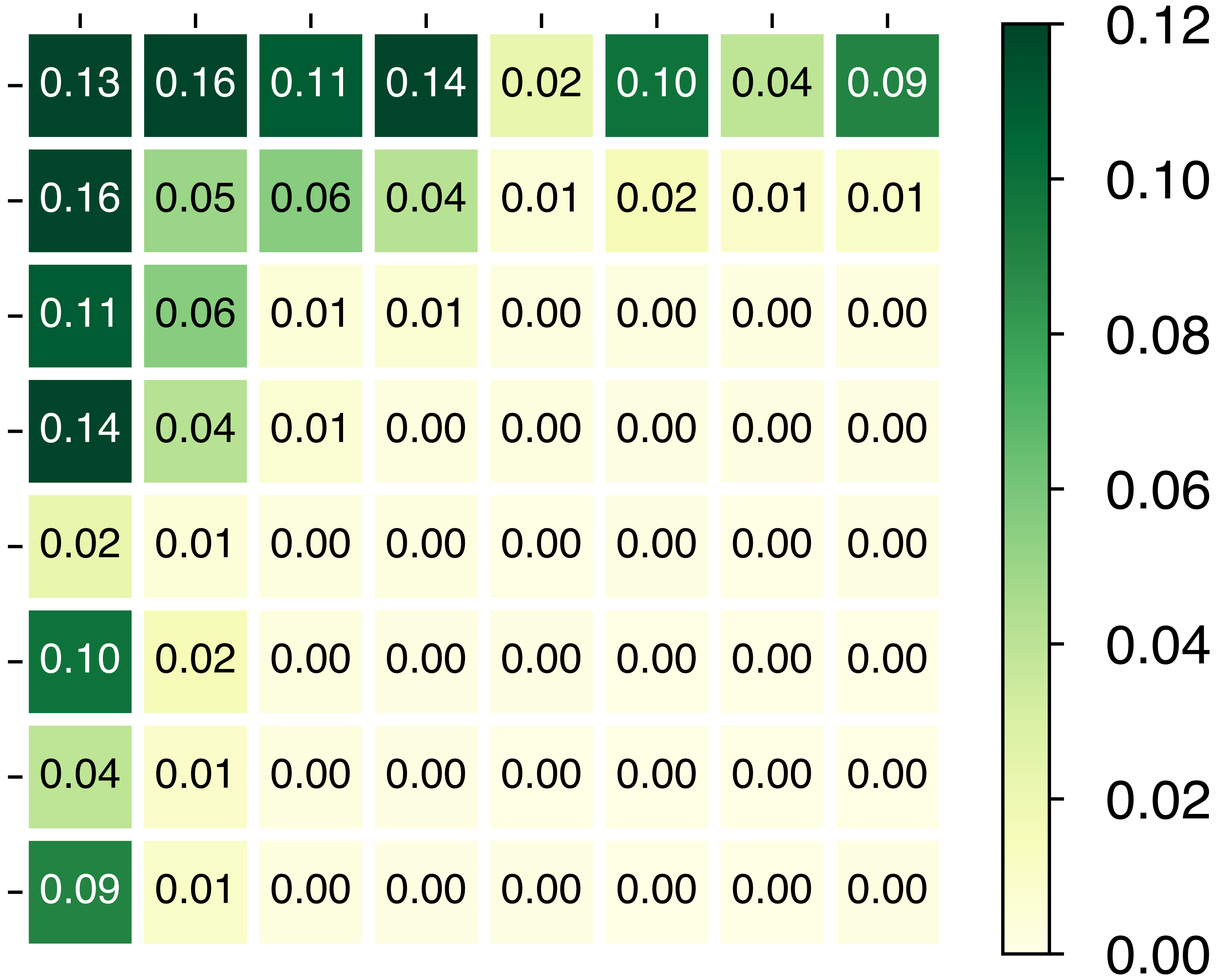
DFOH computes the degree and customer cone size of
every AS in the sampled AS paths

DFOH trains a random forest that it uses
to compute a probability that a given AS path is fake or real

DFOH builds a sample of nonexistent links that is similarly balanced as the set of existing links

Existing links distribution
between different AS categories

AS category: →
↓



0.13	0.16	0.11	0.14	0.02	0.10	0.04	0.09
0.16	0.05	0.06	0.04	0.01	0.02	0.01	0.01
0.11	0.06	0.01	0.01	0.00	0.00	0.00	0.00
0.14	0.04	0.01	0.00	0.00	0.00	0.00	0.00
0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00
0.10	0.02	0.00	0.00	0.00	0.00	0.00	0.00
0.04	0.01	0.00	0.00	0.00	0.00	0.00	0.00
0.09	0.01	0.00	0.00	0.00	0.00	0.00	0.00

Sampled nonexistent links distribution
when using ***DFOH***'s balanced sampling

0.13	0.14	0.10	0.11	0.03	0.09	0.05	0.09
0.14	0.05	0.05	0.04	0.01	0.02	0.02	0.02
0.10	0.05	0.01	0.01	0.01	0.00	0.00	0.00
0.11	0.04	0.01	0.01	0.00	0.00	0.00	0.00
0.03	0.01	0.01	0.00	0.00	0.00	0.00	0.00
0.09	0.02	0.00	0.00	0.00	0.00	0.00	0.00
0.05	0.02	0.00	0.00	0.00	0.00	0.00	0.00
0.09	0.02	0.00	0.00	0.00	0.00	0.00	0.00

DFOH uses a random forest classifier to classify an AS link as fake or legitimate

1. ***DFOH*** samples 30k existing and nonexistent AS links

2. ***DFOH*** estimates the best parameters using a cross-validated grid search on 25% of the sampled AS links

3. ***DFOH*** trains the classifier with the remaining 75% of the AS links

DFOH repeats this process every day
to ensure that its inferences remain accurate over time