https://faelix.link/linxman2021

STEERING EYEBALLS CLOSER TO CONTENT

WHOLESALE NETWORKS TRICKS FOR LOCAL TRAFFIC DELIVERY

About Marek

- Stuff I do:
  - CTO @FAELIX – https://faelix.net/
  - PC @uknof – https://uknof.uk/
  - Crew @net_mcr – https://www.netmcr.uk/
- Me — @maznu – @NetworkMoose

About Faelix

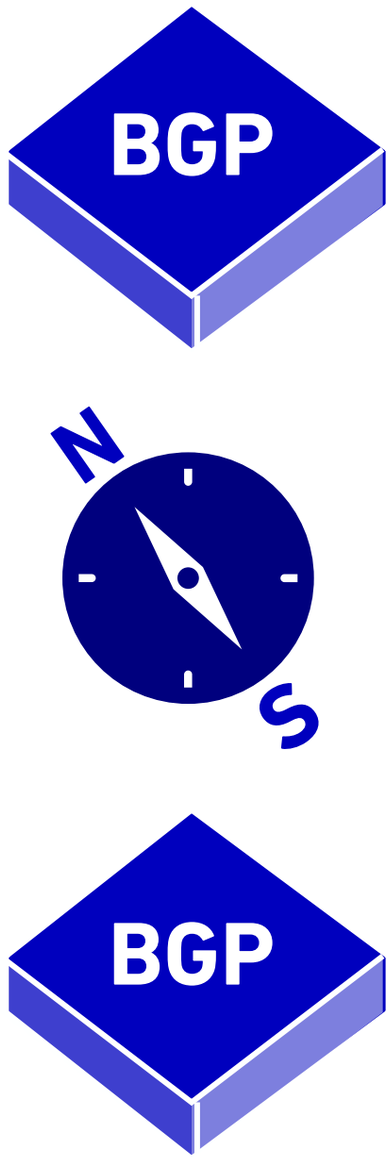
- LINX Silver Partner (LON1, LON2, Manchester)
- Supporting alt-nets (e.g. UKWISPA Consultants)
  - Colocation and interconnection
  - Transit and peering
  - VNF hosting (NFV VPSs)
  - Wholesale connectivity
- Look out for the "AS41495 Faelix Limited" SSID

This Talk

- Aimed at Internet access providers, or networks with a predominantly inbound traffic ratio.
- Consuming wholesale “national ethernet” or L2TP broadband services, or building their own last mile networks using fixed wireless or fibre technologies.
- With multiple “core” sites where their ASN takes transit and is peering on national and regional IXs.

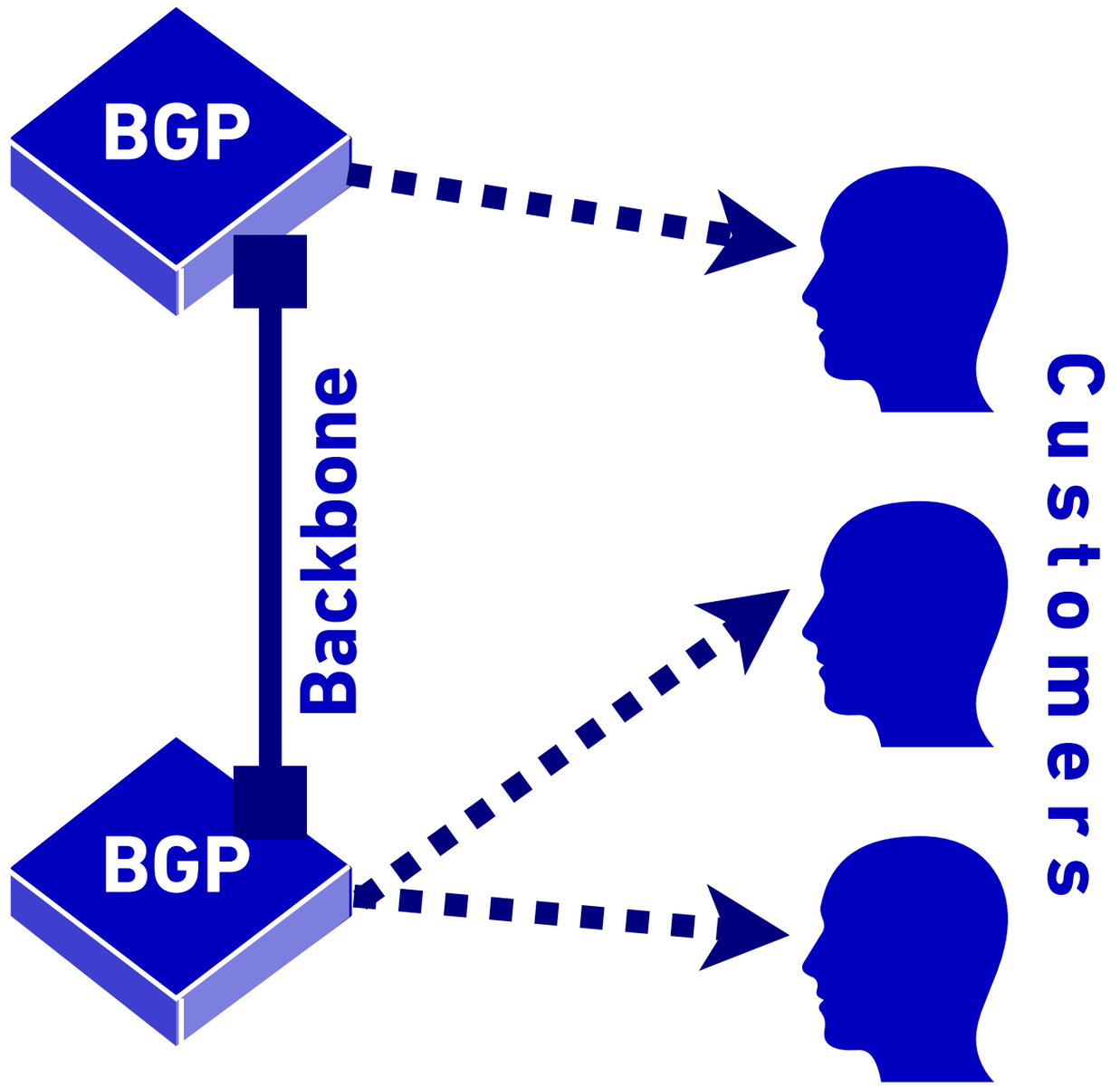
EXAMPLE ACCESS ISP

Overview

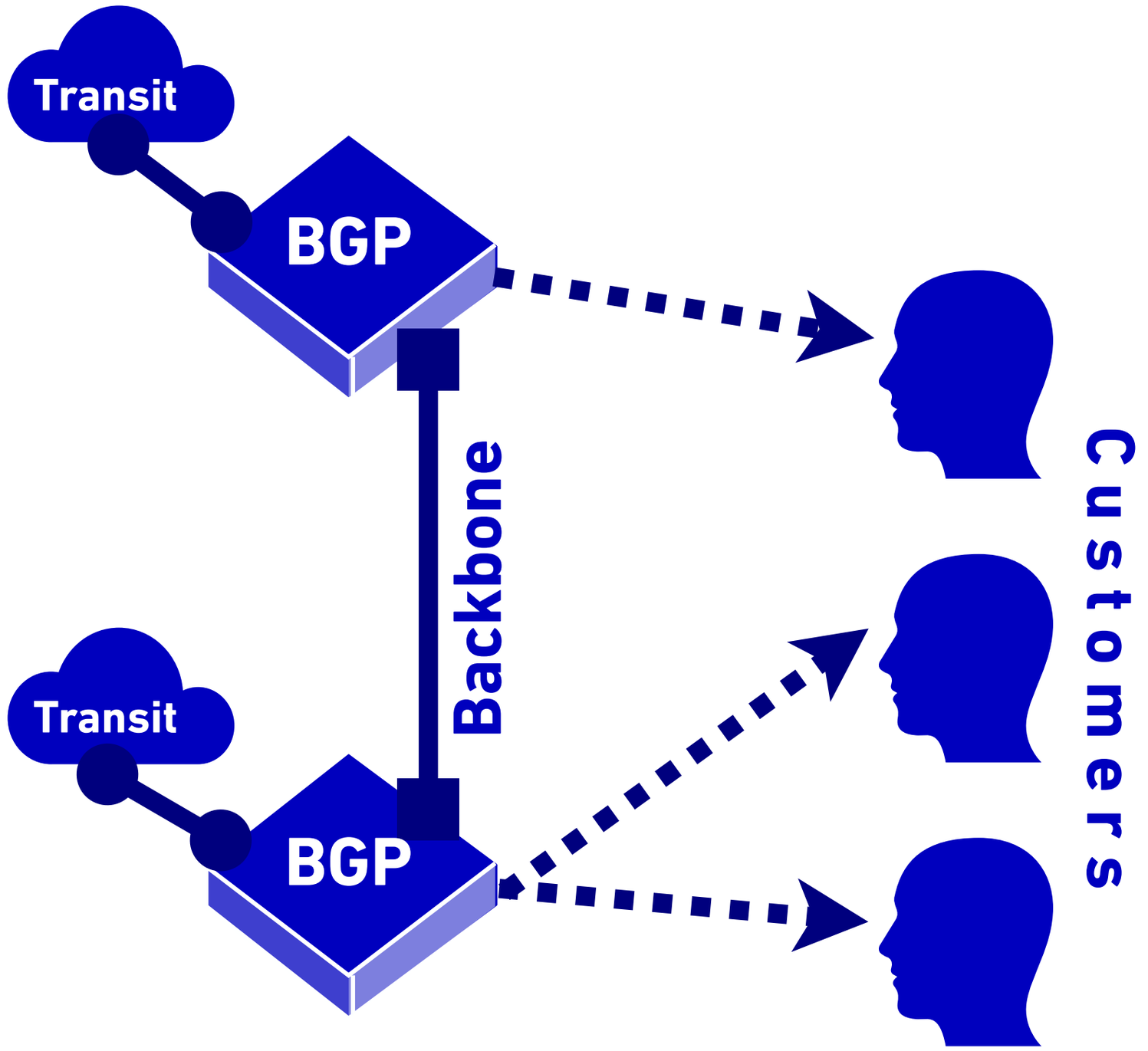


Overview

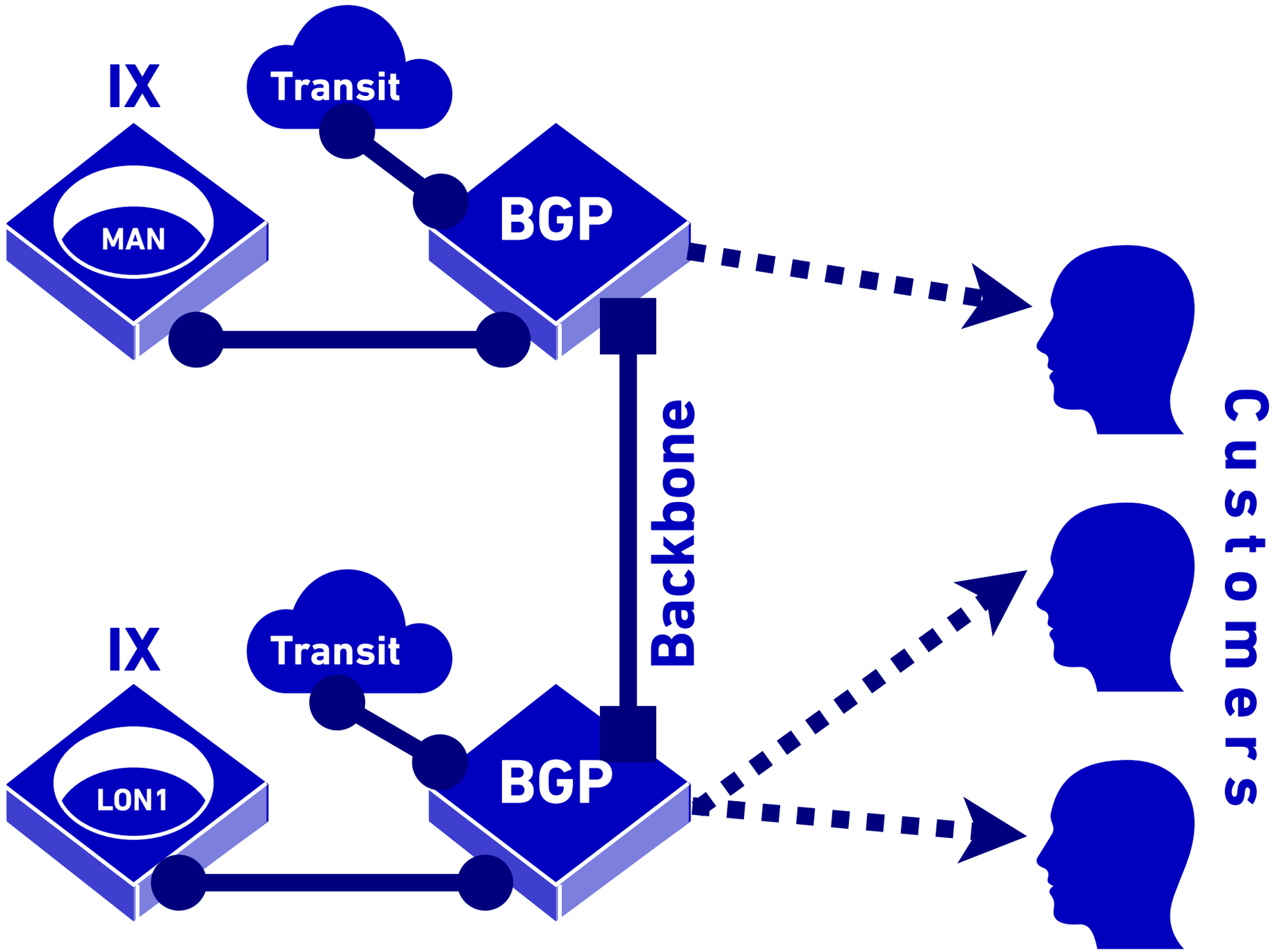
Overview



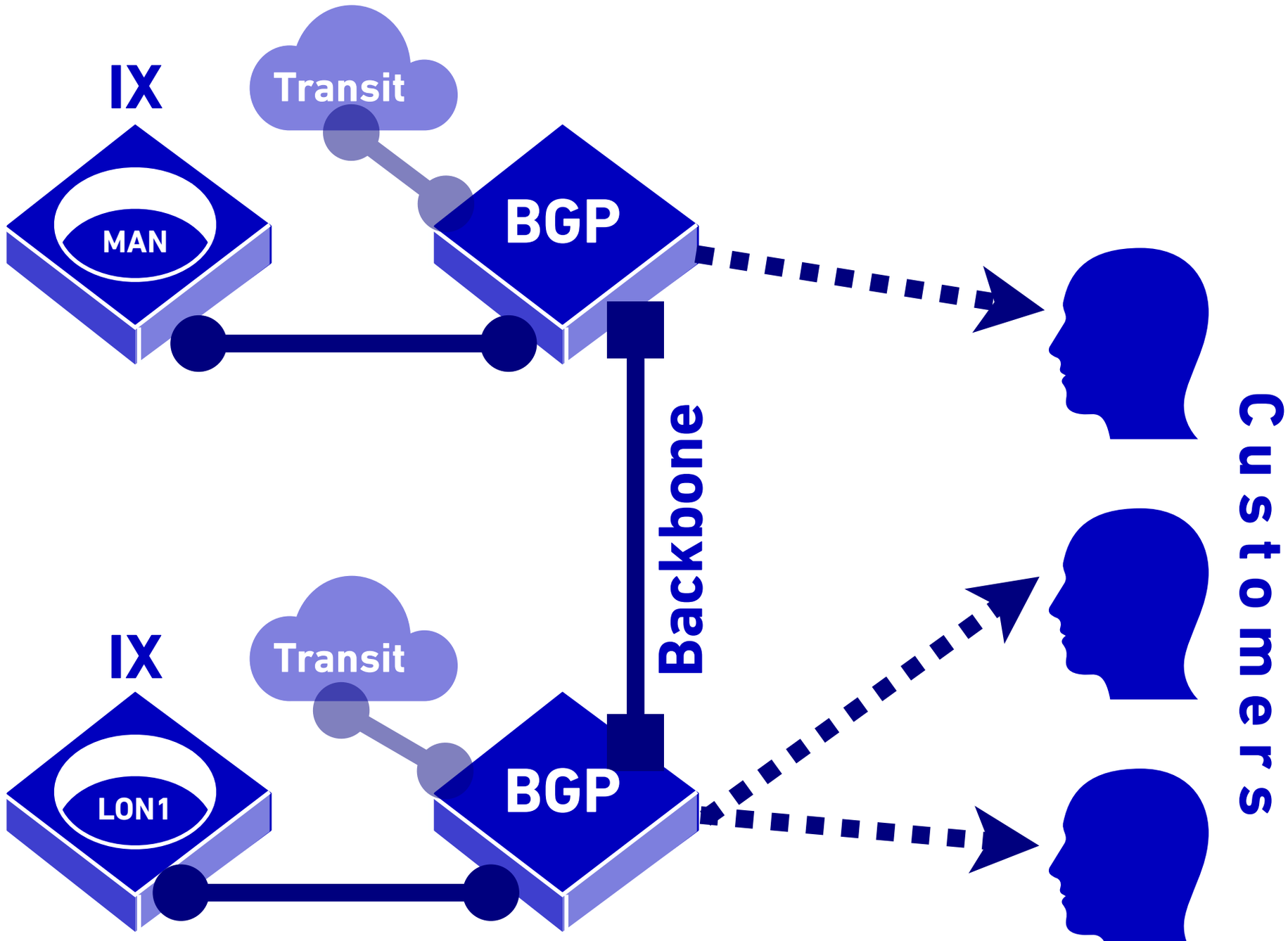
Overview



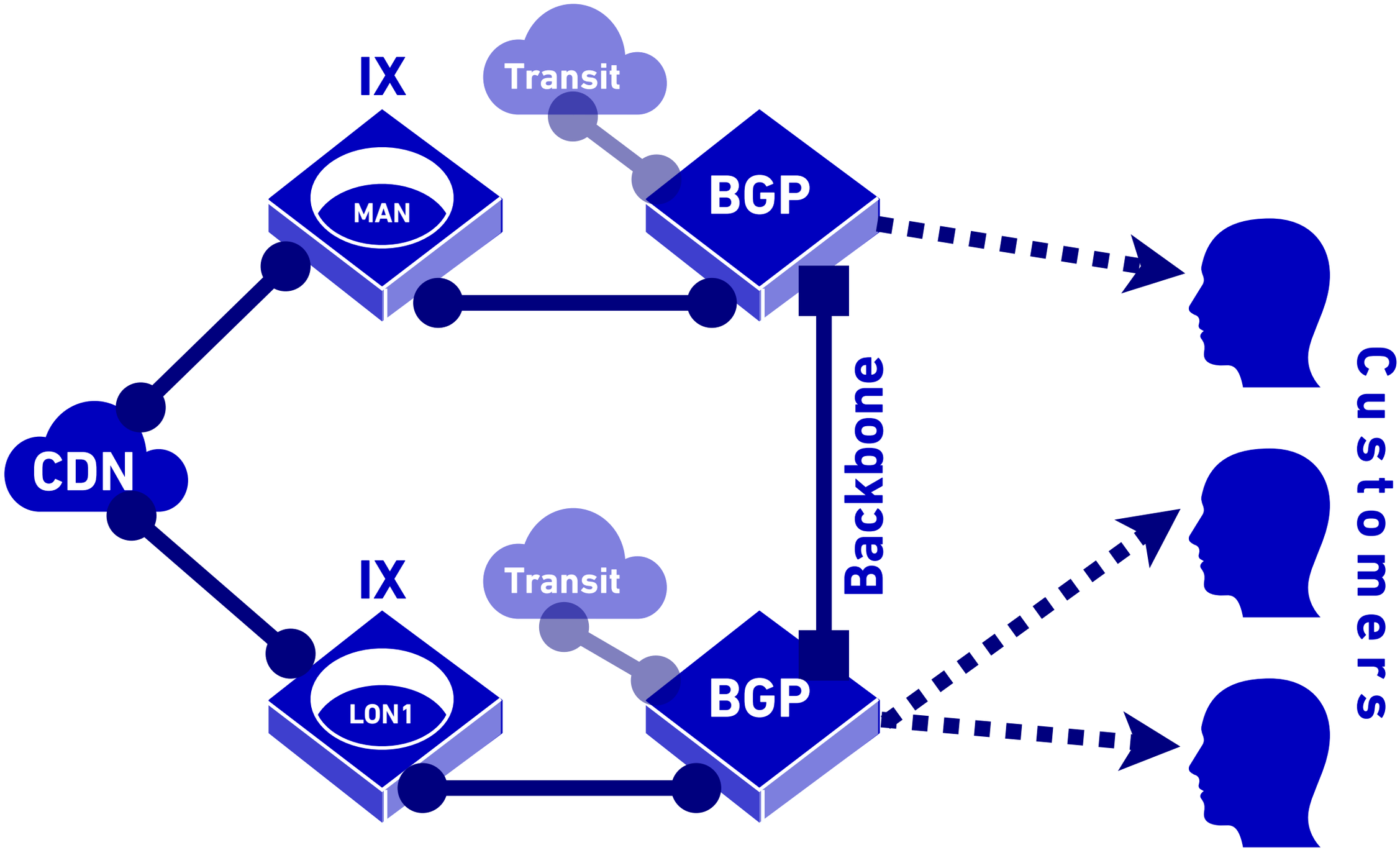
Overview



Overview

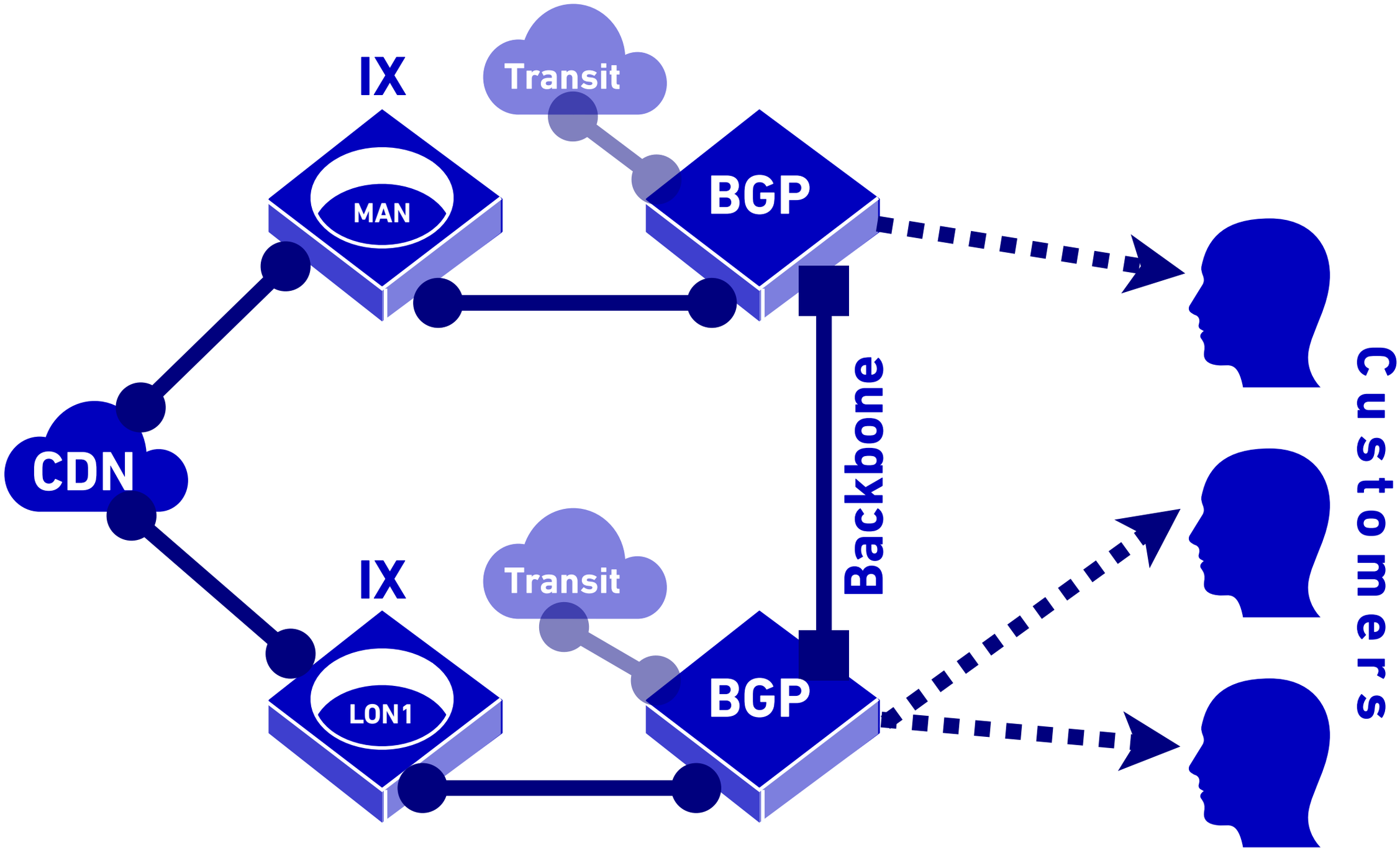


Overview

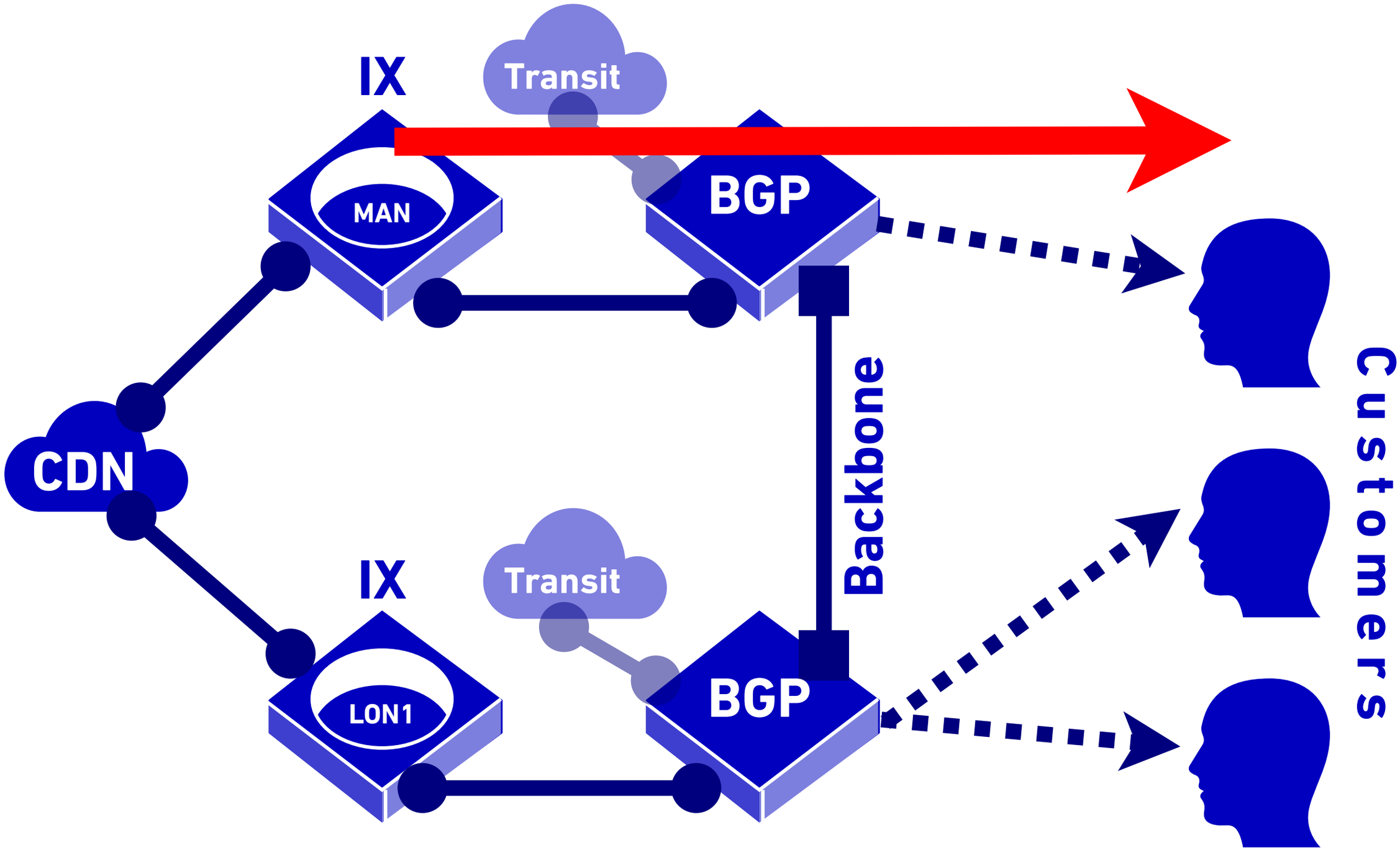

IX
Transit
MAN
BGP
CDN
Backbone
Customers
IX
Transit
LON1
BGP

TYPICAL TRAFFIC FLOWS

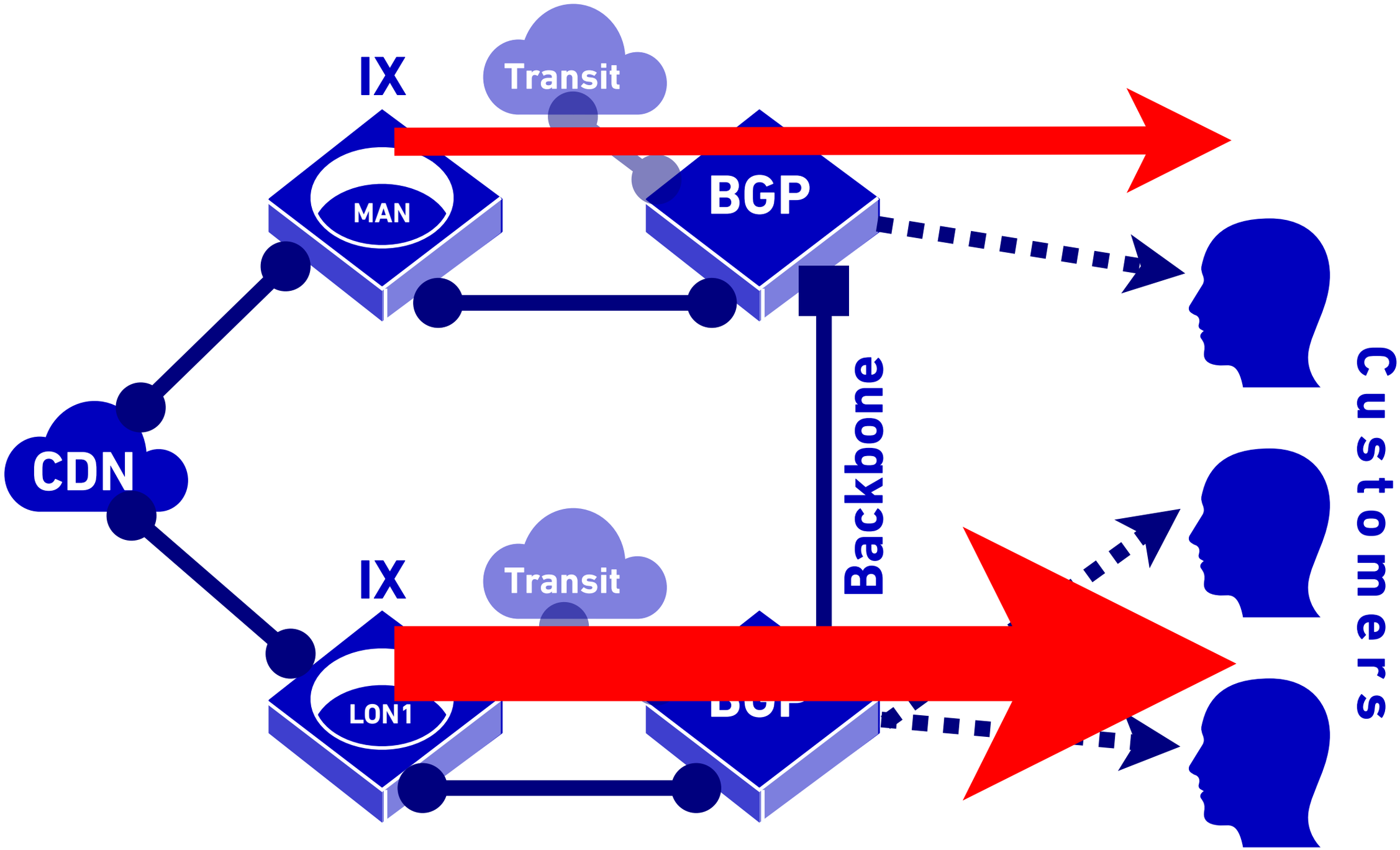
Overview


IX
Transit
MAN
BGP
CDN
Backbone
Customers
IX
Transit
LON1
BGP

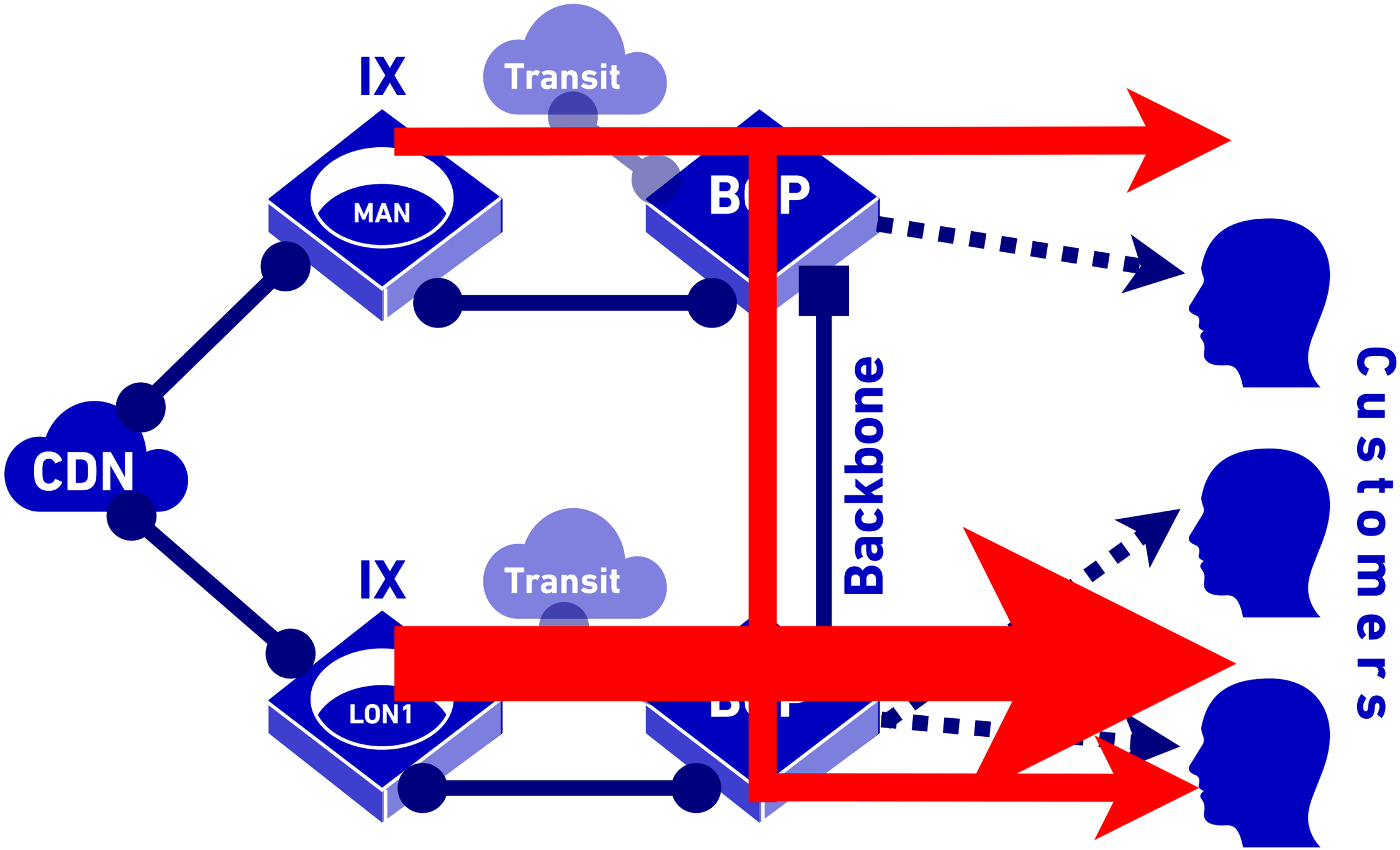
Overview


IX
Transit
MAN
BGP
CDN
Backbone
Customers
IX
Transit
LON1
BGP

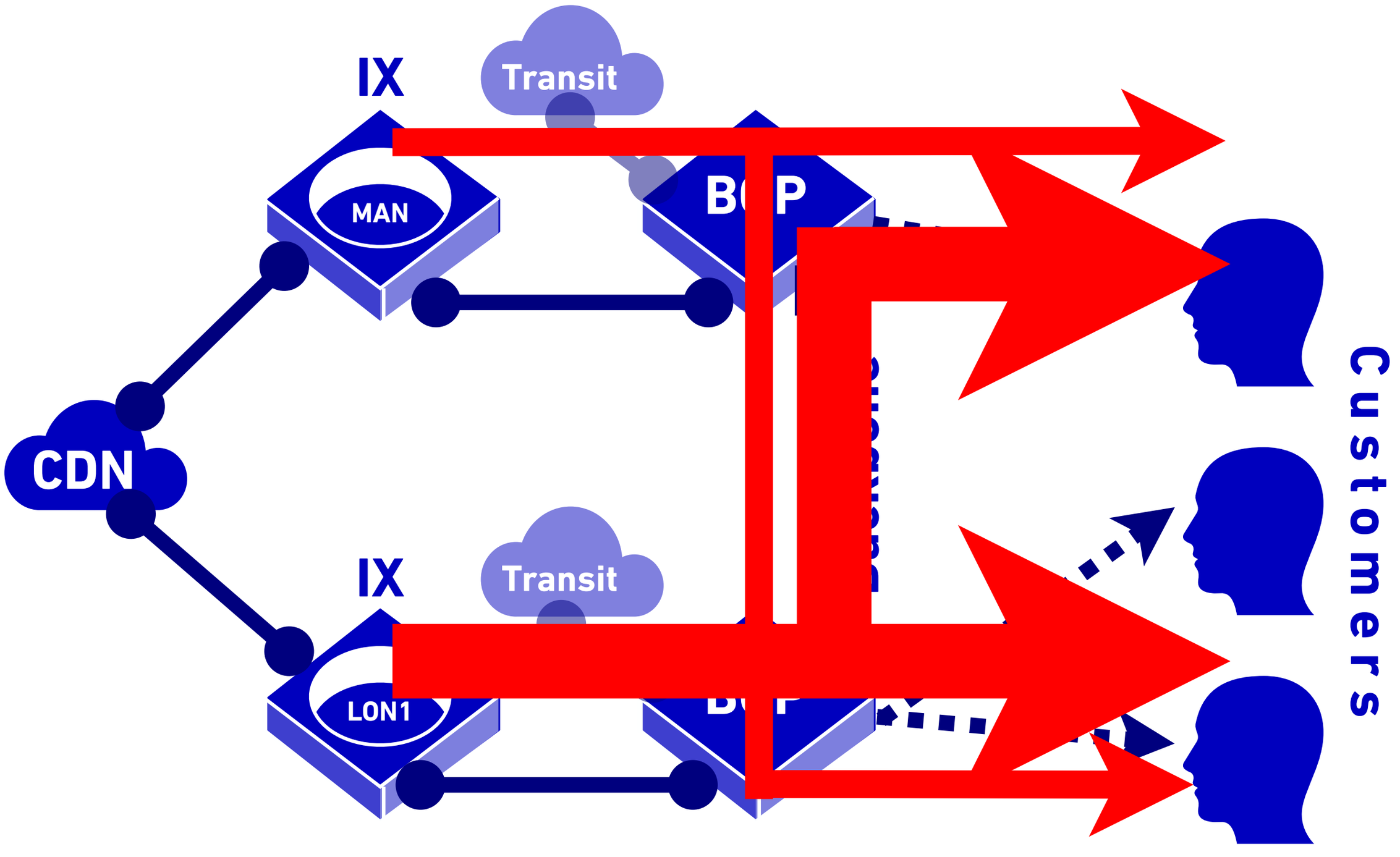
Overview


IX
Transit
MAN
BGP
CDN
Backbone
Customers
IX
Transit
LON1
BGP

Overview



Overview



Challenges

- If predominant traffic were outbound, could TE.
  - How do we help the CDNs deliver traffic locally?
- “National” IX traffic dominates “Regional” IX.
  - How to balance traffic from CDNs on both IXs?
- Backbone capacity roughly equal to peering capacity.
  - Why did we even bother building outside $capital?
  - How do we retrofit $capital into our local alt-net?

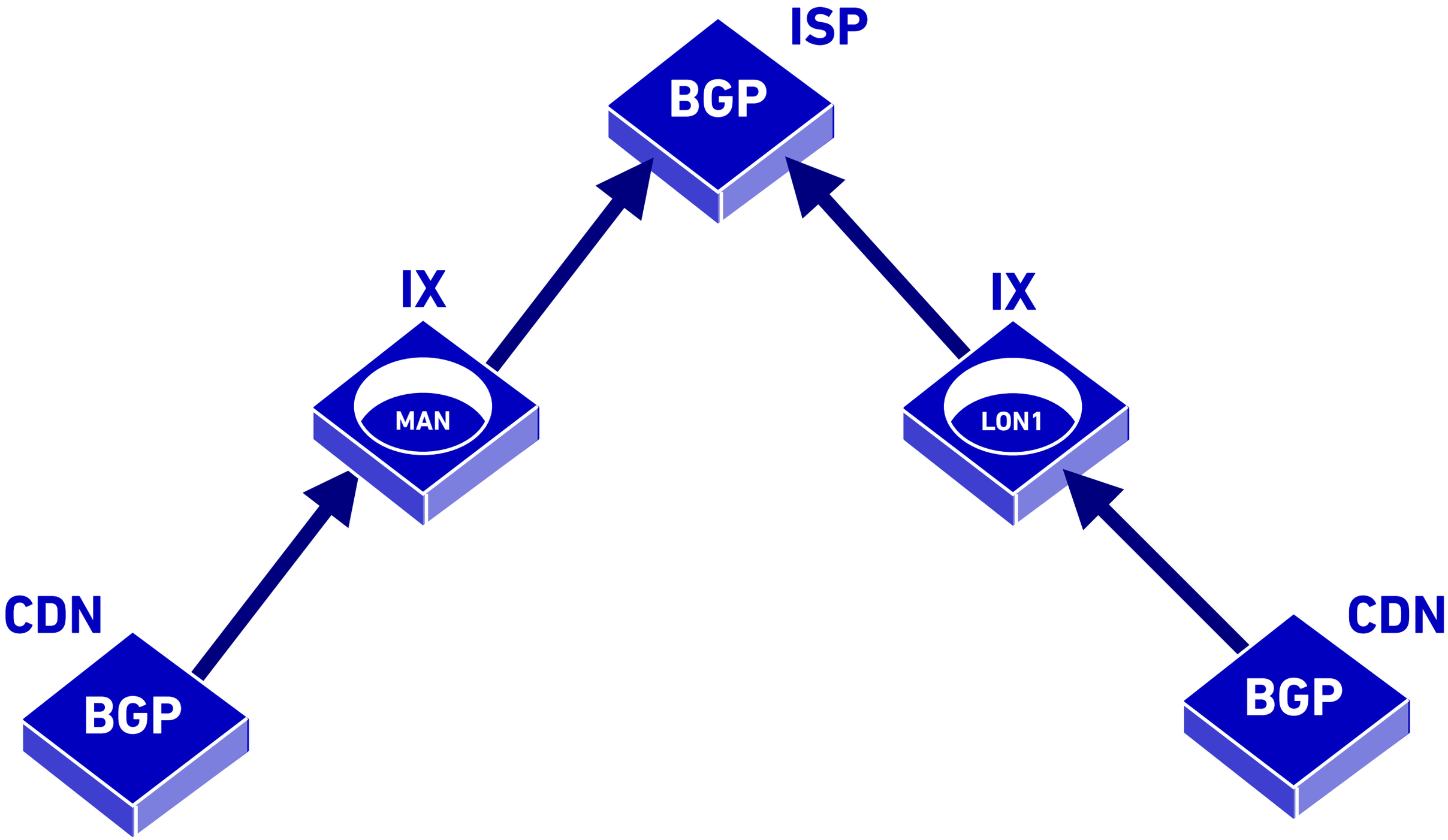
CONTENT NETWORKS

CDN Assumptions

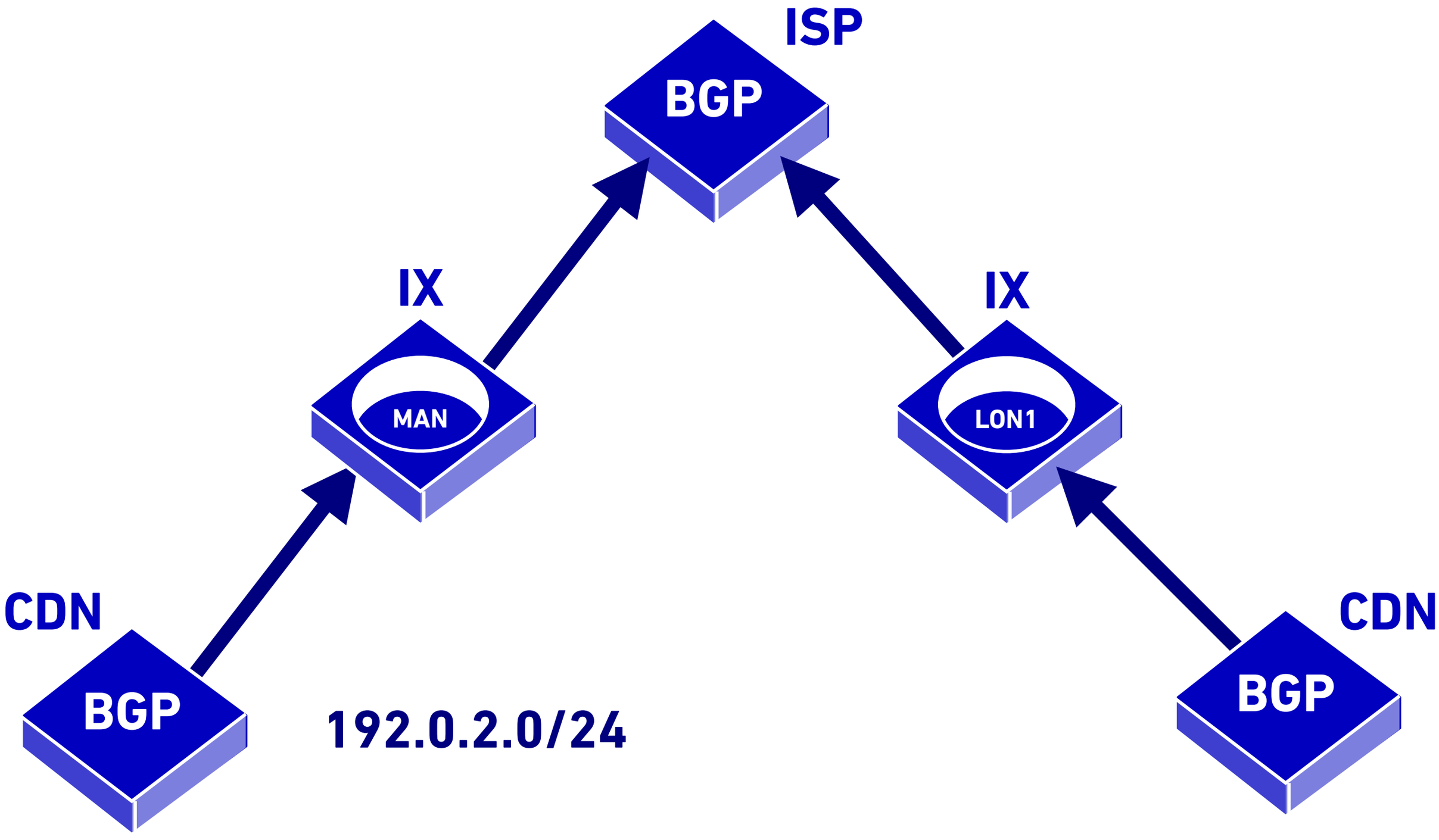
- Eyeballs make a DNS query to CDN.
- CDN responds with something that is nearby.
- How they do this is their secret sauce.

- Mix of anycast, GeoIP, and control plane magic.

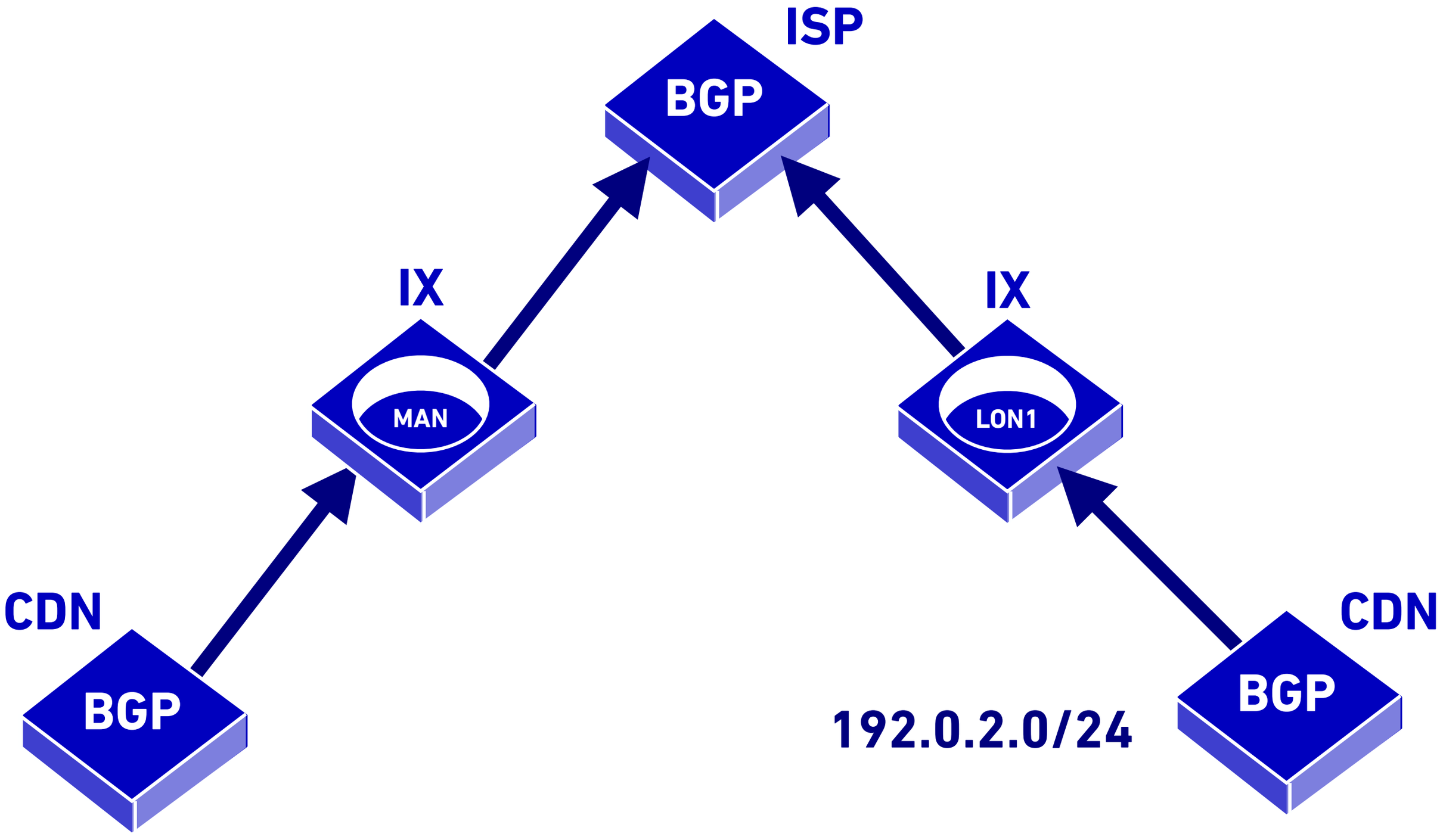
CDN Assumptions



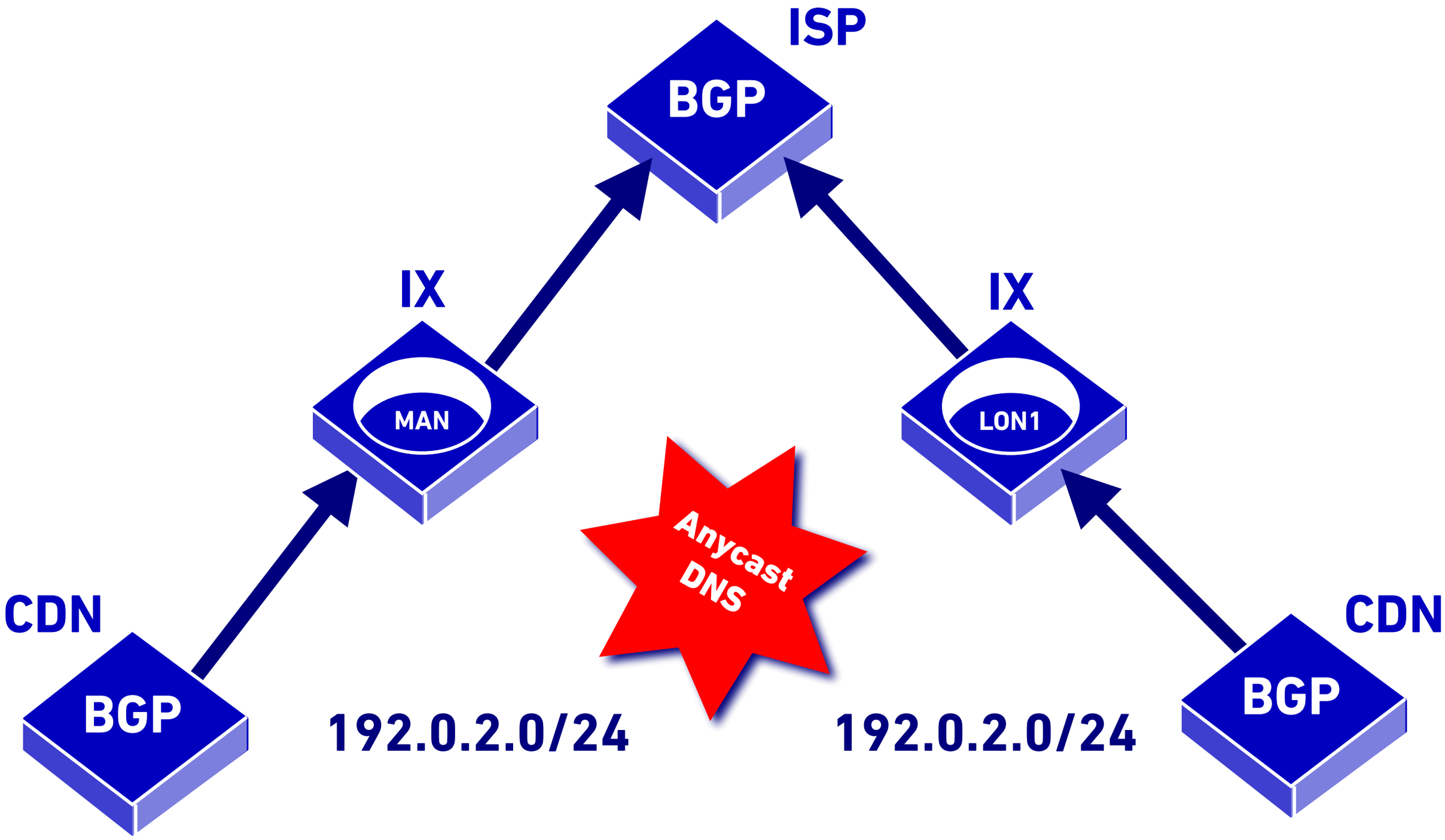
CDN Assumptions



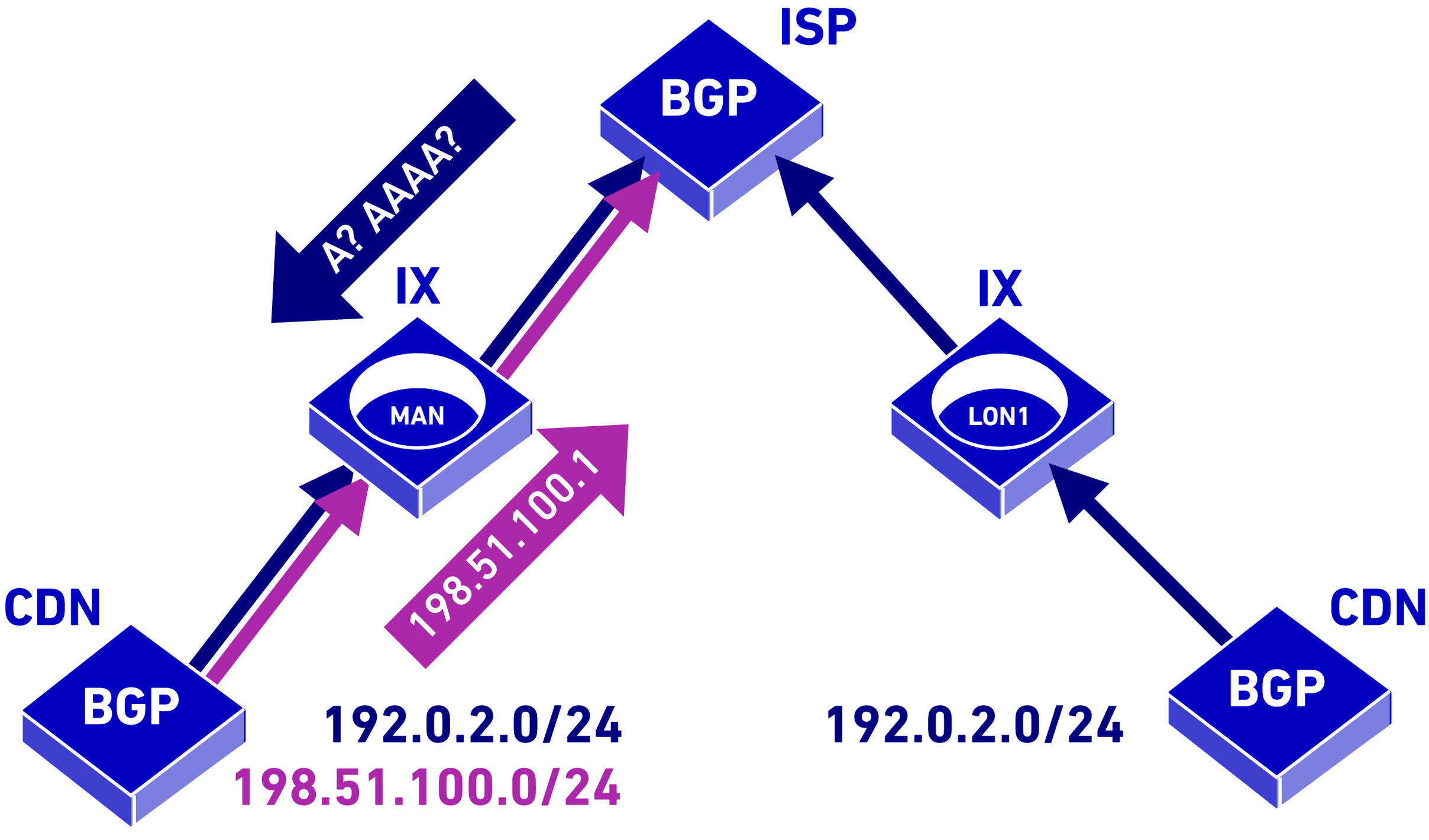
CDN Assumptions



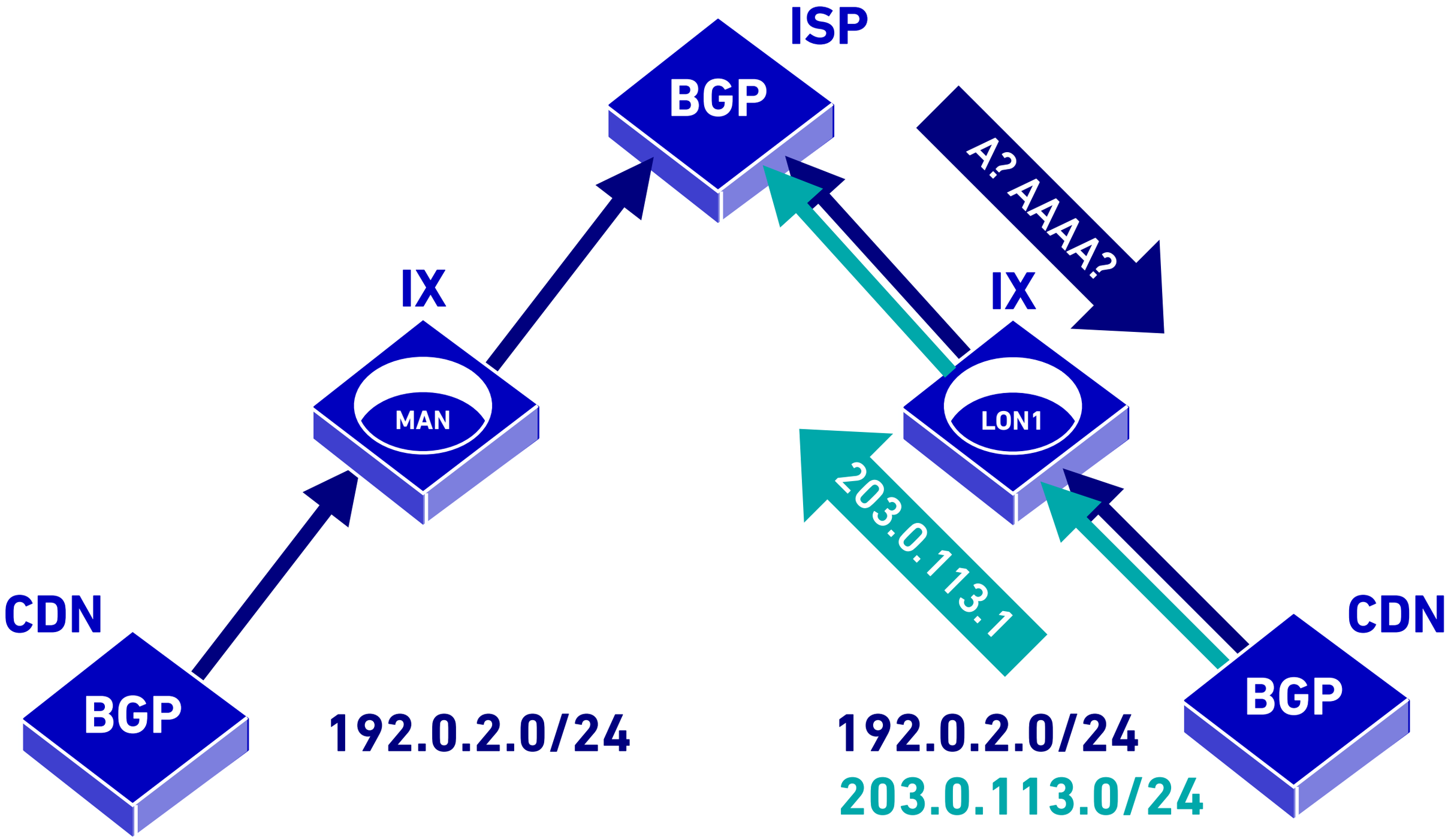
CDN Assumptions



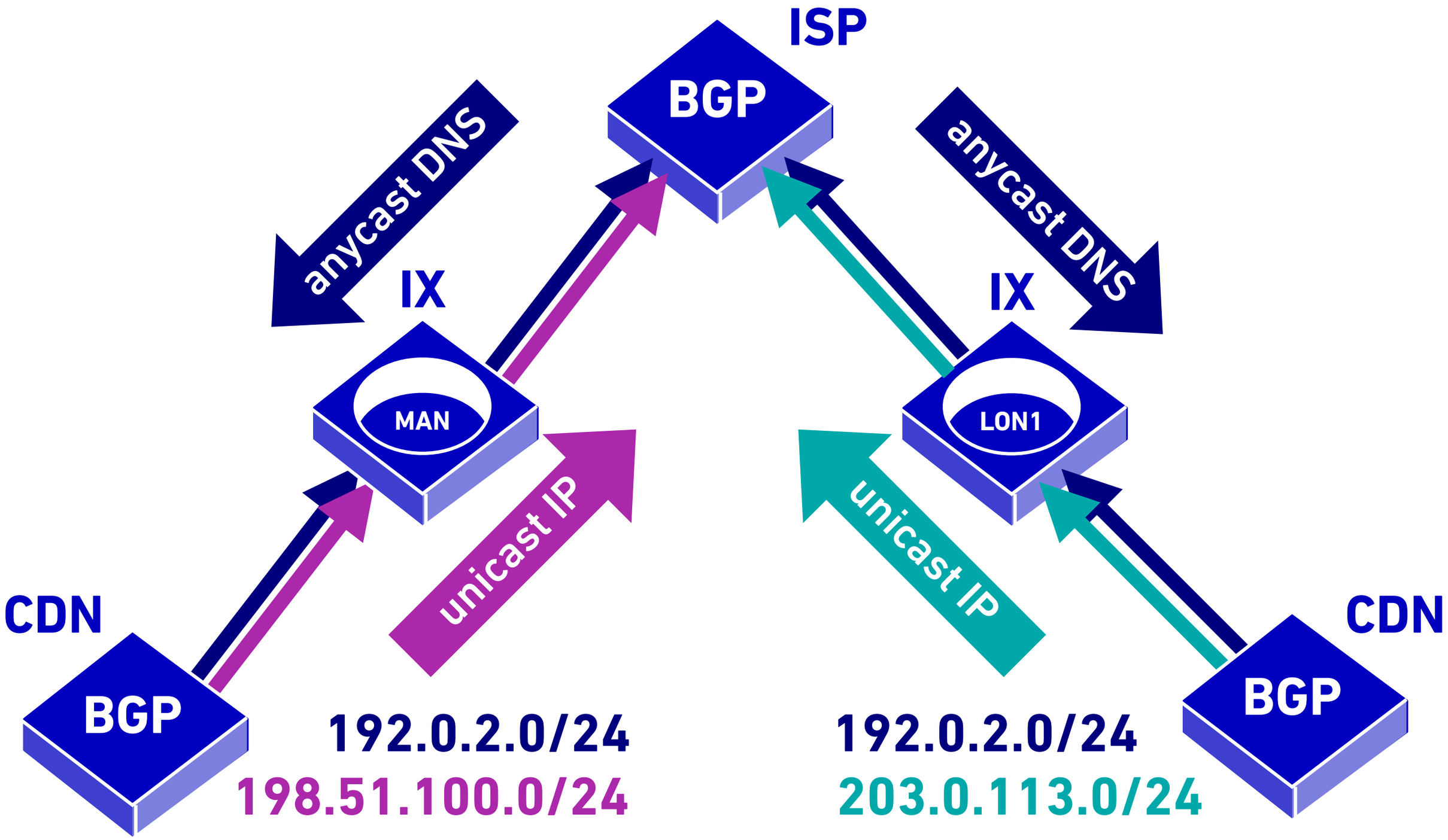
CDN Assumptions



CDN Assumptions



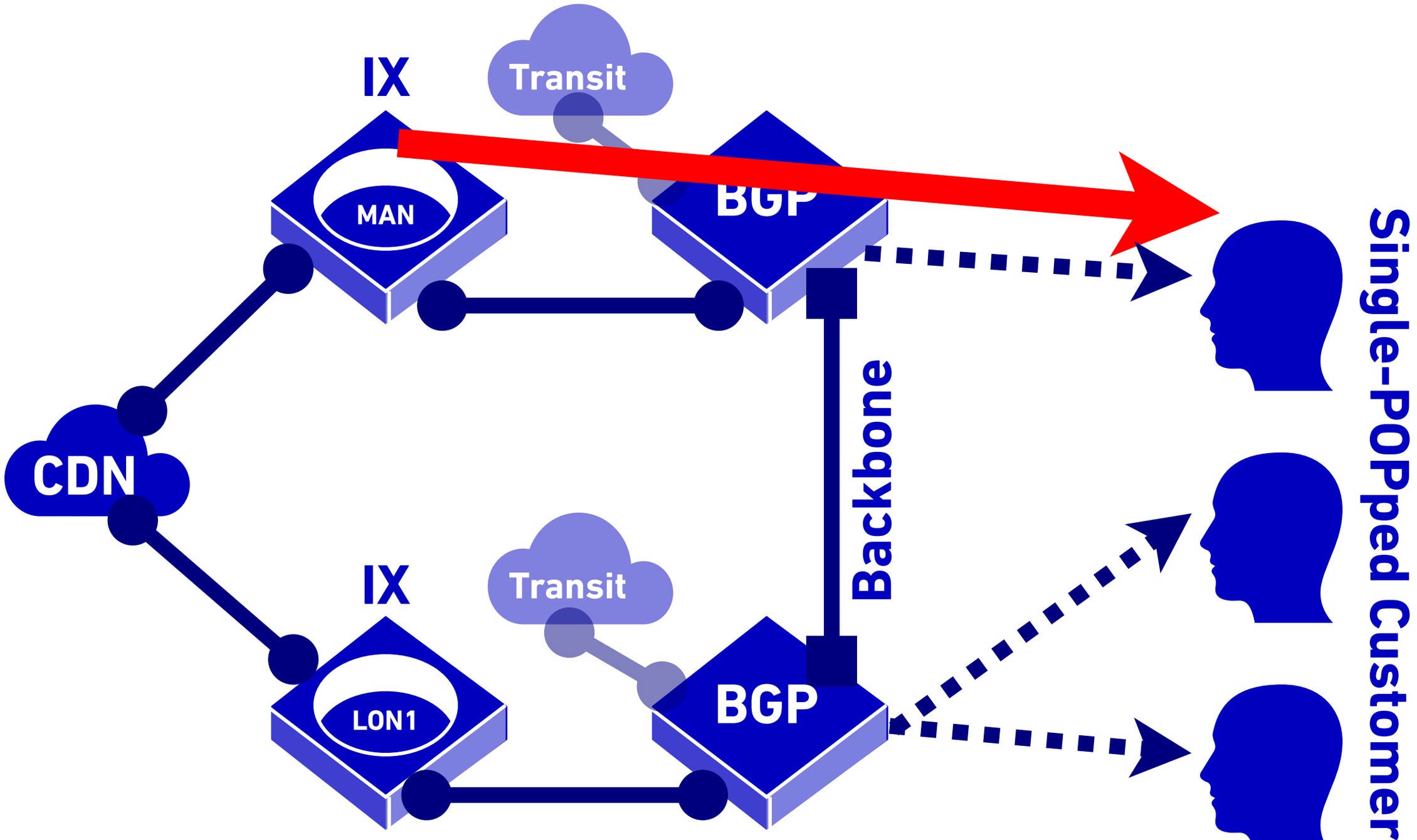
CDN Assumptions



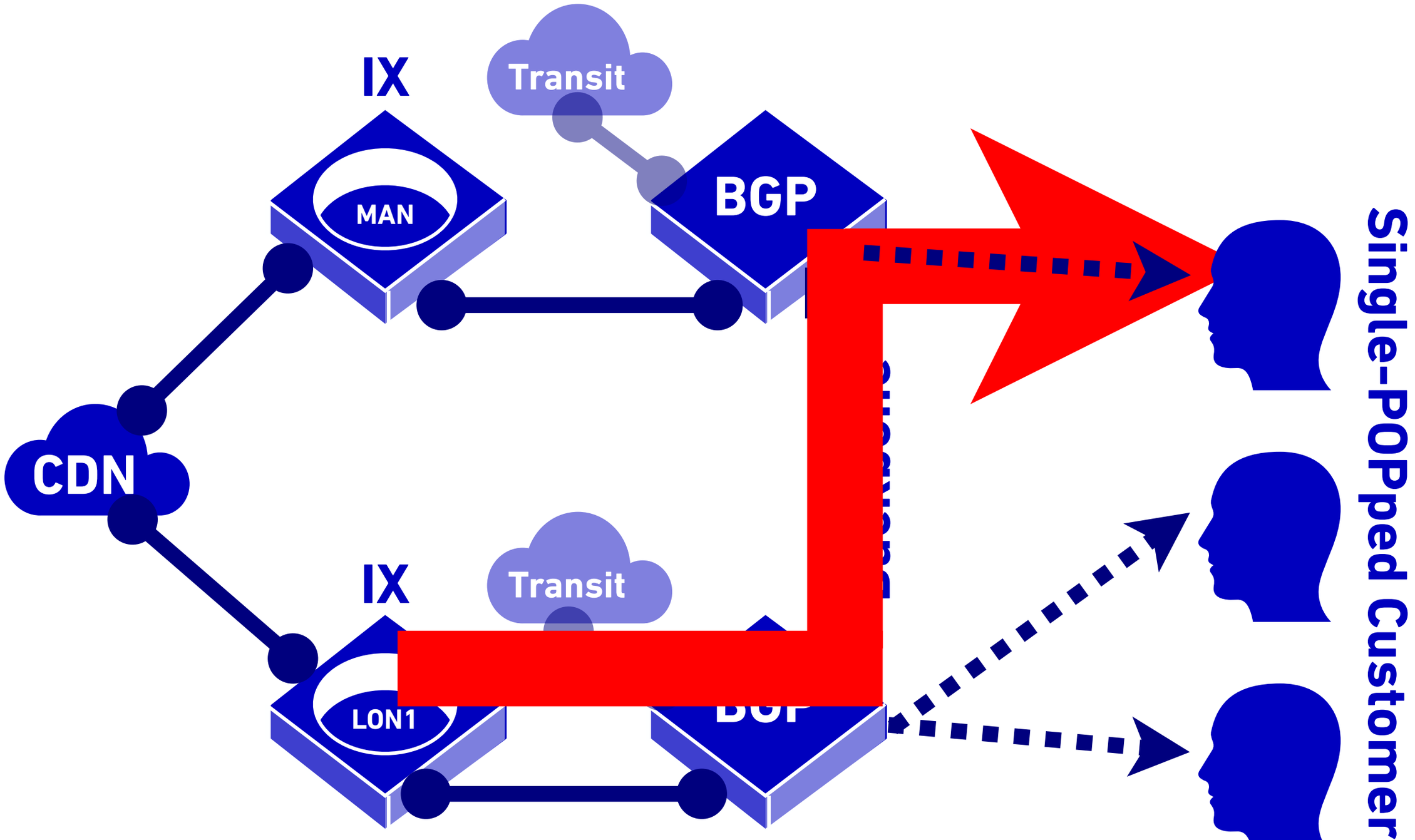
Inspiration

- Thomas Mangin (Exa Networks) LINX 96 talk about using DNS resolvers and announcing more-specific POP-local prefixes to attract inbound traffic to POPs.
- Most CDNs will accept more specifics.
  - CDNs probably build prefix-list filters from your RIPE DB route objects, and may not apply “le 24”.
  - But not all CDNs do iBGP on backbone, so some CDN POPs will only receive your aggregates, and serve traffic to you anyway!

That Helps With...

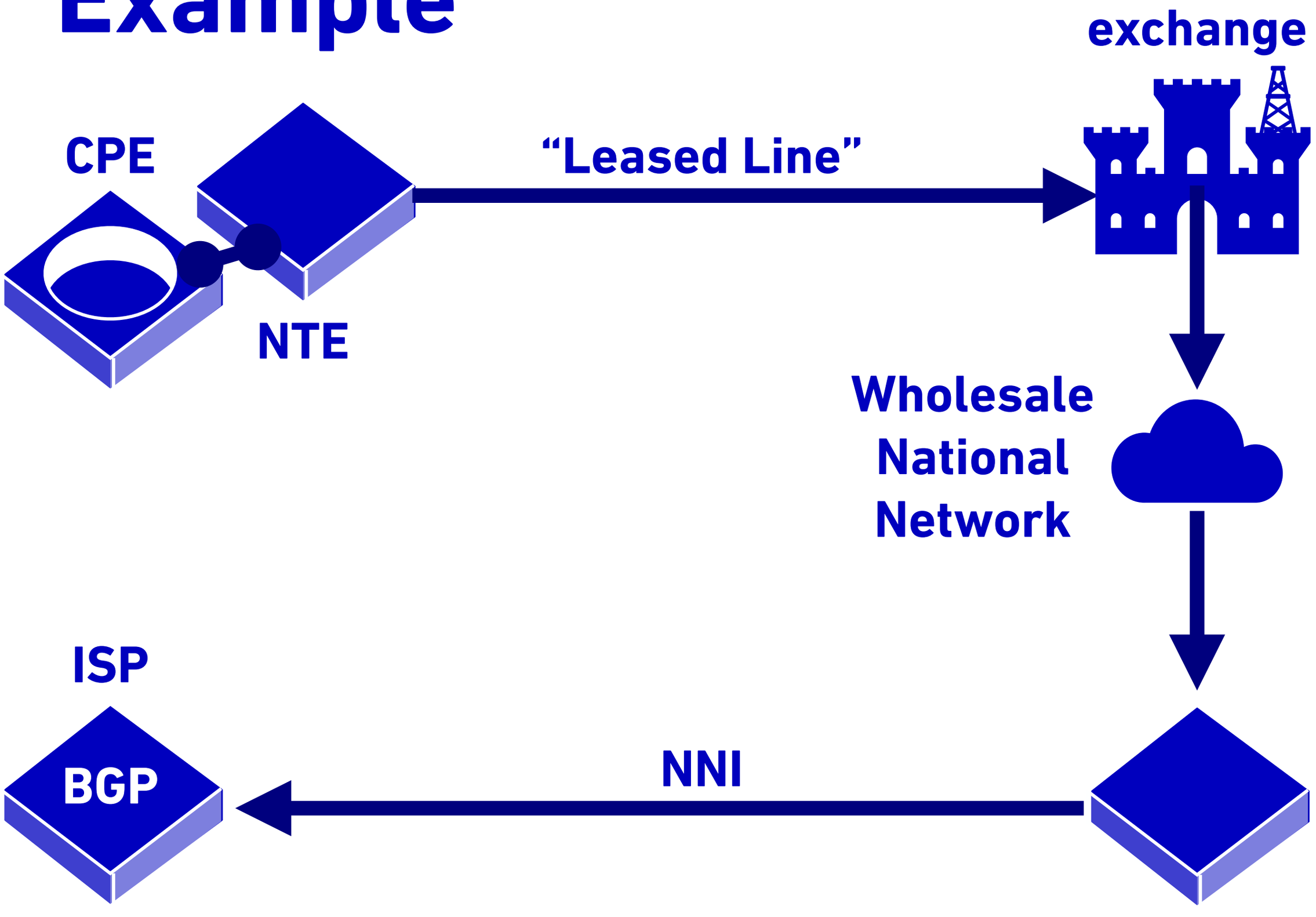


But Doesn't Help With...

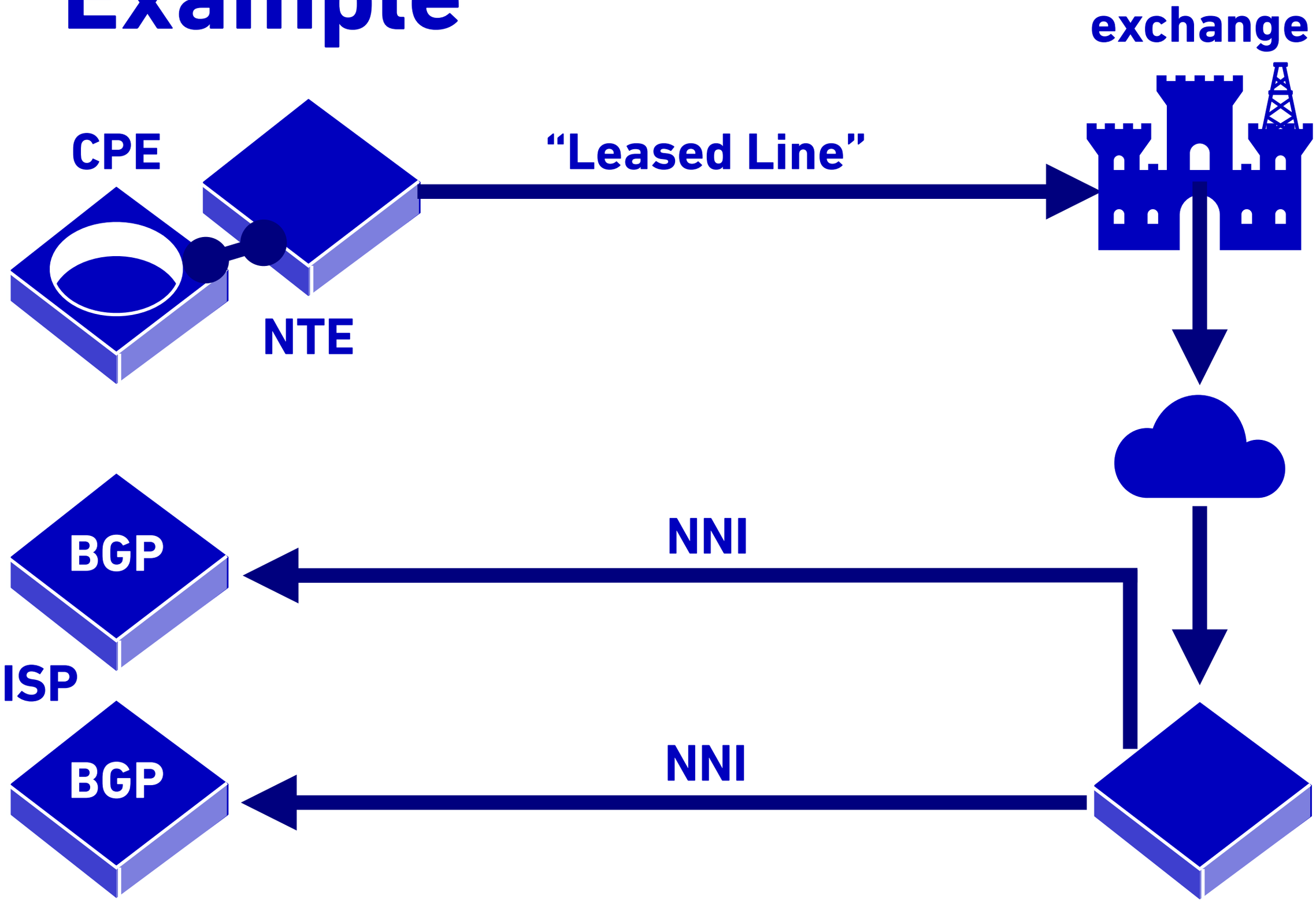


NATIONAL ETHERNET

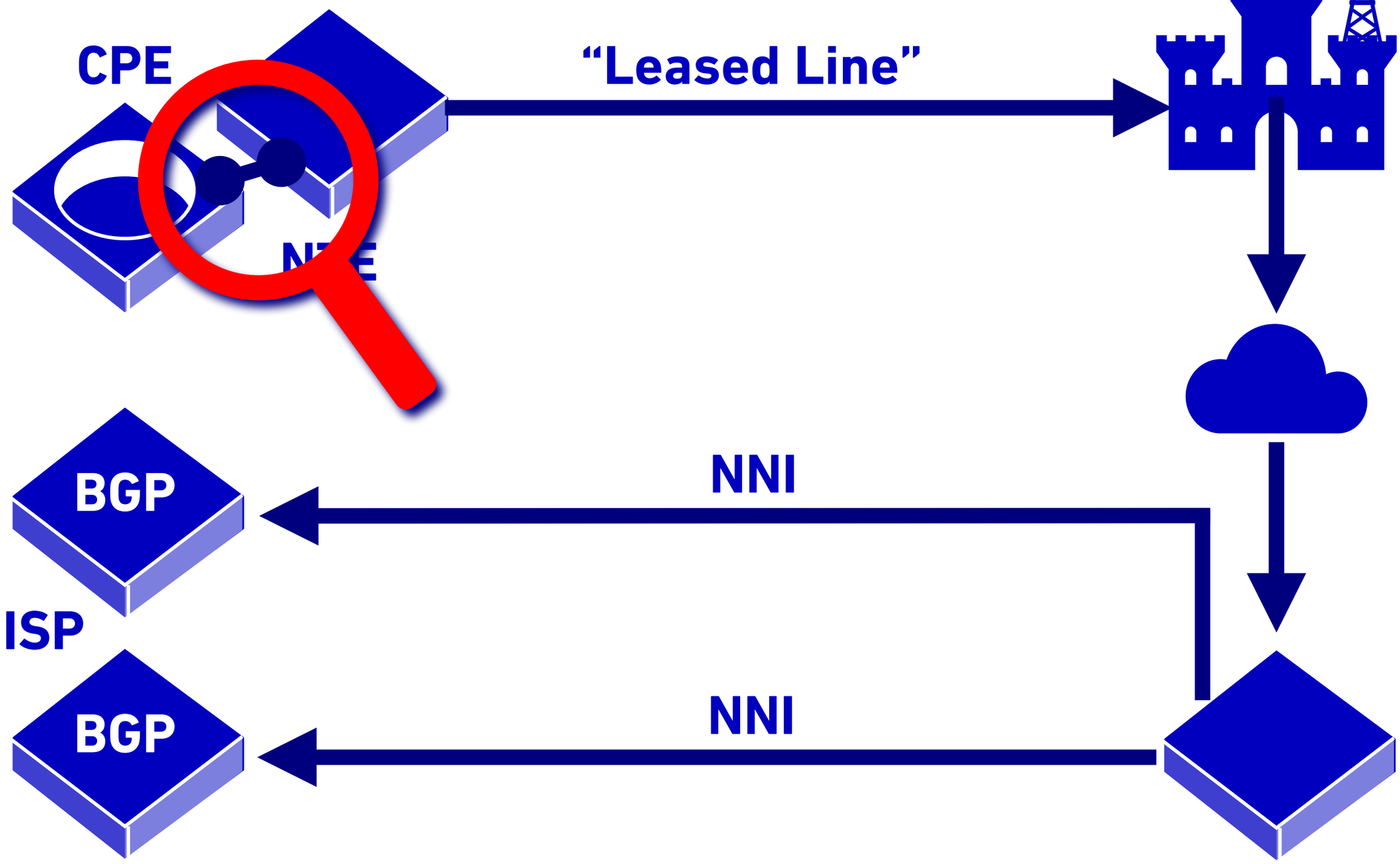
Example


exchange
CPE
“Leased Line”
NTE
Wholesale
National
Network
ISP
BGP
NNI

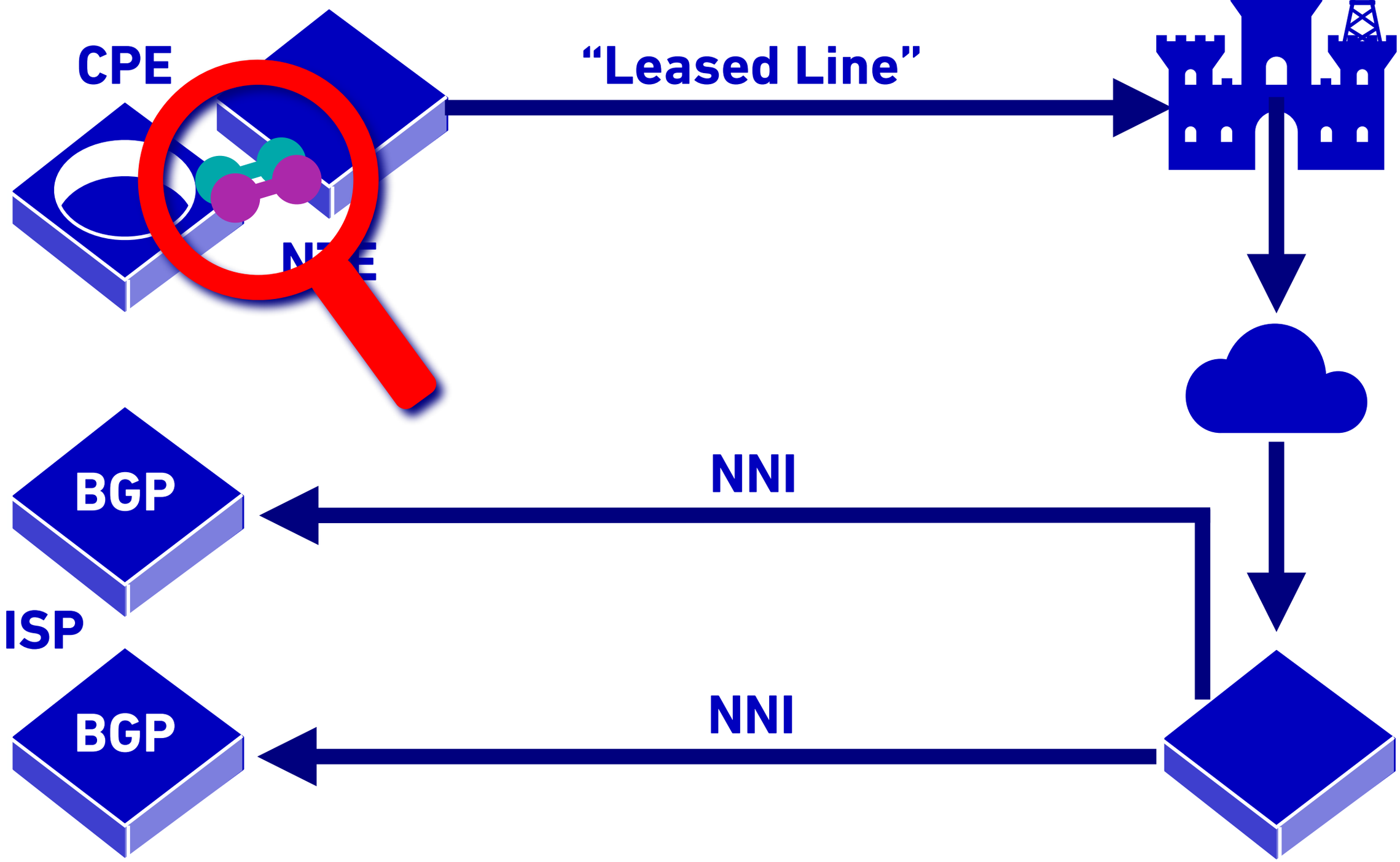
Example



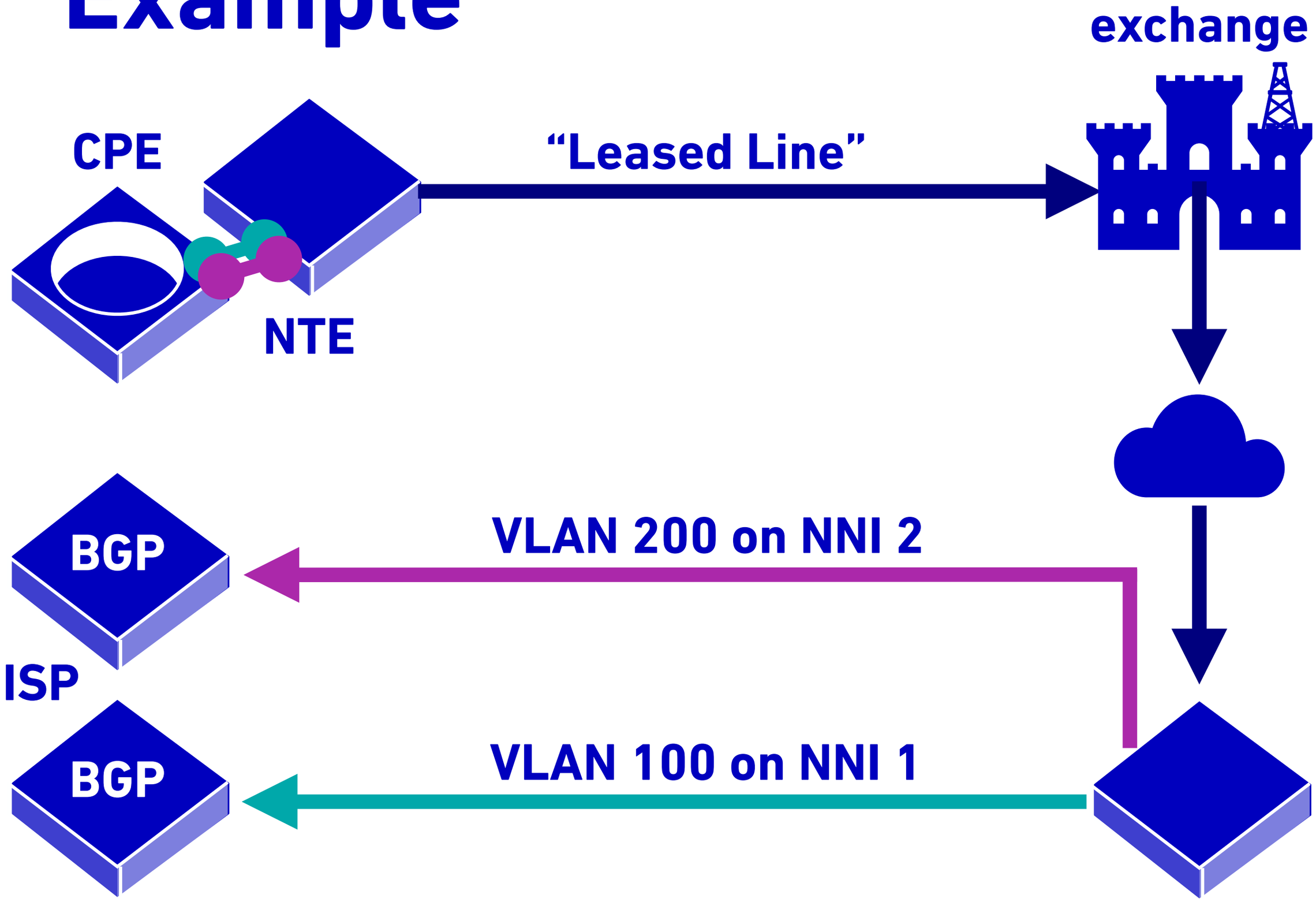
Example



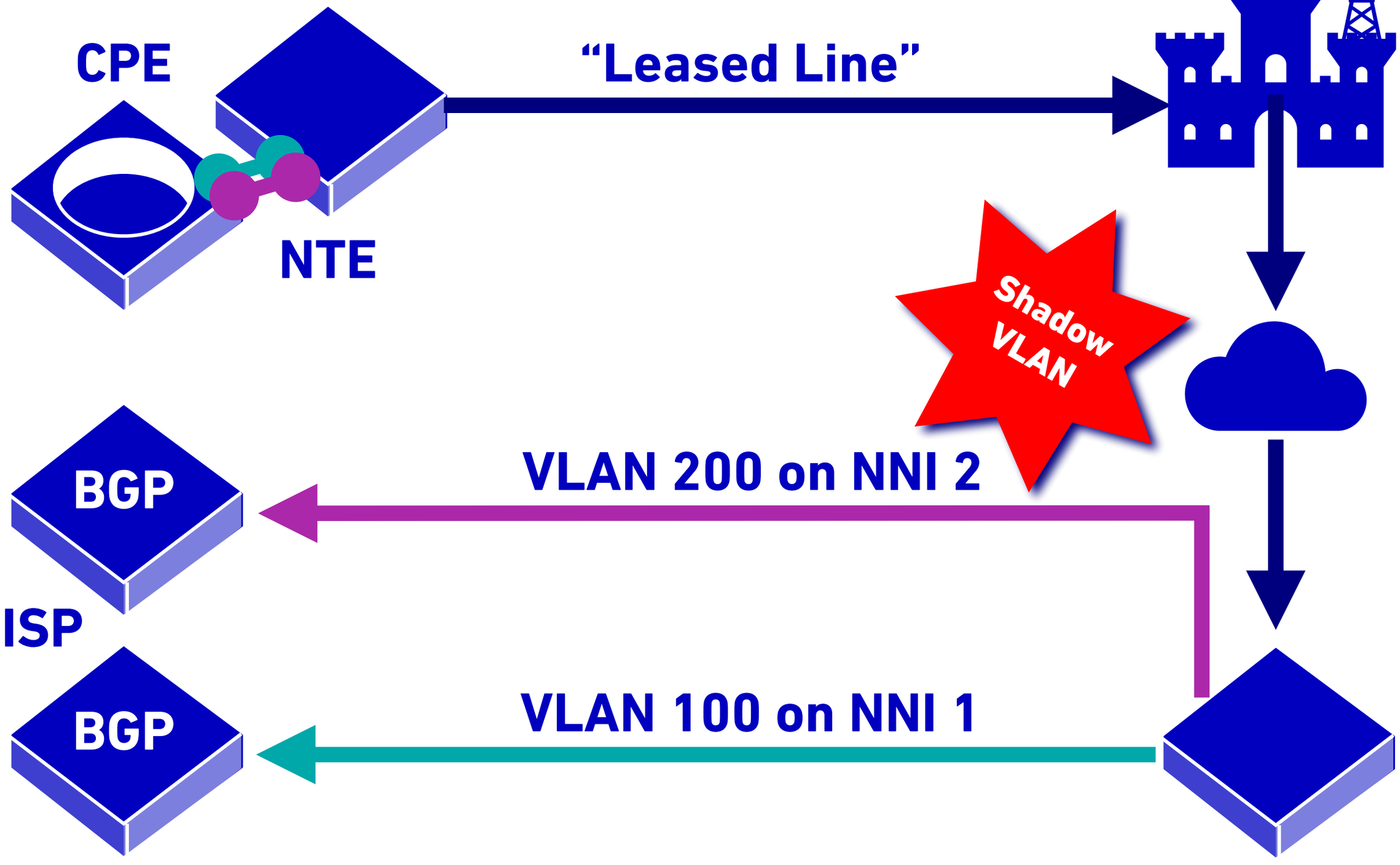
Example



Example


exchange
CPE
“Leased Line”
NTE
BGP
VLAN 200 on NNI 2
ISP
BGP
VLAN 100 on NNI 1

Example

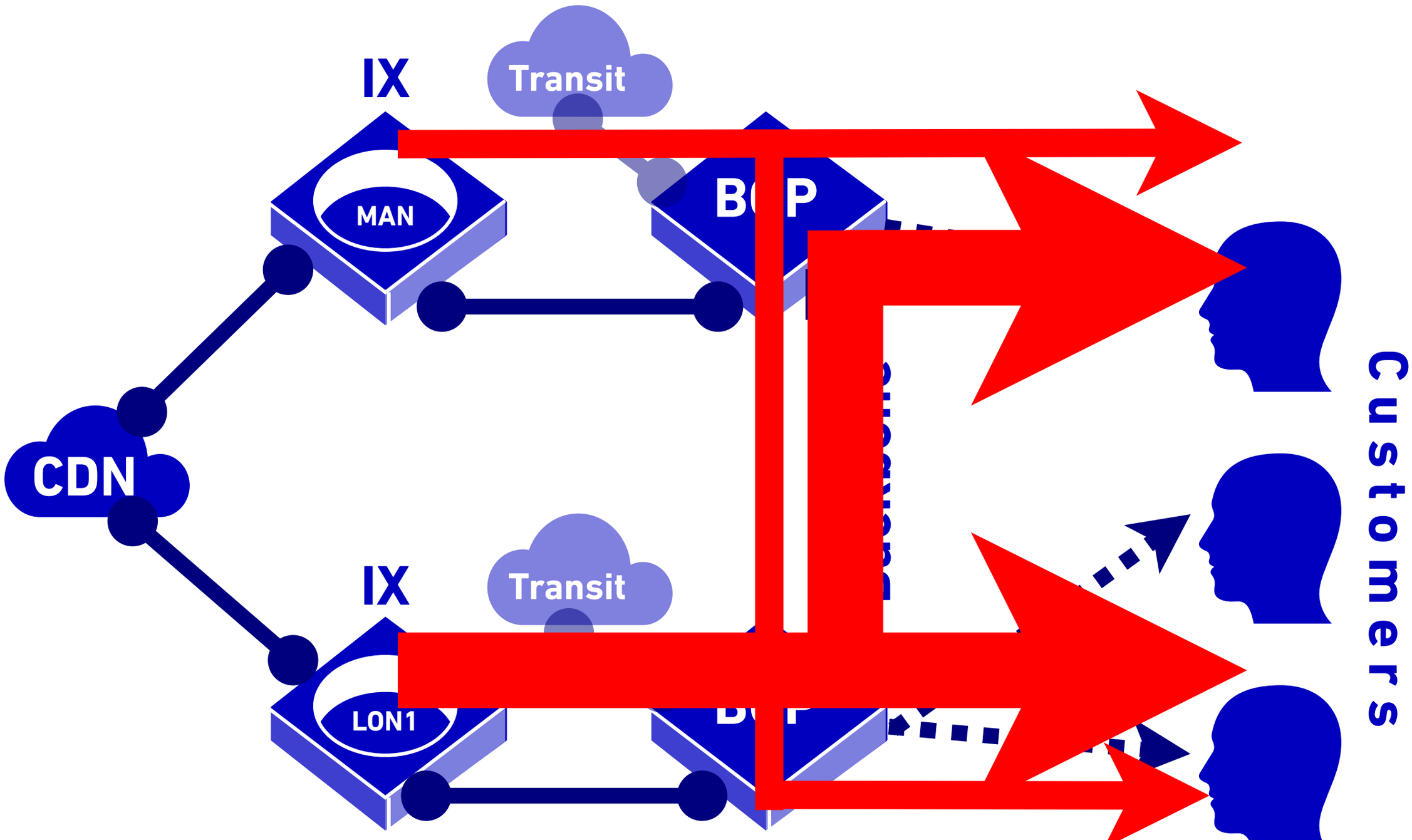

exchange
CPE
"Leased Line"
NTE
Shadow VLAN
BGP
VLAN 200 on NNI 2
ISP
BGP
VLAN 100 on NNI 1

NOT-SO-SHADOW VLAN

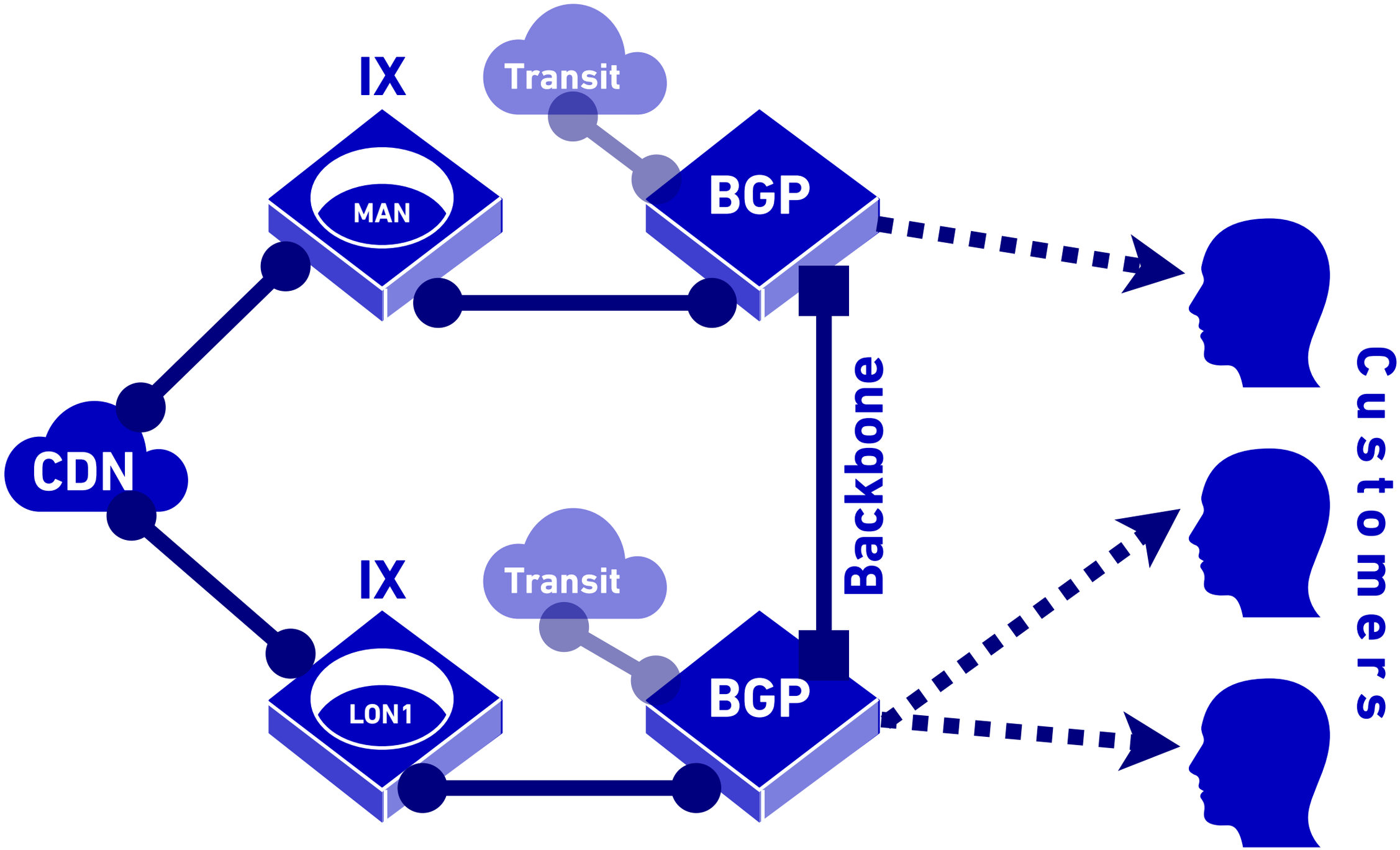
Lightbulb Moment

- National Ethernet providers (SSE, TTB, Virgin, others) market “shadow VLAN” for resilience.
- But pricing for a tail circuit does not vary depending on which NNI(s) it is presented upon.
- Cost of national backhaul is “baked in”.
- “Shadow” VLAN is allowed to be used for carrying production traffic alongside “Main” VLAN.
- So why not use both?

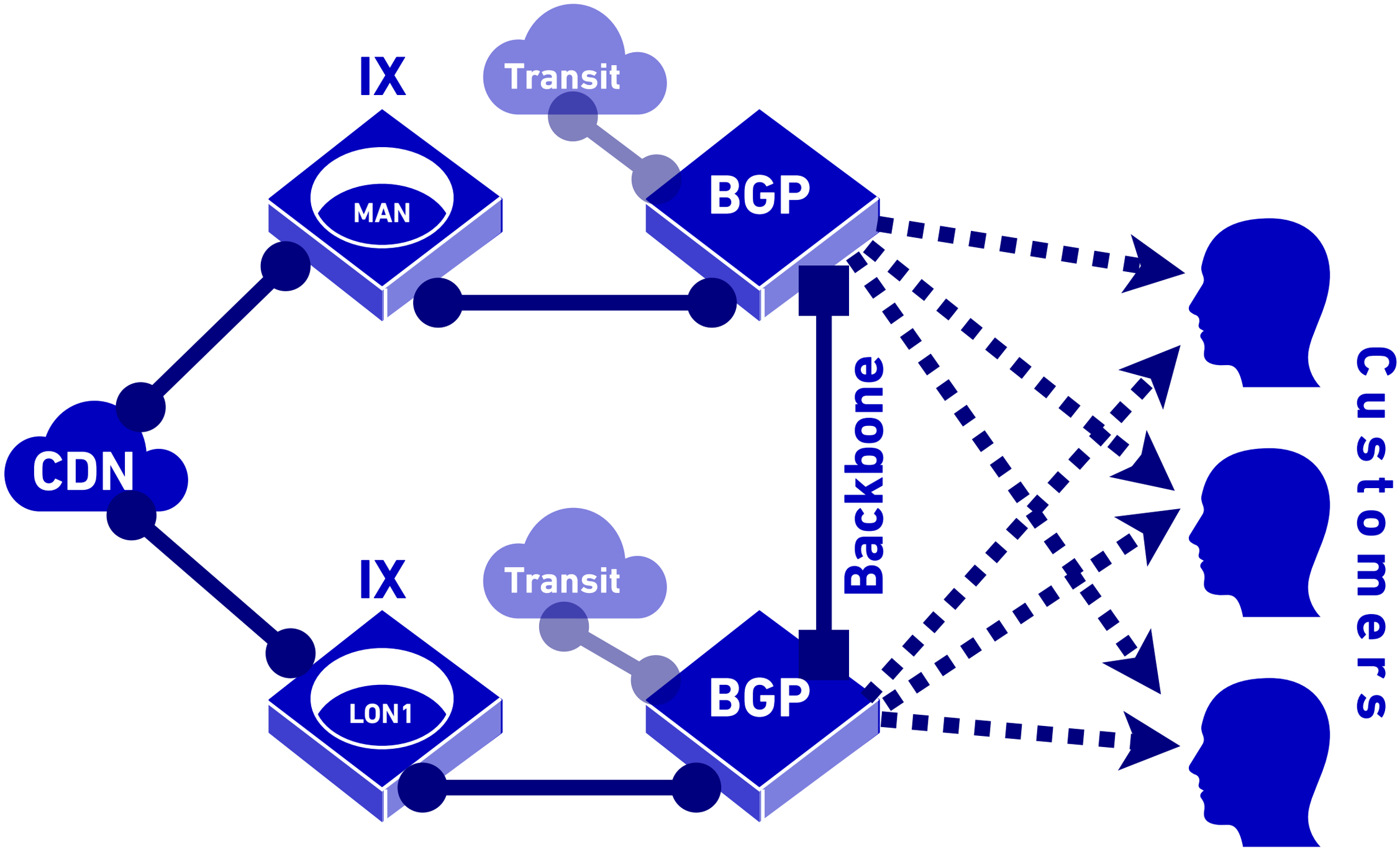
Challenge



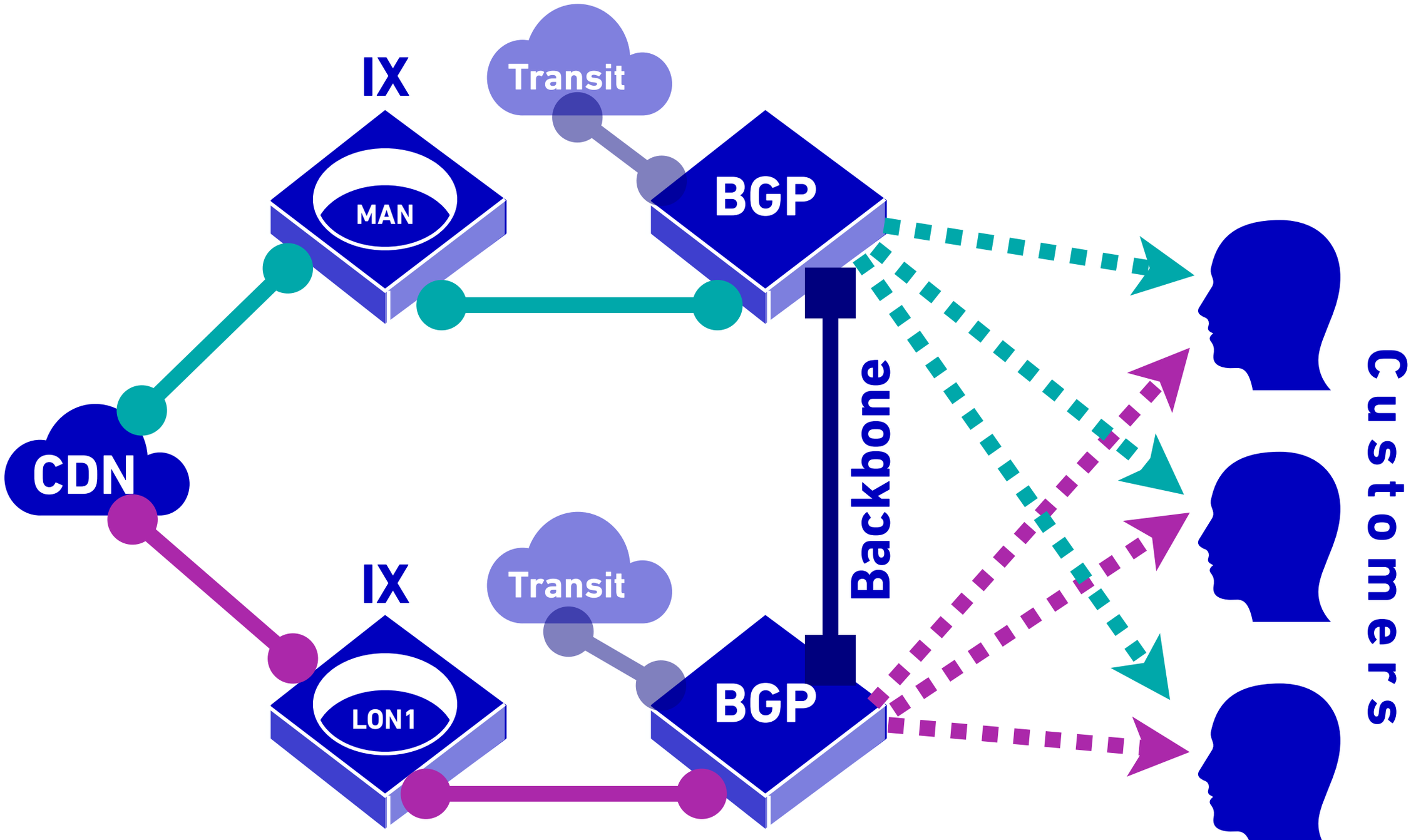
Challenge


IX
Transit
MAN
BGP
CDN
Backbone
Customers
IX
Transit
LON1
BGP

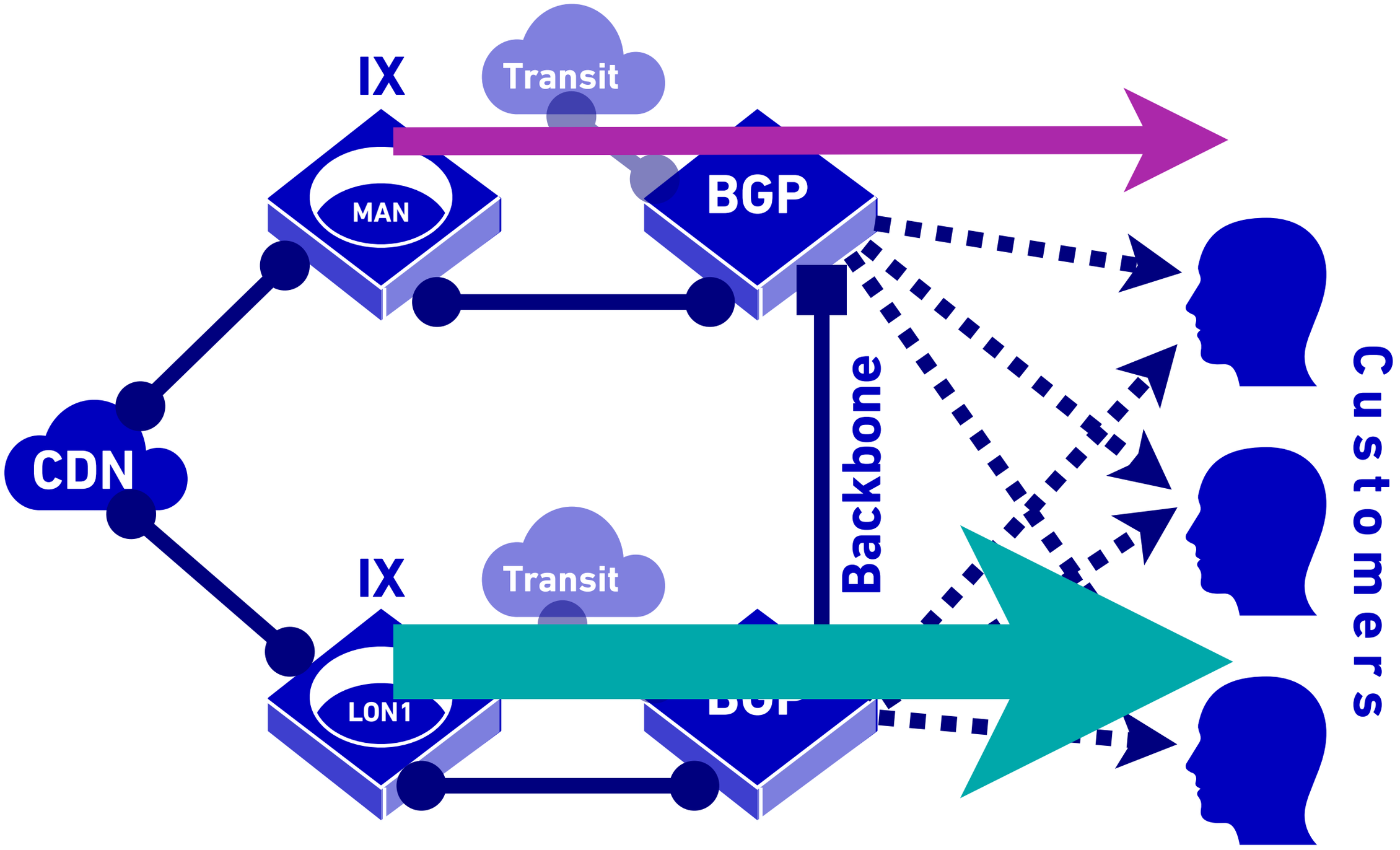
Shadow VLANs



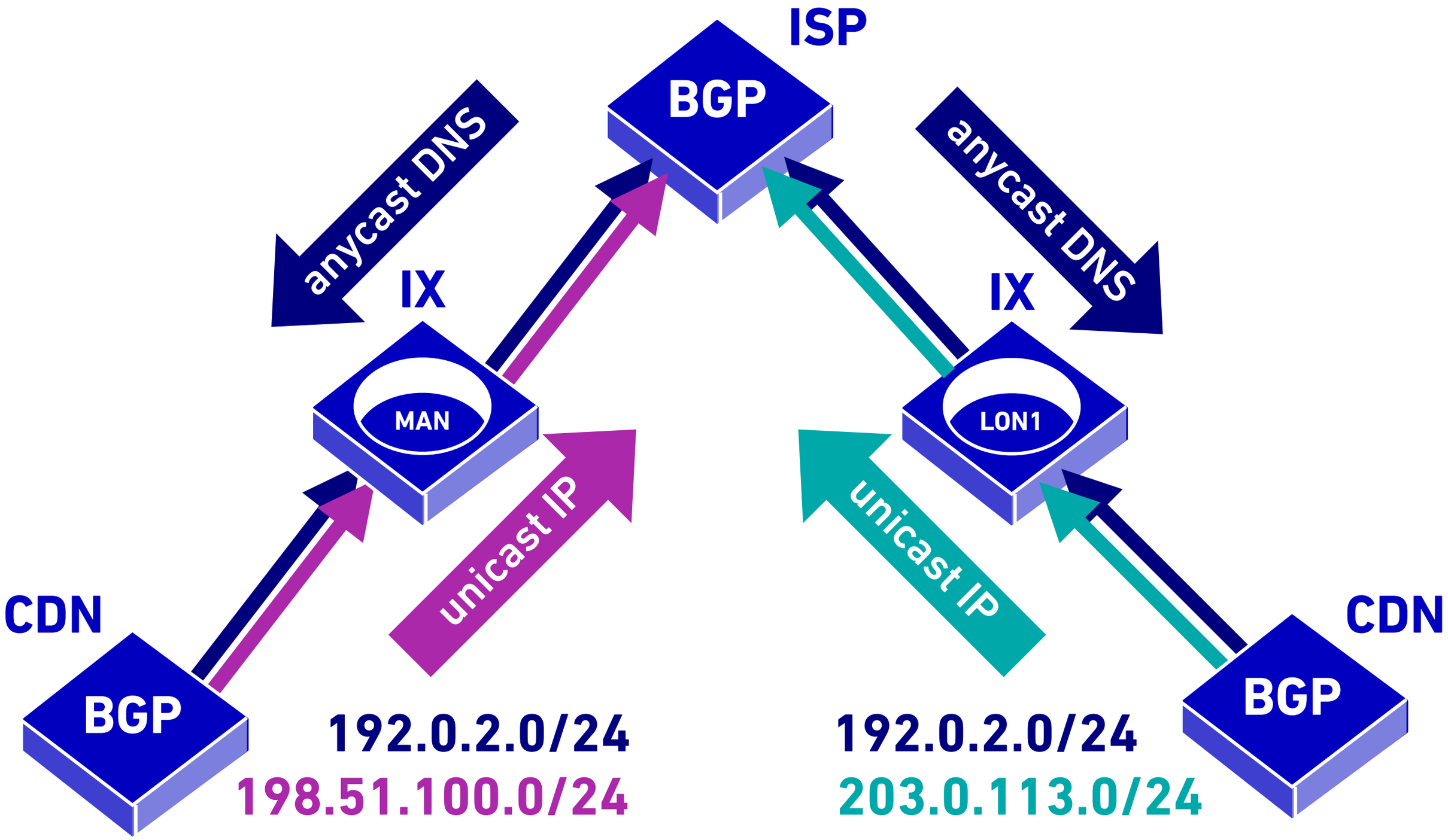
Not-So-Shadow VLANs



Net Result

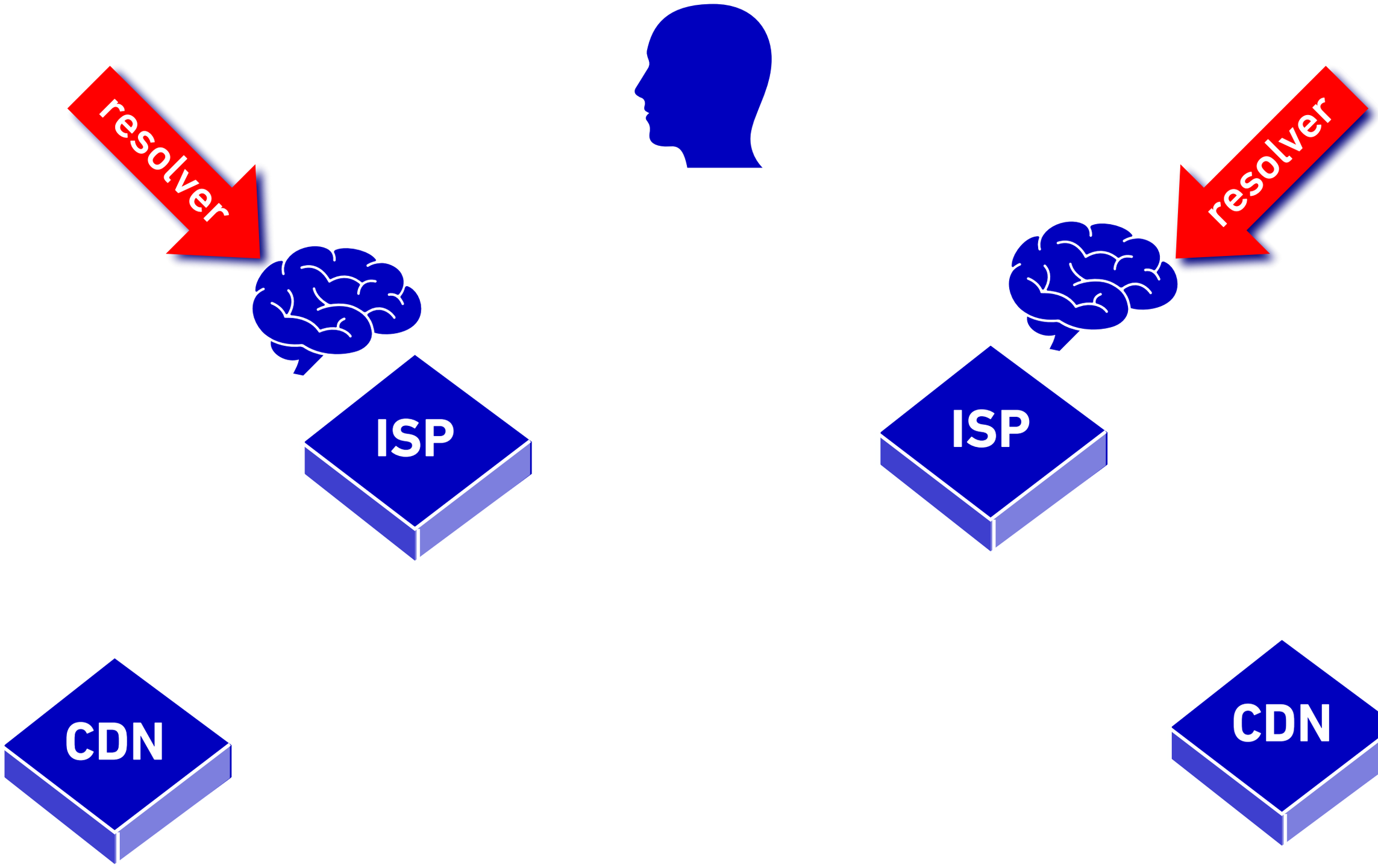

IX
Transit
MAN
BGP
CDN
IX
Transit
LON1
BGP
Backbone
Customers

But What About Anycast?

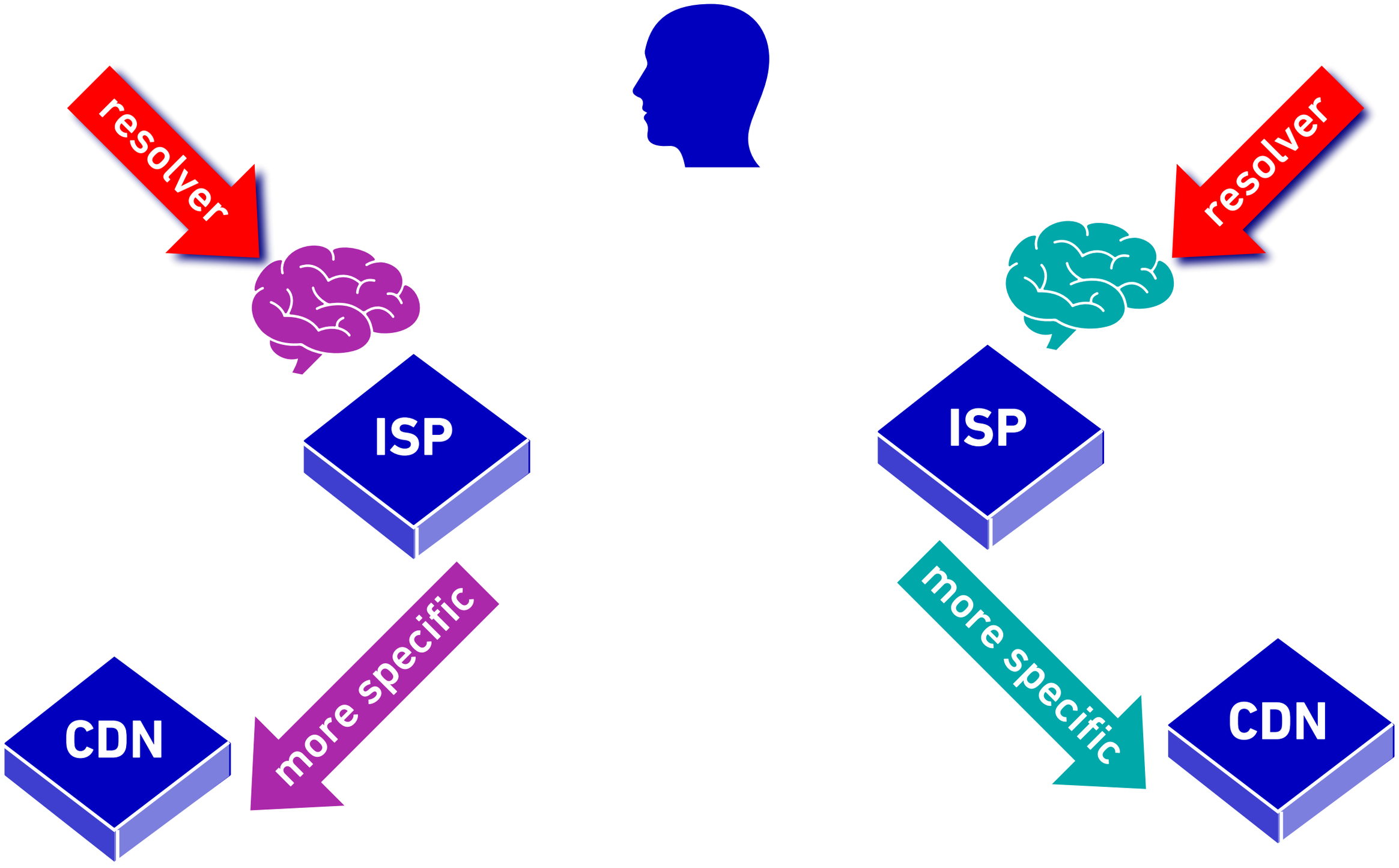


ANYCAST vs ANYCAST

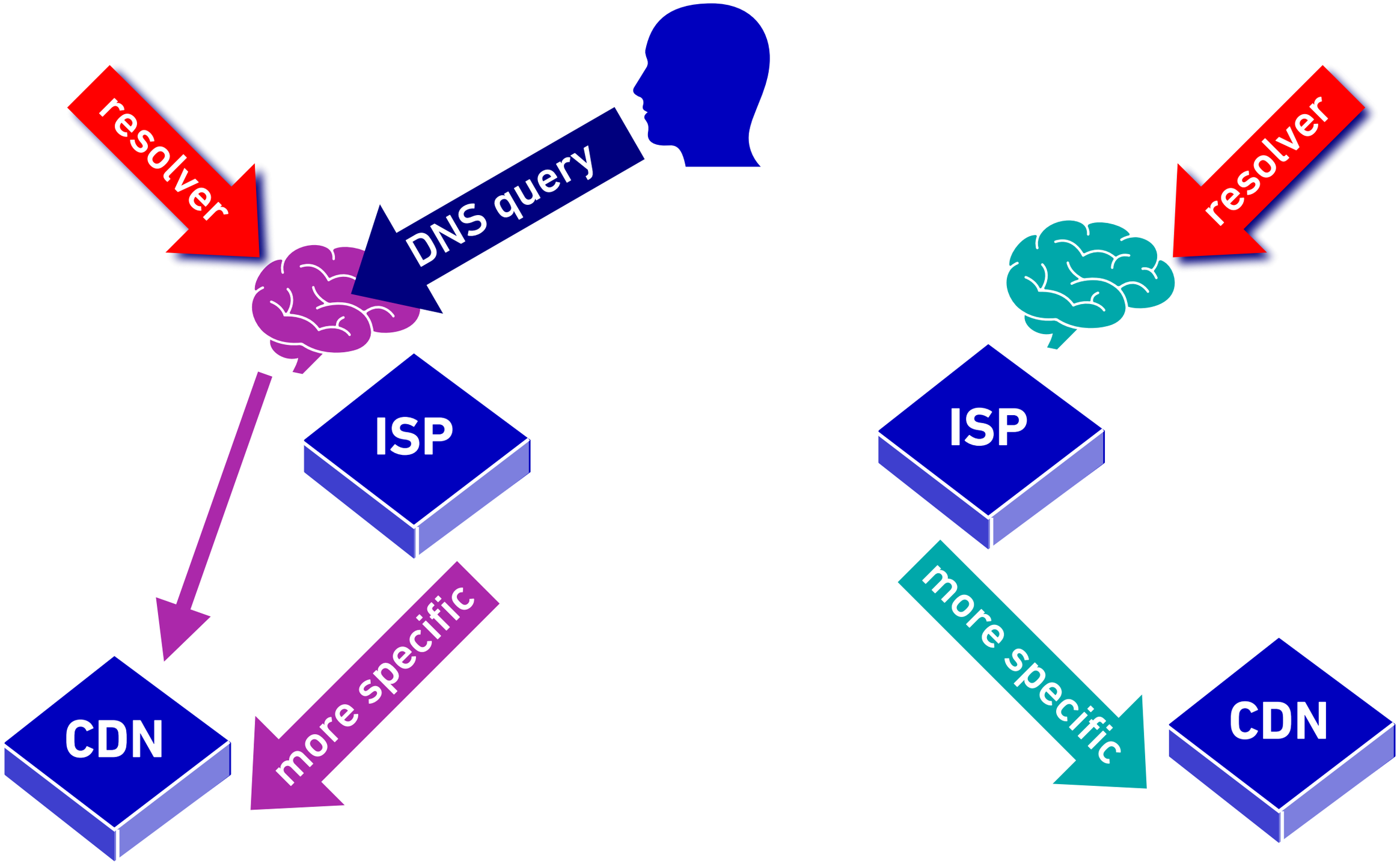
DNS Assumptions



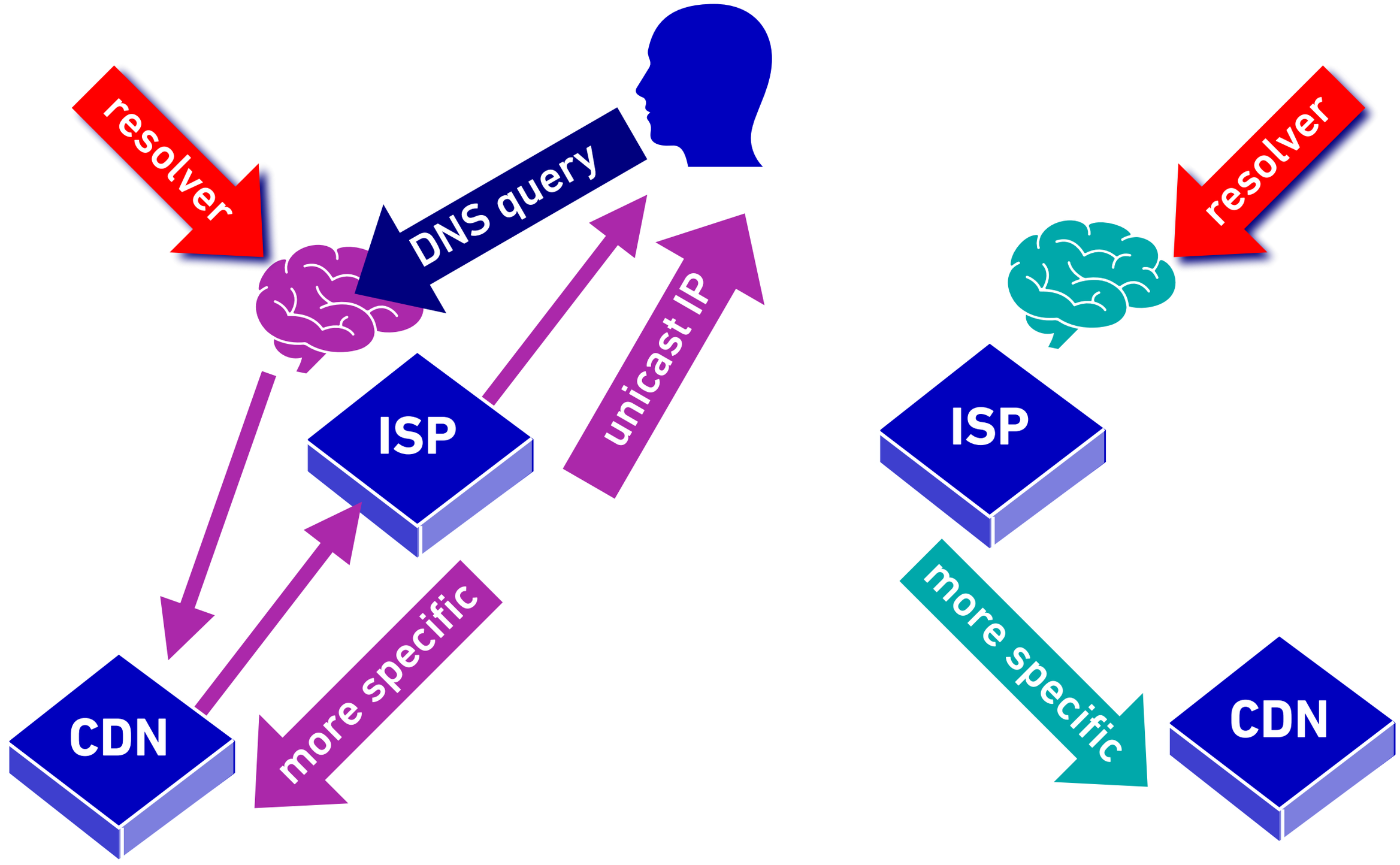
DNS Assumptions



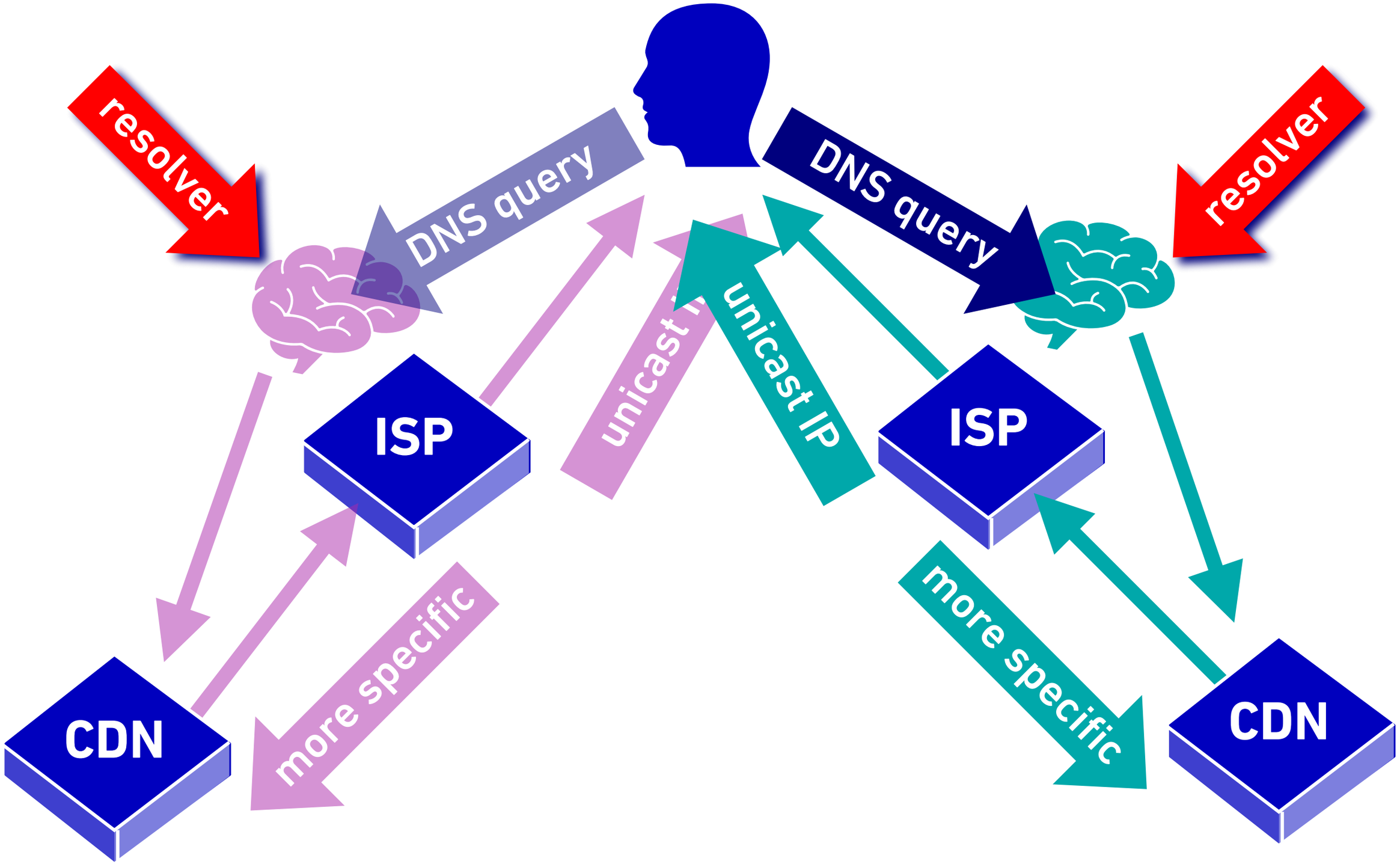
DNS Assumptions



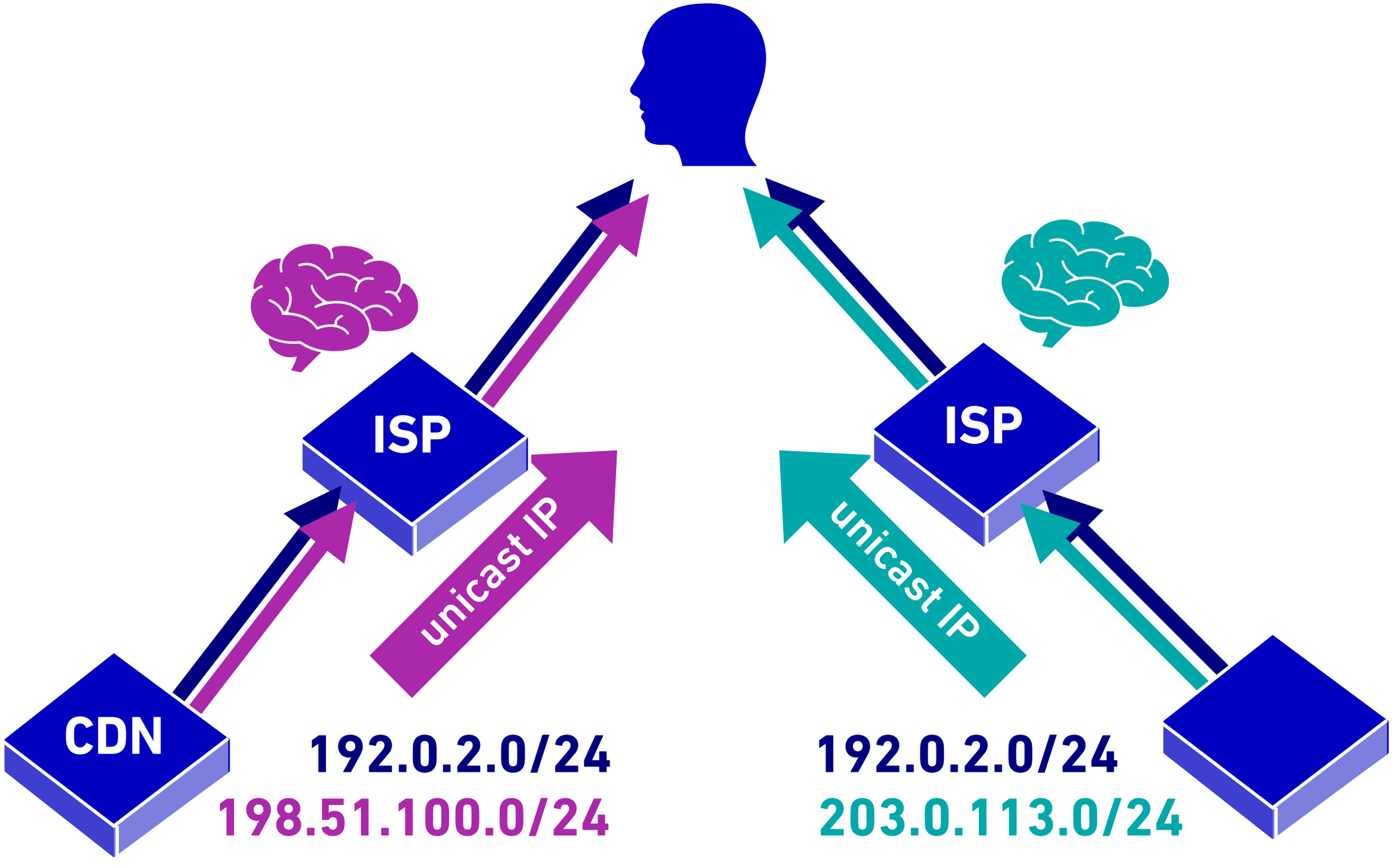
DNS Assumptions



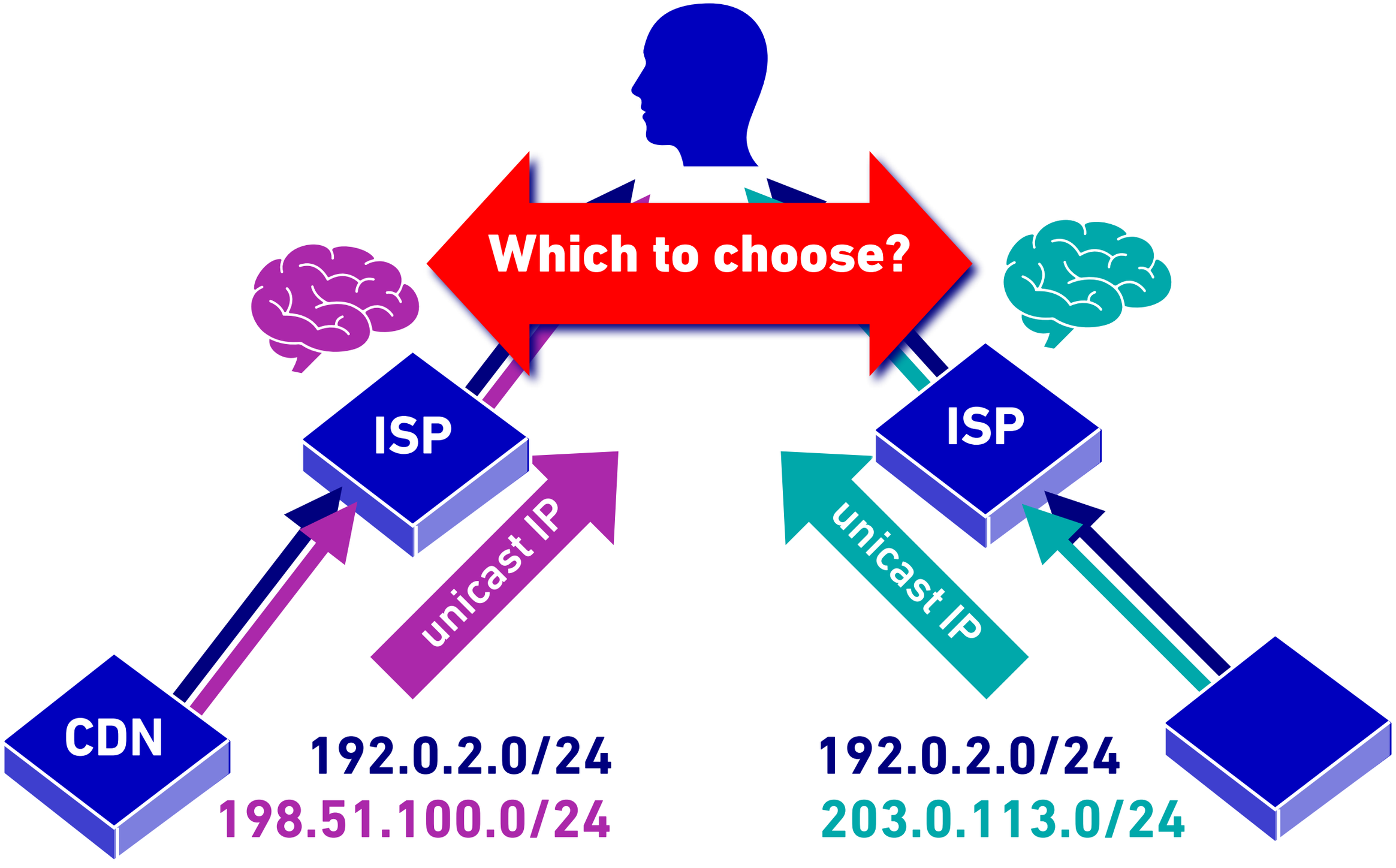
DNS Assumptions



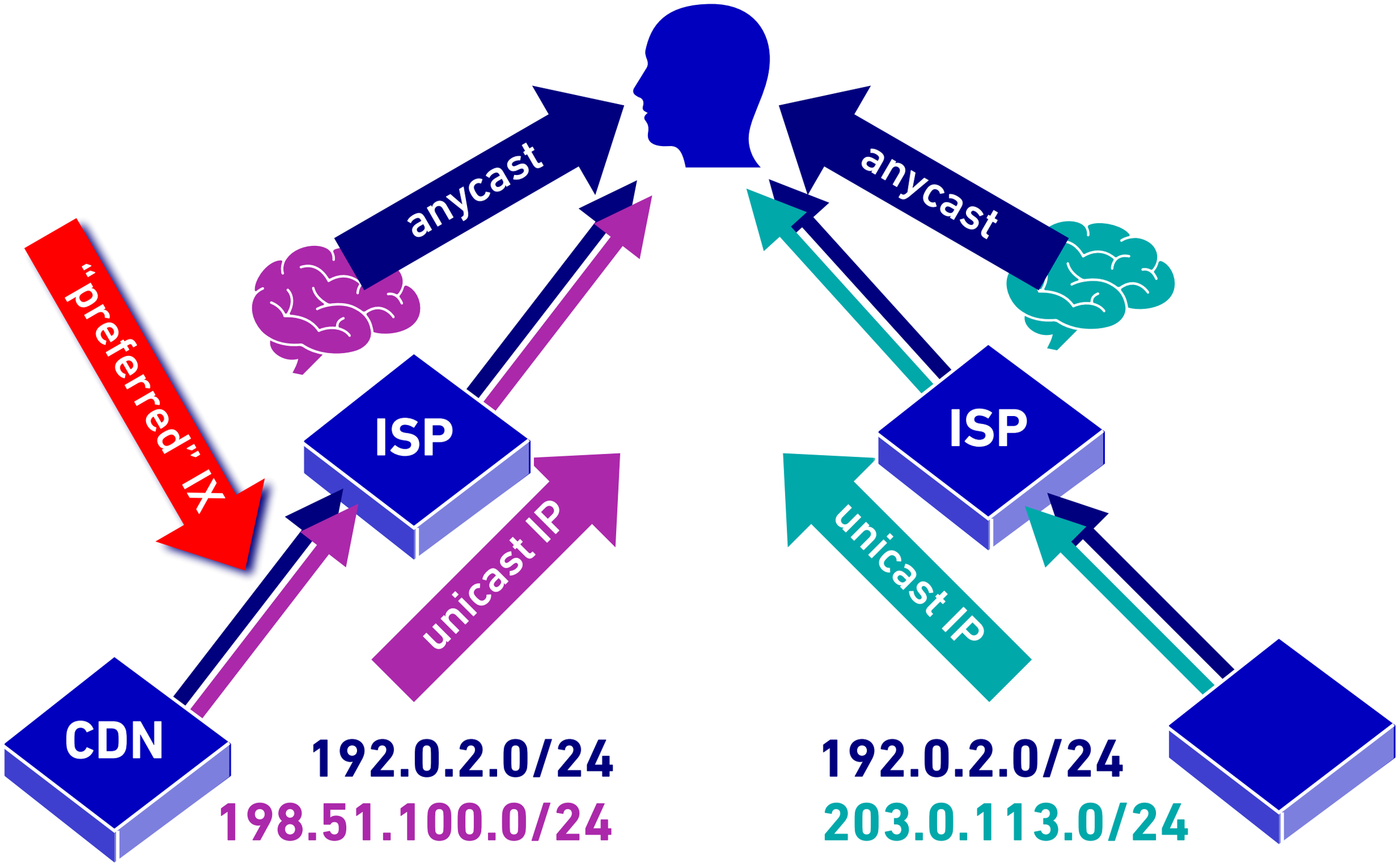
DNS Assumptions



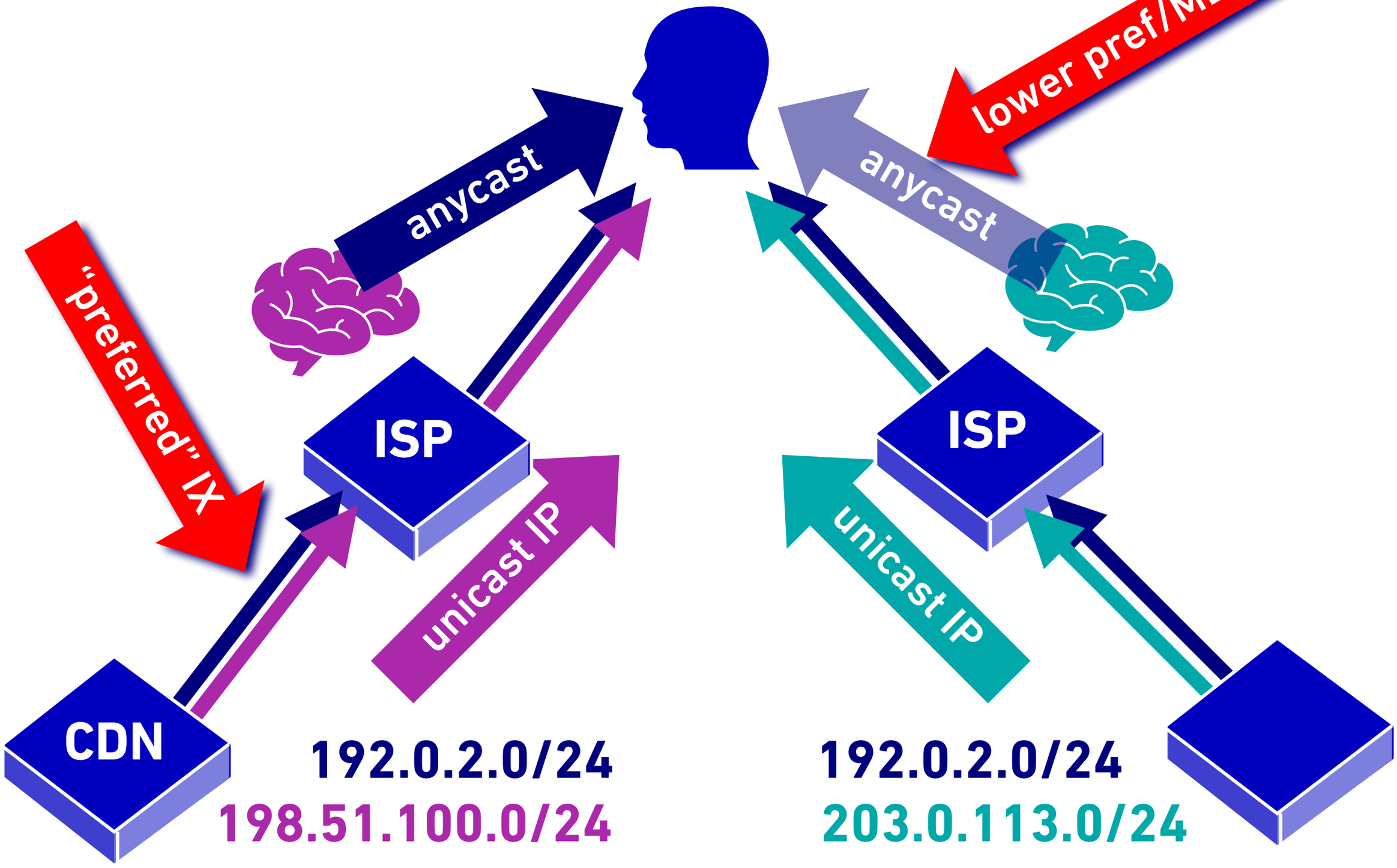
DNS Assumptions



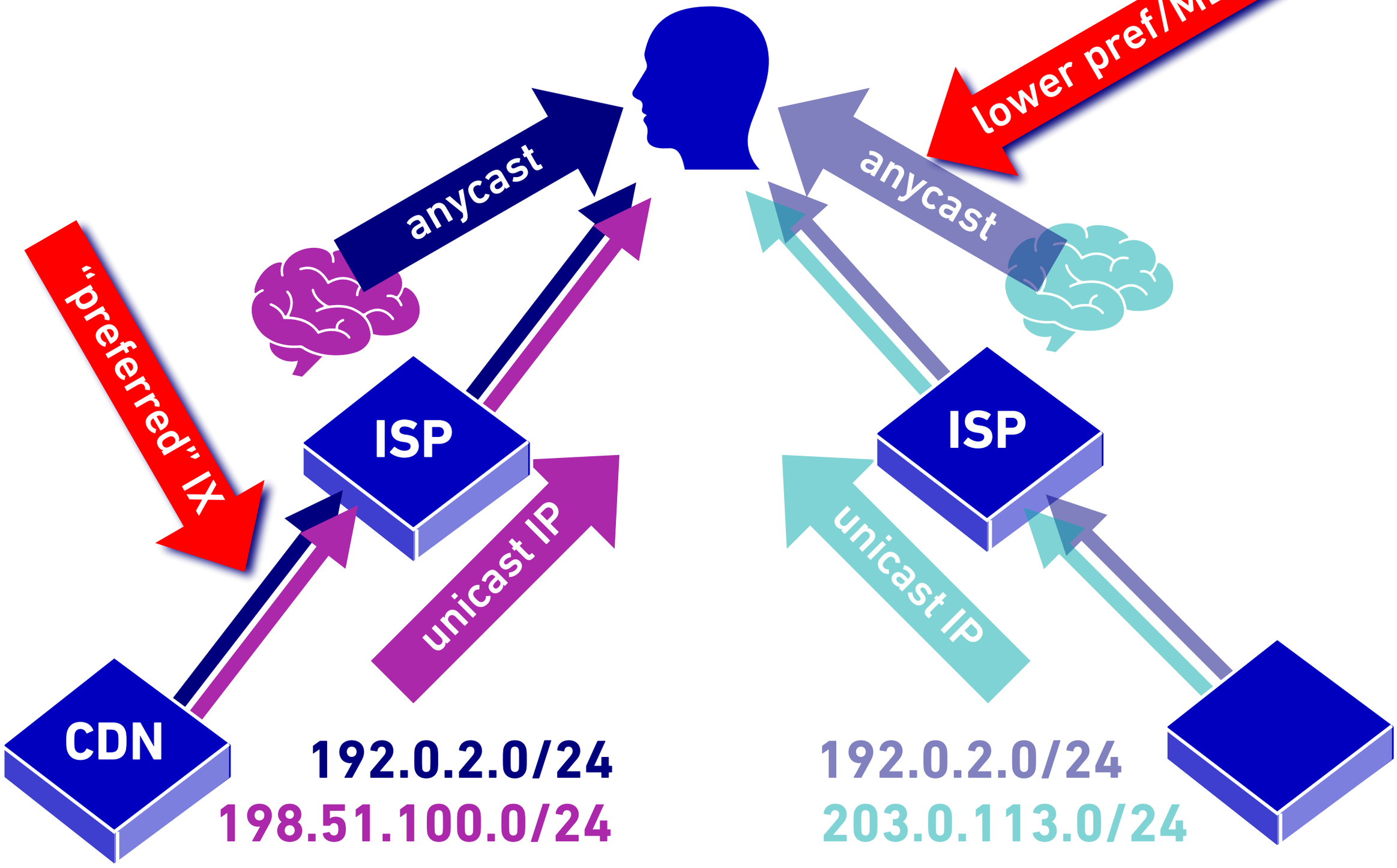
DNS Assumptions



DNS Assumptions



DNS Assumptions



Challenges

- Are you able to notionally allocate a /24 per POP?
  - Do you have enough /24s?
  - Are your customers' addresses already allocated (e.g. sequentially) and there is no locality to /24s?
- What do you do when you run out of addresses for customers from one POP's /24?
- Is your IPAM, outgoing prefix-list and route-map automation going to be able to accommodate this?

BGP CDN SHADOW VLAN

But What About Anycast?

- CPE no longer has single default route.
  - Or accepting default from multiple BGP sessions.
- Which of main/shadow VLANs is preferred “default”?
  - Closest?
  - Most IX traffic?
- Want to also propagate “golden subnets” to CPEs:
  - Including CDNs’ anycast DNS prefixes.
  - Destinations for outbound traffic.

Golden Prefixes

- We already use BIRD route servers.
- We have a route server in each POP.
- We already automate the RS configuration and deployment using SaltStack.
- Simple enough to add a special filter:
  - Was this prefix learned from IX/PNI near this NNI?
  - Is this prefix “golden” CDN/anycast/DNS?
  - Or a “big destination”?

How We Decide “Golden”

- Do we peer with origin ASN on LON1/2 and MAN?
  - If yes: accept both, maybe local-pref MAN.
- Is peering, and origin ASN in $list_of_CDN_ASNs?
  - If yes: accept, so CPE has more specific.
- Is peering, and origin ASN in “top 20 destinations”?
  - If yes: accept, so CPE prefers over default route.

BIRD: Golden Prefixes

```
function is_golden_cdn( bgppath p ) {
    {% for asn in salt['pillar.get']('golden_cdns',[]) %}
      if p ~ [= * {{ asn }} * =] then return true;
    {% endfor %}
    return false;
}
```

```
filter golden_cdn {
    if is_golden_cdn( bgp_path ) then {
      {% for prefix in salt['pillar.get']('local_ixs_ipv4',[]) %}
        if bgp_next_hop ~ {{ prefix }} then accept;
      {% endfor %}
    }
    reject;
}
```

BIRD: Golden Prefixes

- **function** is_golden_cdn(**bgppath** p) {
  {% for asn in salt['pillar.get']('golden_cdns',[]) %}
  if p ~ [= * {{ asn }} * =] **then return true**;
  {% endfor %}
  return false;
  }

- **filter** golden_cdn {
  if is_golden_cdn(**bgp_path**) **then** {
  {% for prefix in salt['pillar.get']('local_ixs_ipv4',[]) %}
  if bgp_next_hop ~ {{ prefix }} **then accept**;
  {% endfor %}
  }
  reject;
  }

BIRD: CPE via eBGP

```
protocol bgp {{ prefix }} {
    local as {{ rs_asn }};
    neighbor {{ remote_address }} as {{ remote_asn }};
    source address {{ rs_address }};
    multihop;
    igp table myself;
    gateway recursive;
    next hop keep;
    import none;
    export filter golden_cdn;
}
```

BIRD: CPE via eBGP

```
protocol bgp {{ prefix }} {
    local as {{ rs_asn }};
    neighbor {{ remote_address }} {{ remote_asn }};
    source address {{ rs_add }};
    multihop;
    igp table myself;
    gateway recursive;
    next hop keep;
    import none;
    export filter golden_cdn;
}
```

BIRD: CPE via eBGP

```
protocol bgp {{ prefix }} {
    local as {{ rs_asn }};
    neighbor {{ remote_address }} as {{ r      e_asn }};
    source ad      s {{ rs_address }};
    multihop;
    igp table myself;
    gateway recursive;
    next hop keep;
    import none;
    export filter golden_cdn;
}
```

want to be like an eBGP route-server

BIRD: CPE via eBGP

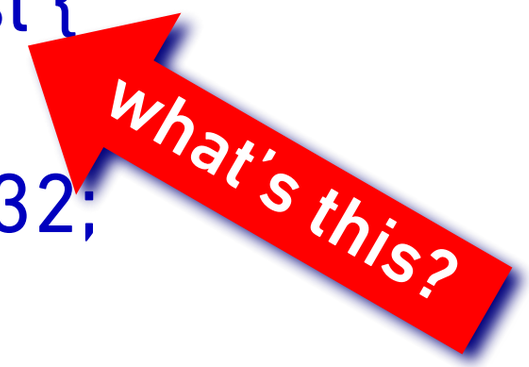
```
protocol bgp {{ prefix }} {
    local as {{ rs_asn }};
    neighbor {{ remote_address }} as {{ remote_asn }};
    source address {{ rs_address }};
    multihop;
    igp table myself;
    gateway recursive;
    next hop keep;
    import none;
    export filter golden_cdn;
}
```

BIRD: CPE via eBGP

```
protocol bgp node200222_cgnv6dsl {
    local as 41495;
    neighbor 46.227.202.104 as 65432;
    source address 46.227.201.12;
    multihop;
    igp table myself;
    gateway recursive;
    next hop keep;
    import none;
    export filter golden_cdn;
}
```

BIRD: CPE via eBGP

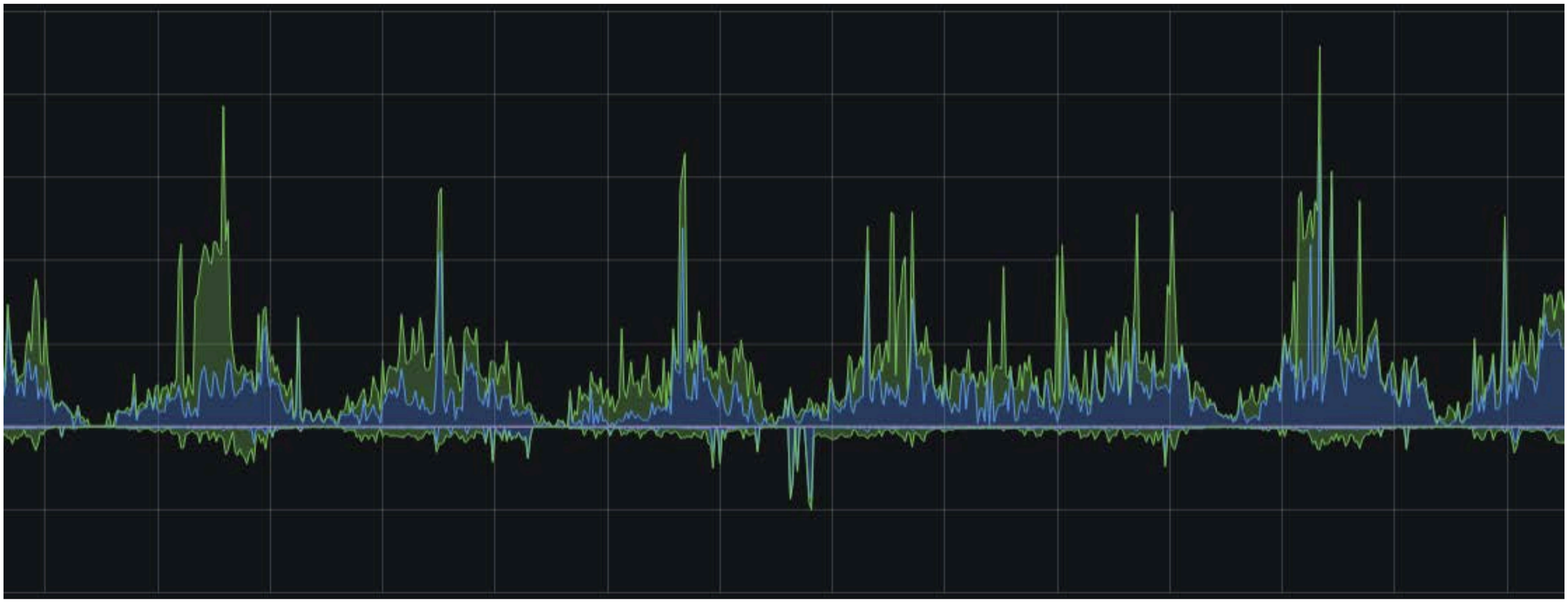
```
protocol bgp node200222_cgnv6dsl {
    local as 41495;
    neighbor 46.227.202.104 as 65432;
    source address 46.227.201.12;
    multihop;
    igp table myself;
    gateway recursive;
    next hop keep;
    import none;
    export filter golden_cdn;
}
```



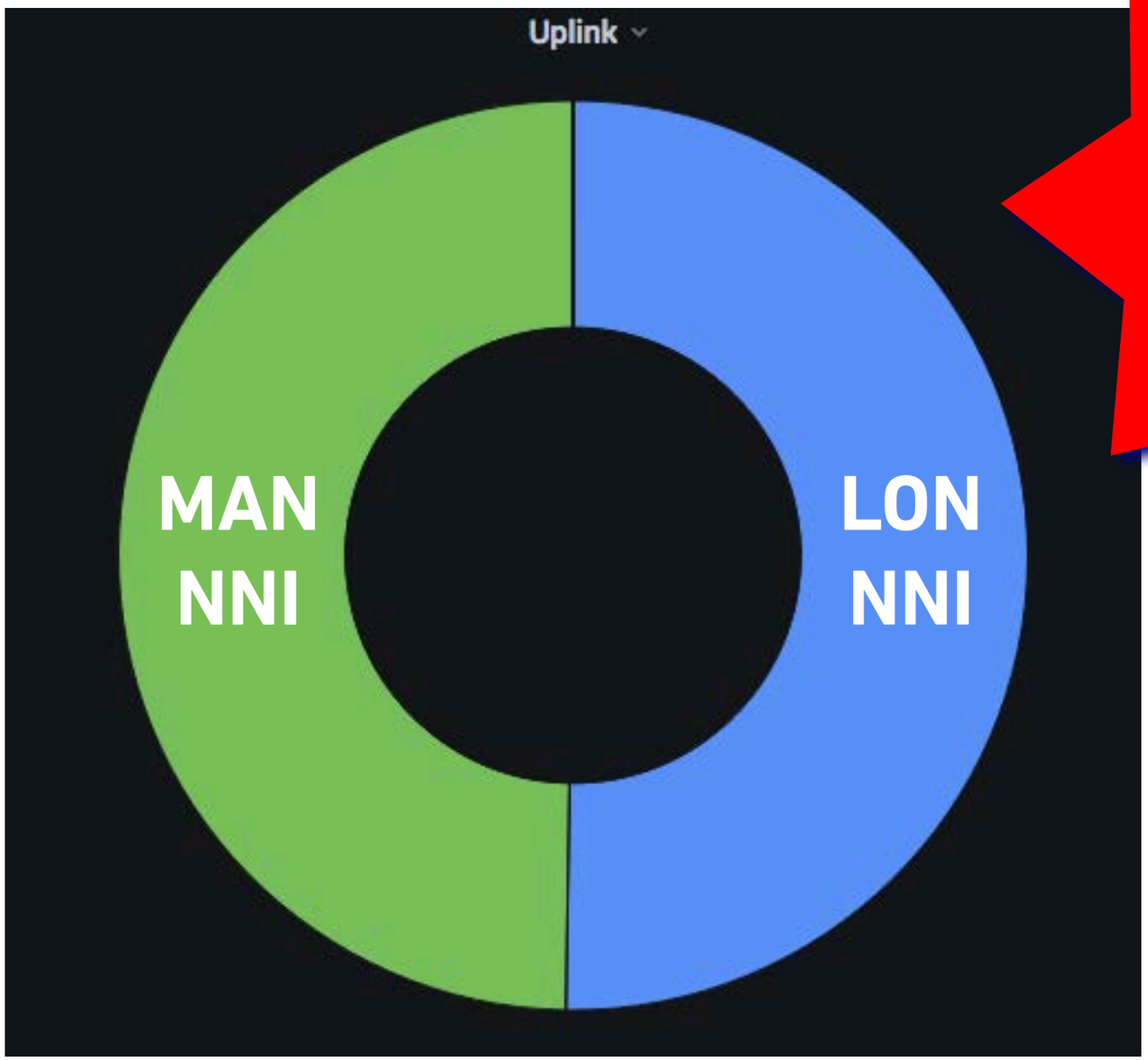
BIRD: CPE via eBGP

- **protocol bgp** node200222_cgnv6dsl {
  local as 41495;
  neighbor 46.227.202.104 **as** 65432;
  source address 46.227.201.12;
  multihop;
  igp table myself;
  gateway recursive;
  next hop keep;
  import none;
  export filter golden_cdn;
}

Show Us The Graphs!



Show Us The Graphs!

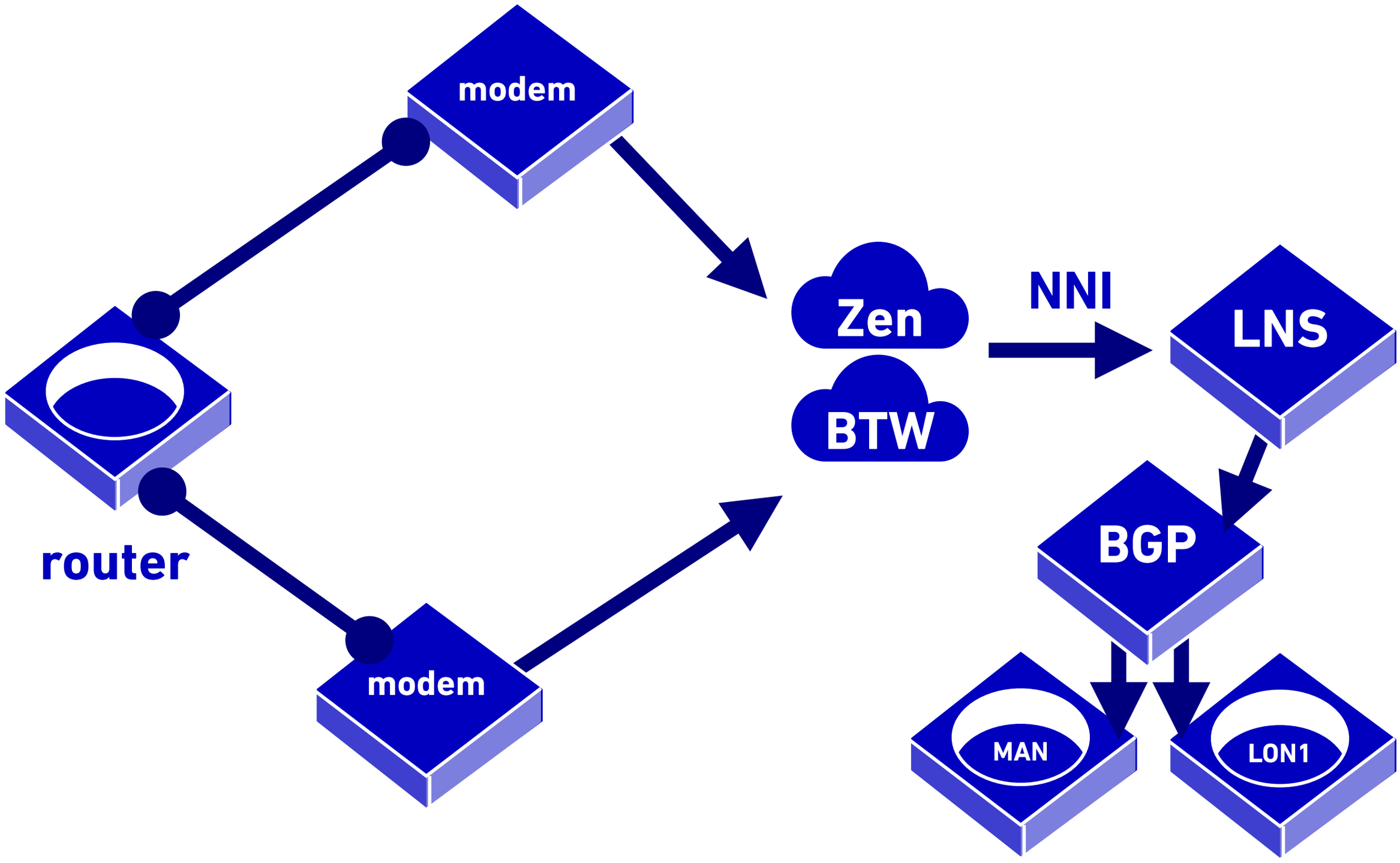


BGP CDN ECMP FTTX

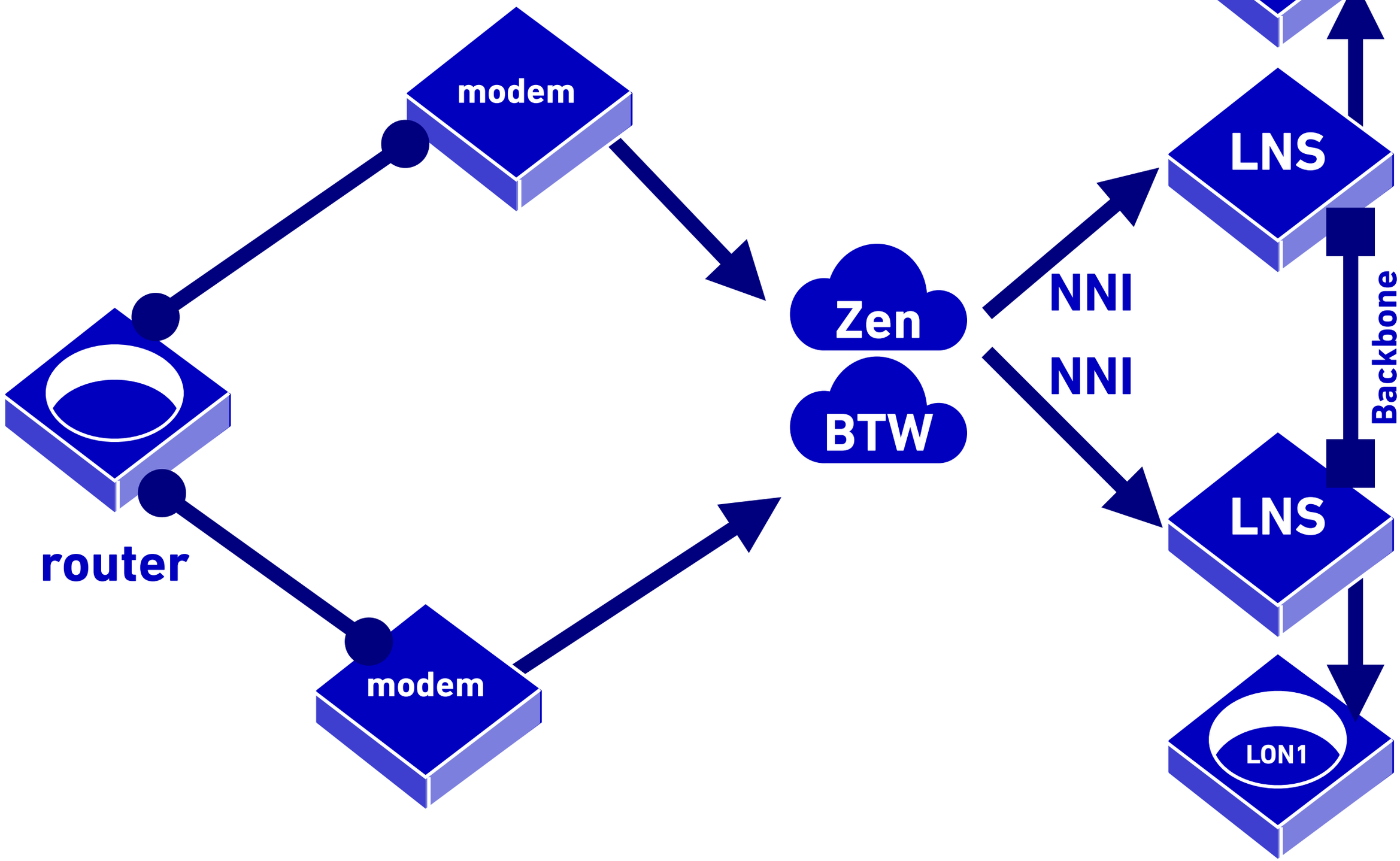
Aggregating XDSL

- We sell IP transit on LLs and FTTX to help alt-nets build out backhaul (either primary or backup).
- FTTP 1000M is great when you can get it, but what if we're limited to GFast or even FTTC?
- MLPPP was in vogue for a while for access providers, but we're more familiar with data-centre technologies than CPE/BNG/LNS inter-op issues.
- Can we find parallels with "Not-So-Shadow VLANs"?

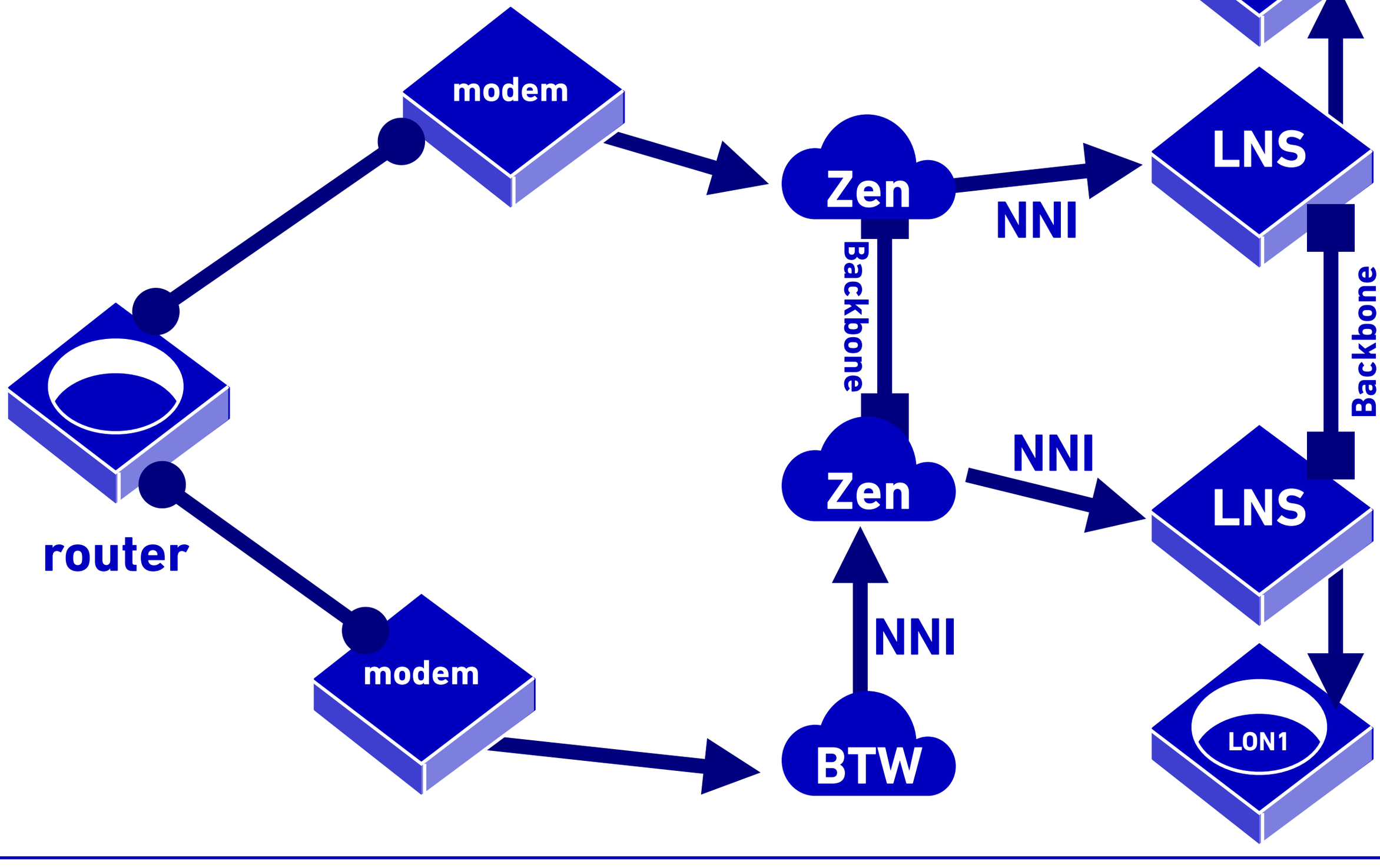
FTTX



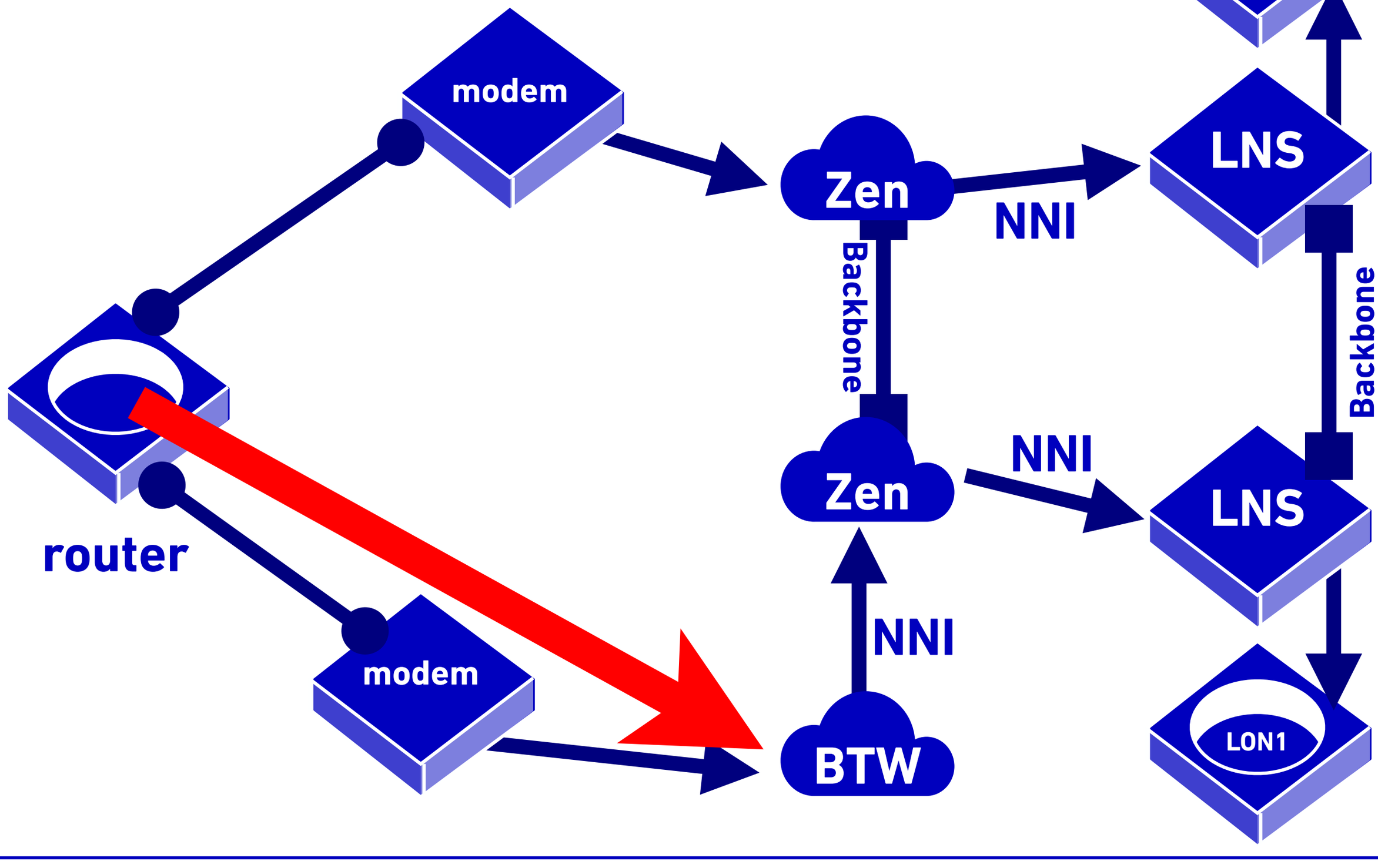
FTTX



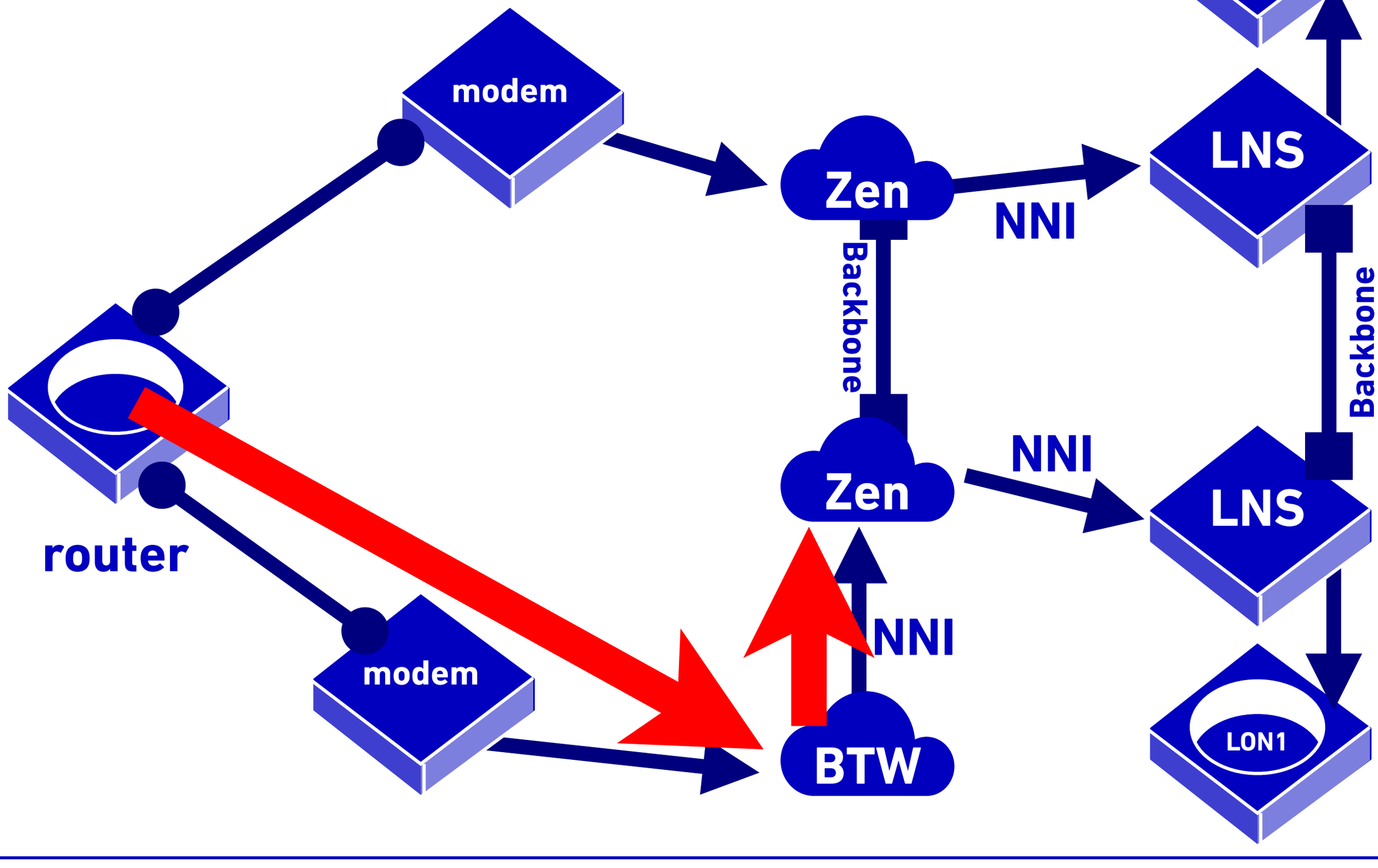
FTTX


modem
router
modem
Zen
BTW
Zen
NNI
NNI
NNI
MAN
LNS
LNS
LON1
Backbone
Backbone

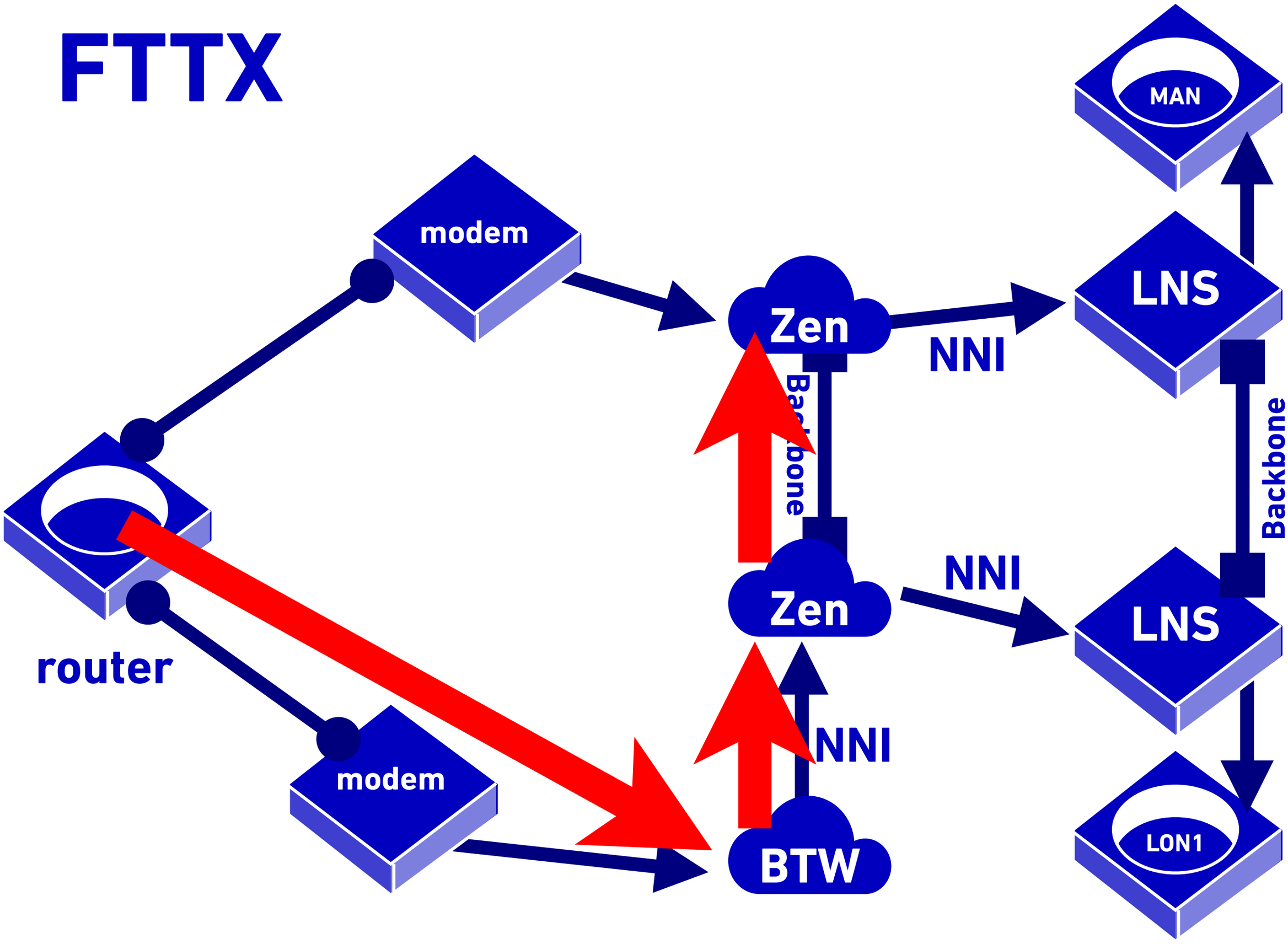
FTTX


MAN
modem
Zen
NNI
LNS
Backbone
Backbone
router
NNI
NNI
modem
BTW
LON1

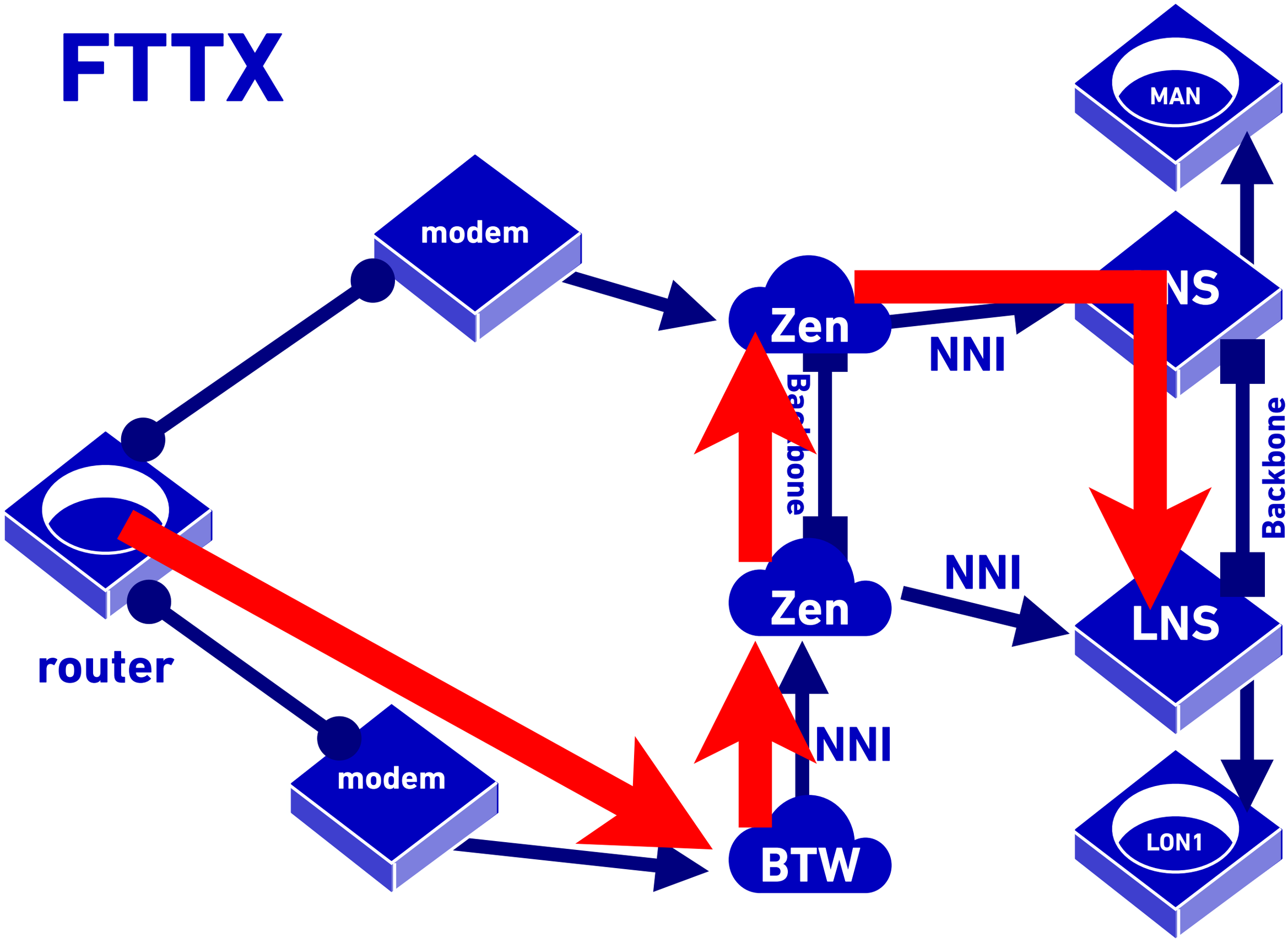
FTTX


modem
router
modem
Zen
BTW
Backbone
NNI
NNI
Zen
NNI
MAN
LNS
Backbone
LNS
LON1

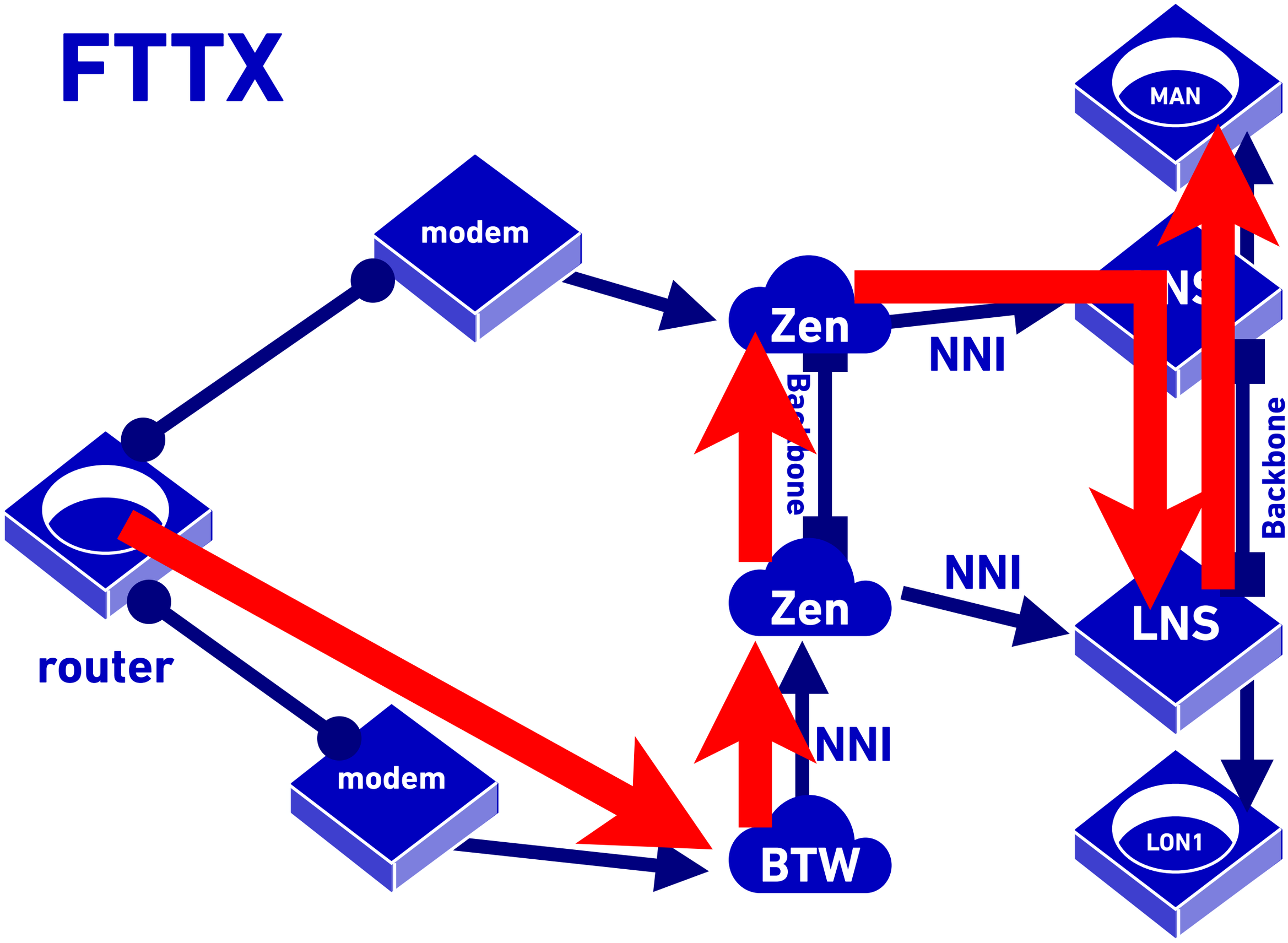
FTTX


MAN
modem
Zen
NNI
LNS
Backbone
router
modem
NNI
Zen
NNI
LNS
BTW
LON1

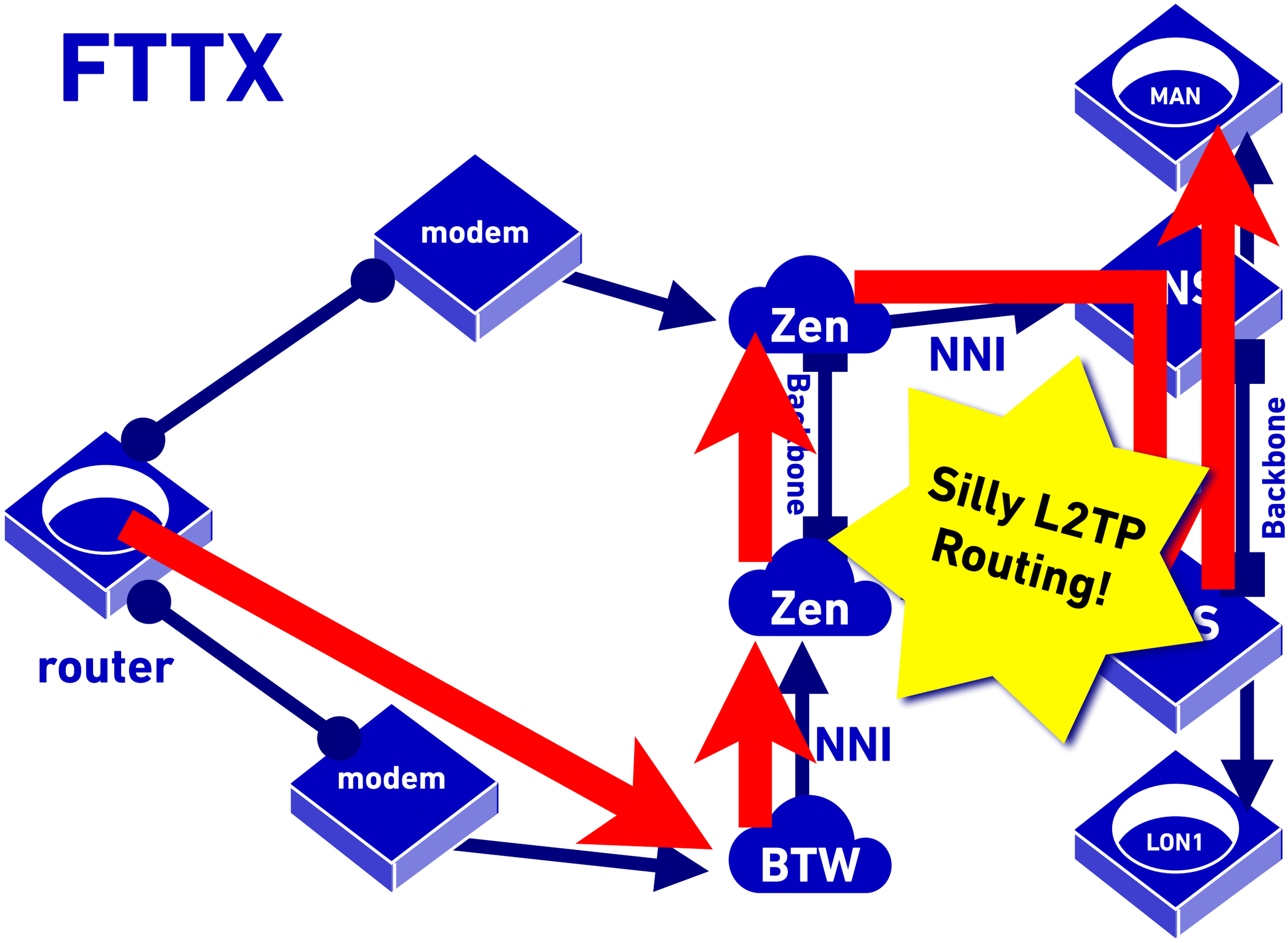
FTTX


MAN
modem
Zen
NNI
NS
Backbone
Backbone
router
modem
NNI
Zen
NNI
LNS
BTW
LON1

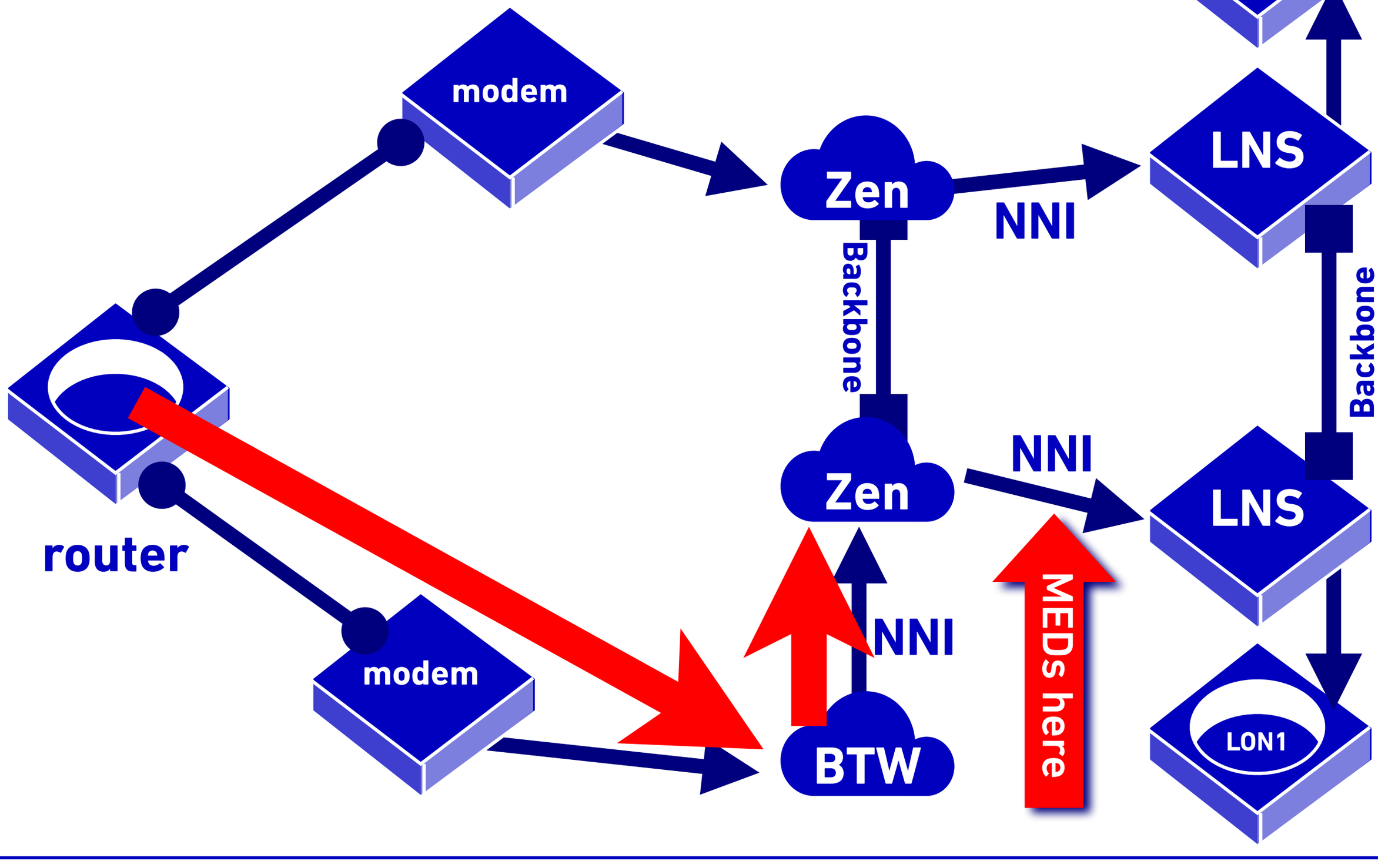
FTTX


modem
router
modem
Zen
BTW
Zen
NNI
NNI
NNI
MAN
LNS
LON1
Backbone
Backbone

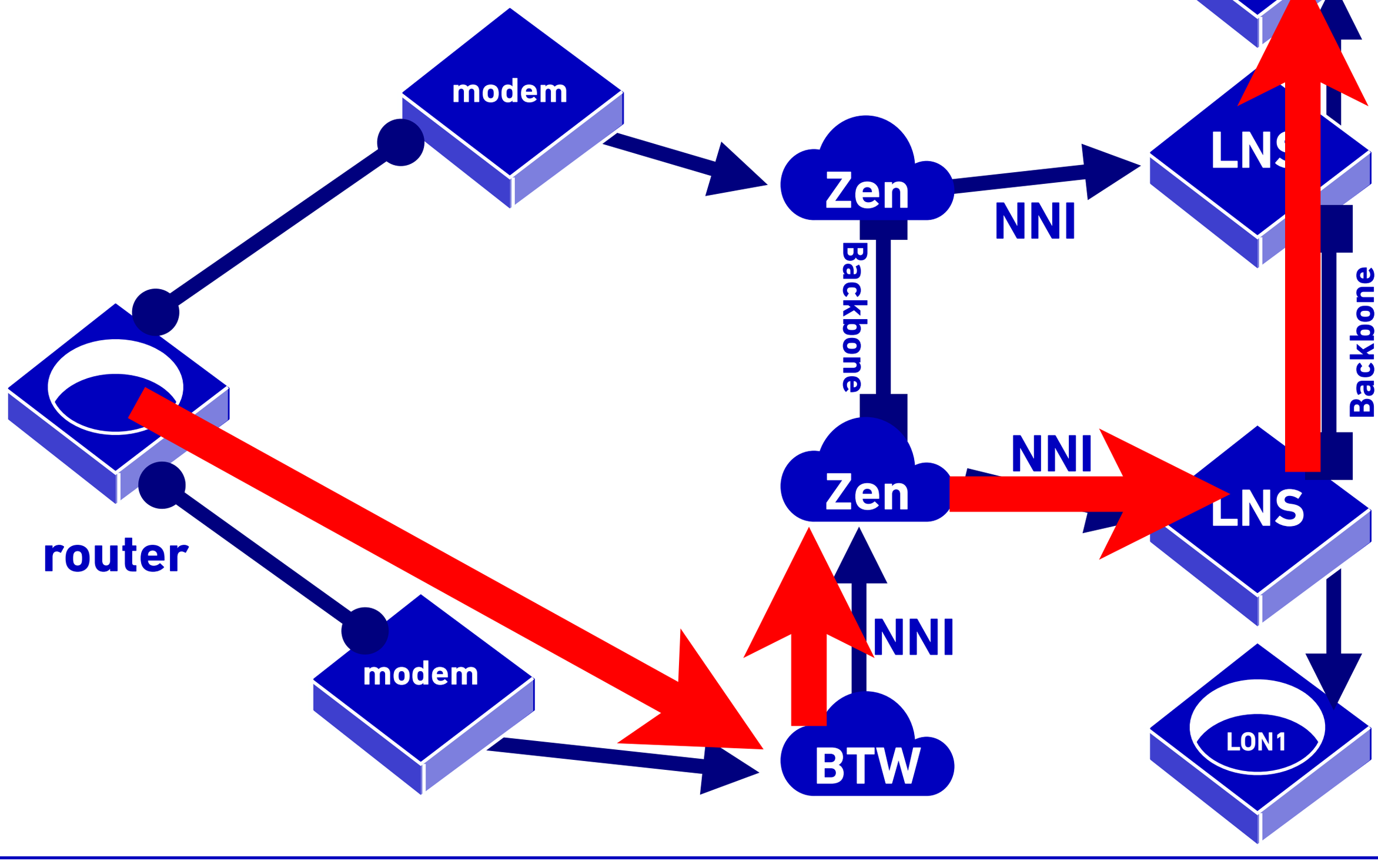
FTTX


modem
router
modem
Zen
NNI
Backbone
Silly L2TP Routing!
MAN
Zen
NNI
BTW
LON1
Backbone

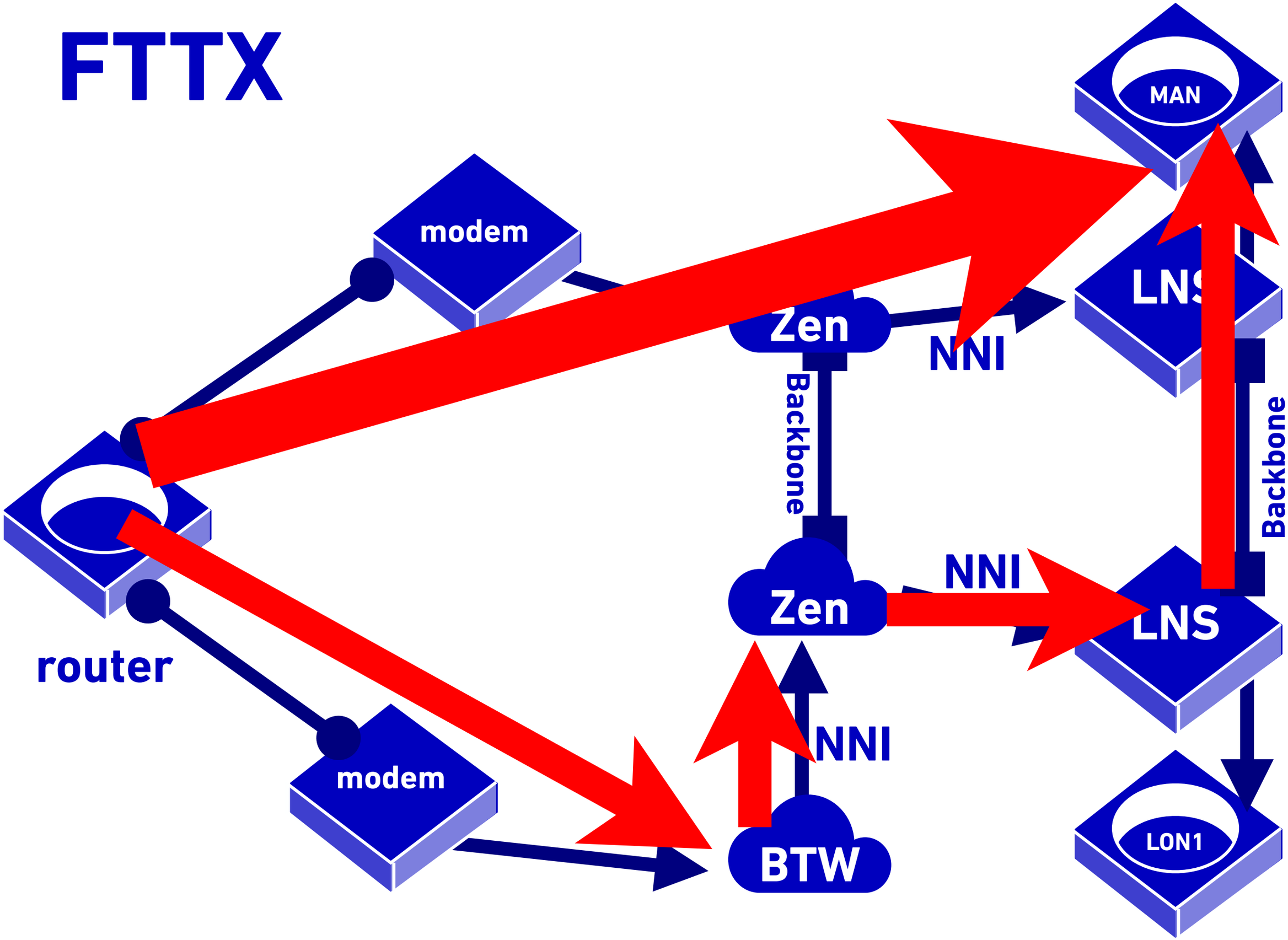
FTTX


MAN
modem
Zen
NNI
LNS
Backbone
Backbone
router
NNI
Zen
NNI
LNS
modem
NNI
MEDs here
BTW
LON1

FTTX


modem
router
modem
Zen
BTW
NNI
NNI
Backbone
Zen
NNI
MAN
LNS
Backbone
LNS
LON1

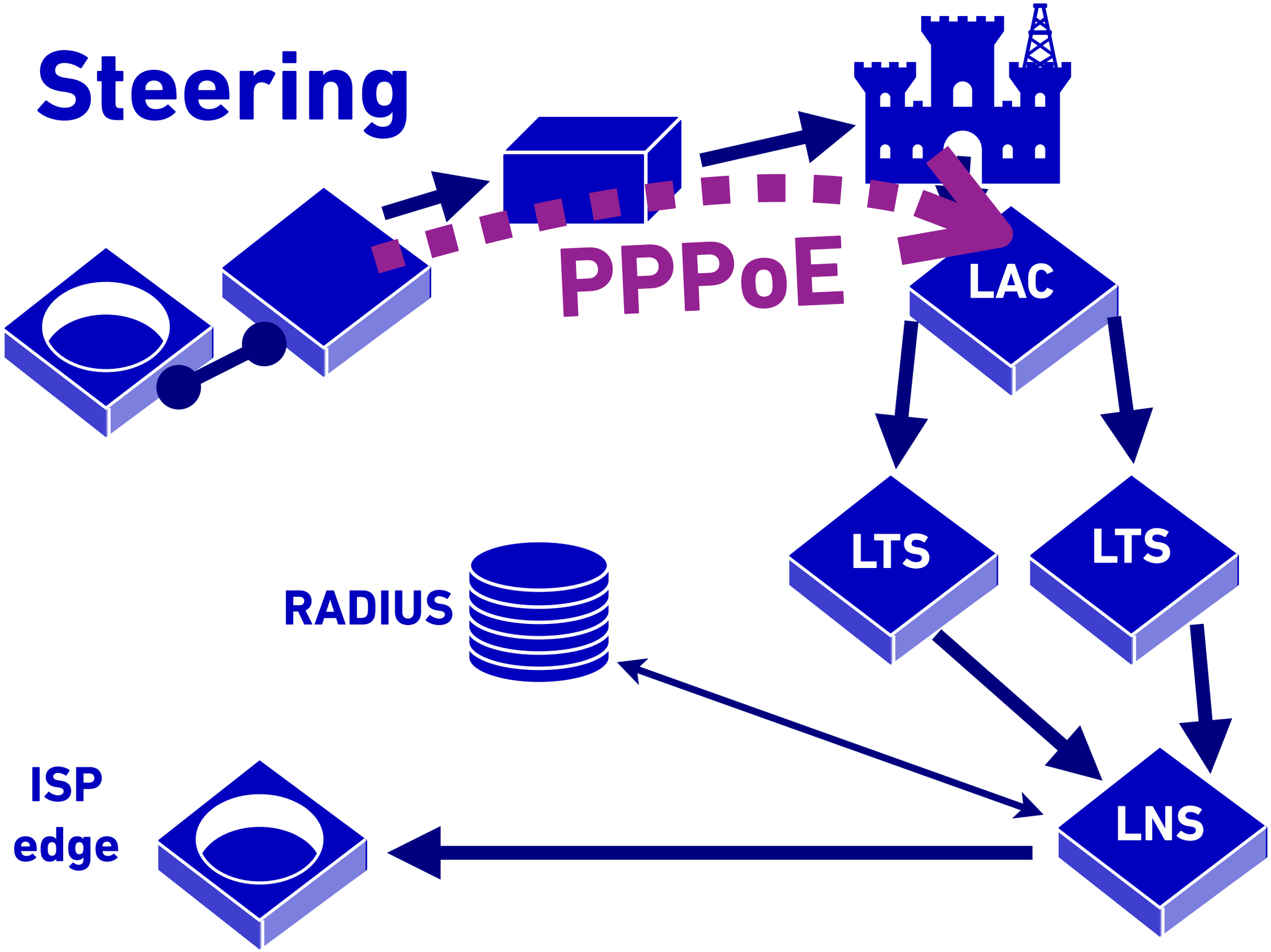
FTTX


MAN
modem
Zen
NNI
Backbone
LNS
router
NNI
Zen
NNI
LNS
modem
BTW
LON1

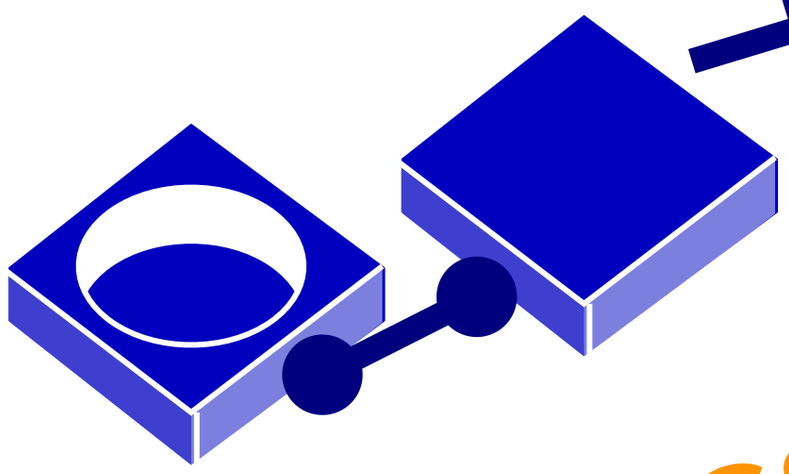
Avoiding Scenic Routing

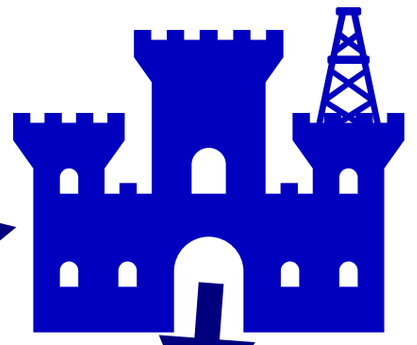
- Much of the UK's broadband infrastructure is aggregated back to London.
- But some LLU operators can break out and hand off circuits elsewhere (Zen, Entanet, others?).
- Can we minimise traffic traversing both our backbone, and our wholesale provider's backbone?

Steering



Steering





LAC

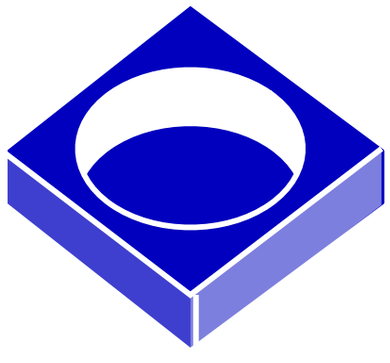
Access-Request

RADIUS

LTS

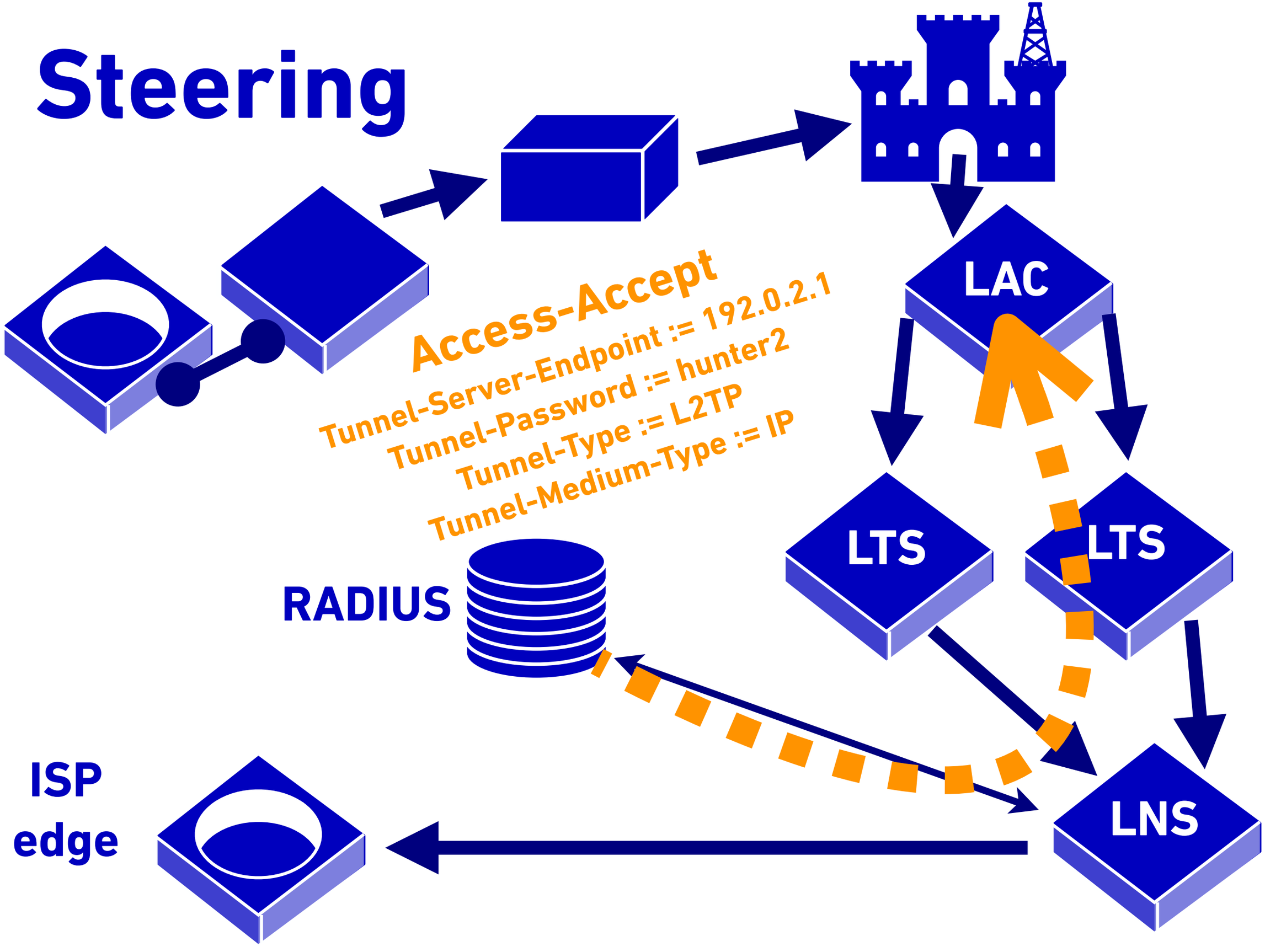
LTS

ISP
edge

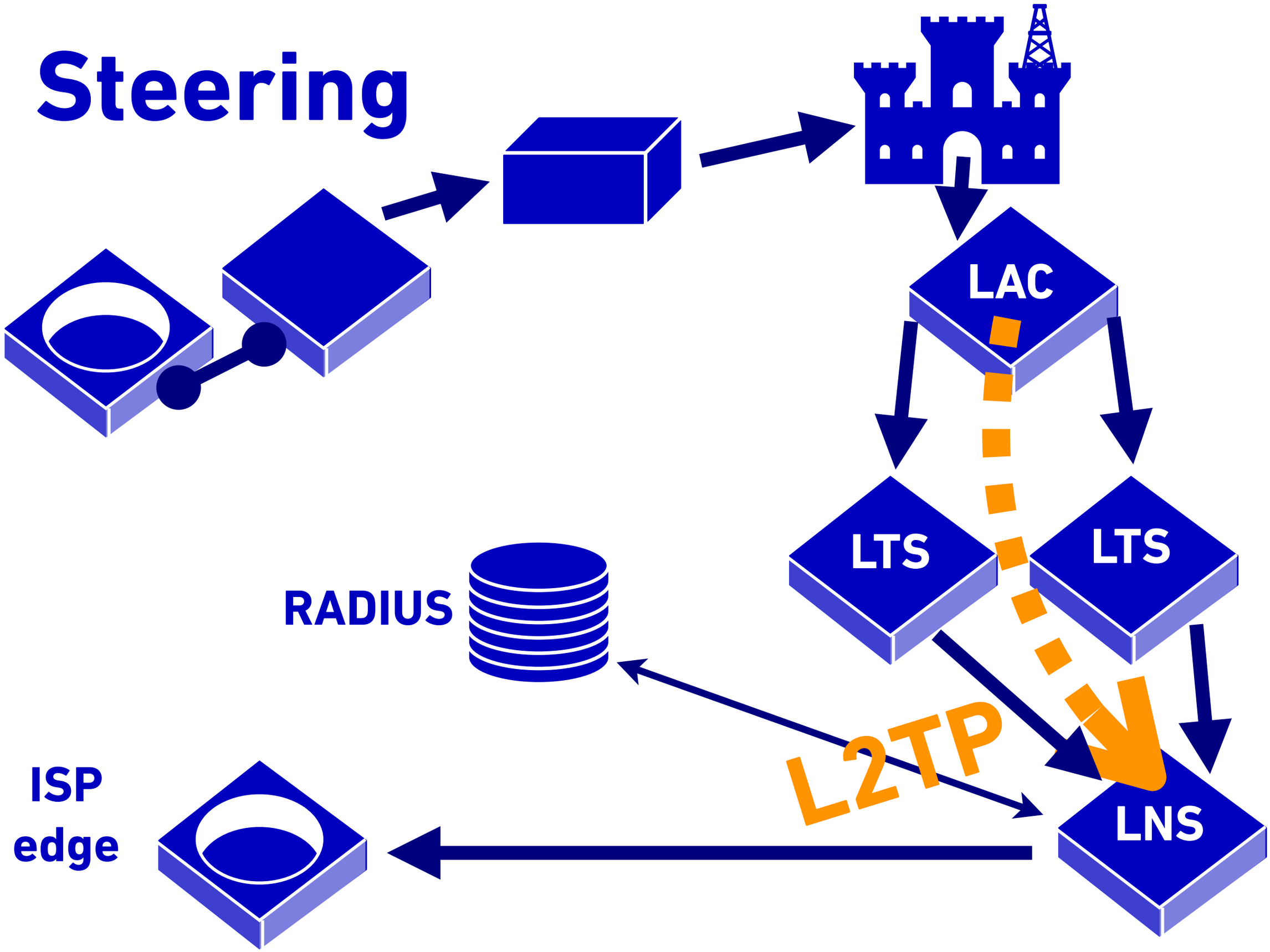


LNS

Steering


Access-Accept
Tunnel-Server-Endpoint := 192.0.2.1
Tunnel-Password := hunter2
Tunnel-Type := L2TP
Tunnel-Medium-Type := IP
LAC
LTS
LTS
RADIUS
LNS
ISP
edge

Steering



Steering

Access-Request

LAC

LTS

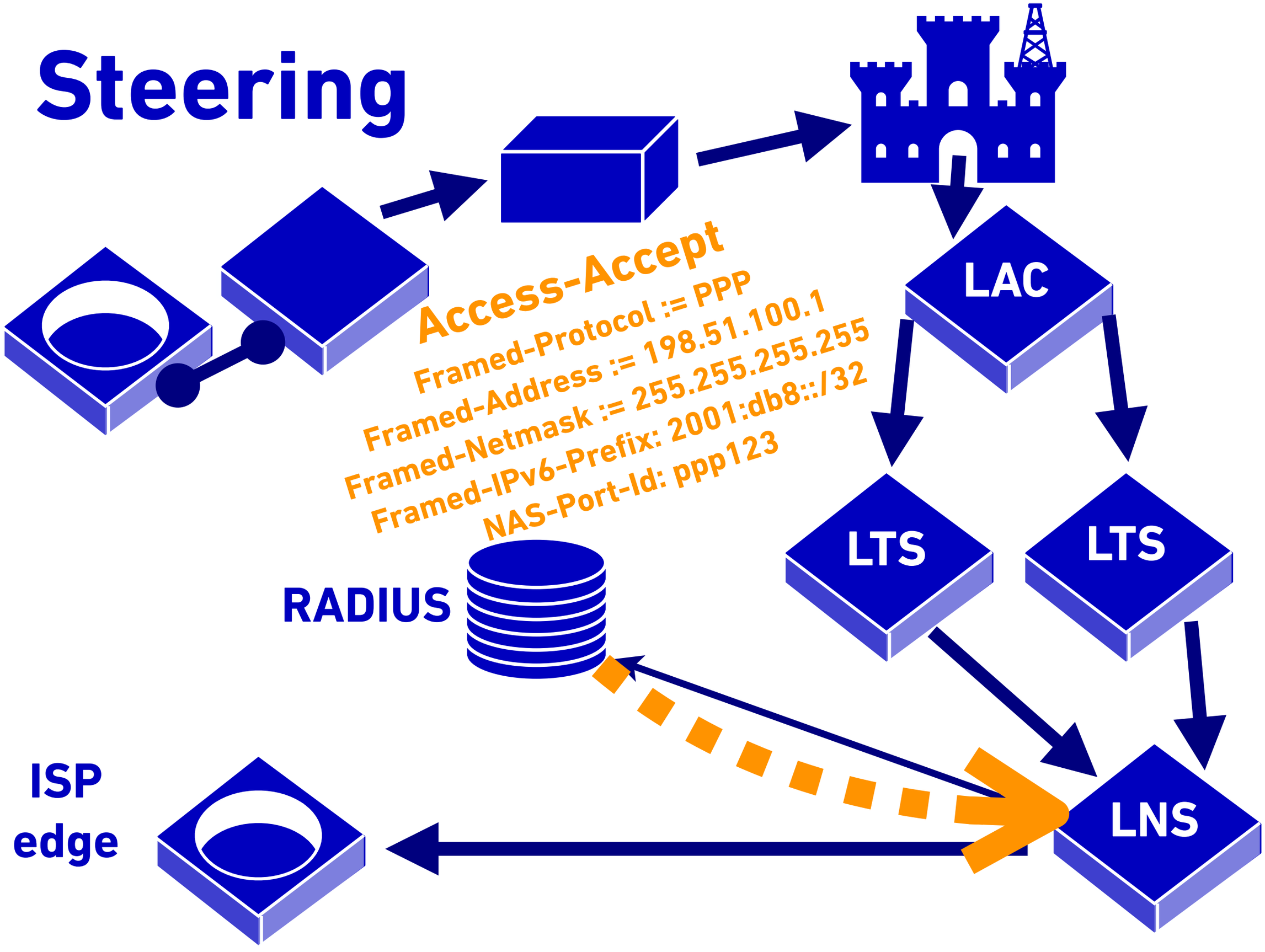
LTS

RADIUS

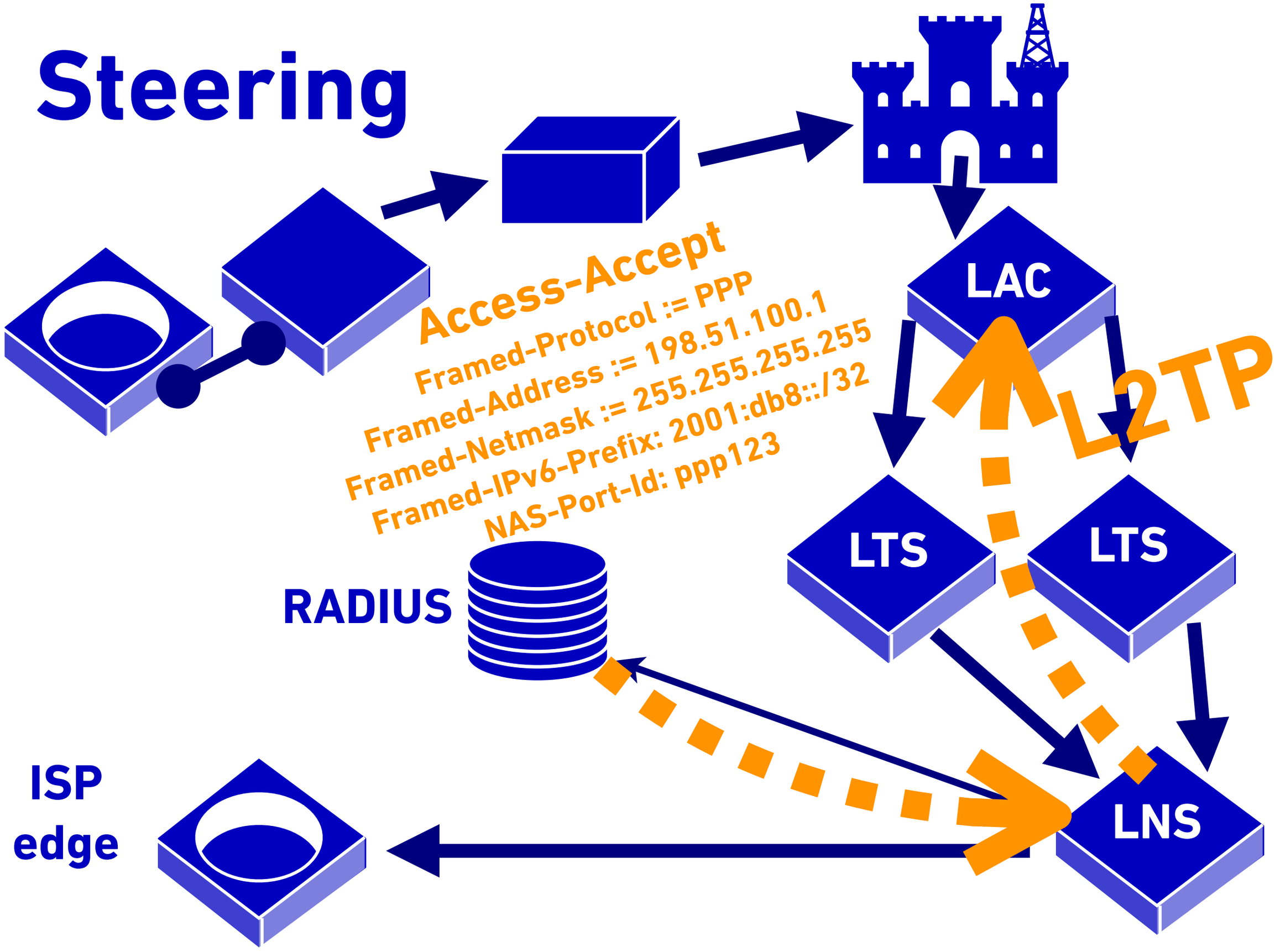
ISP
edge

LNS

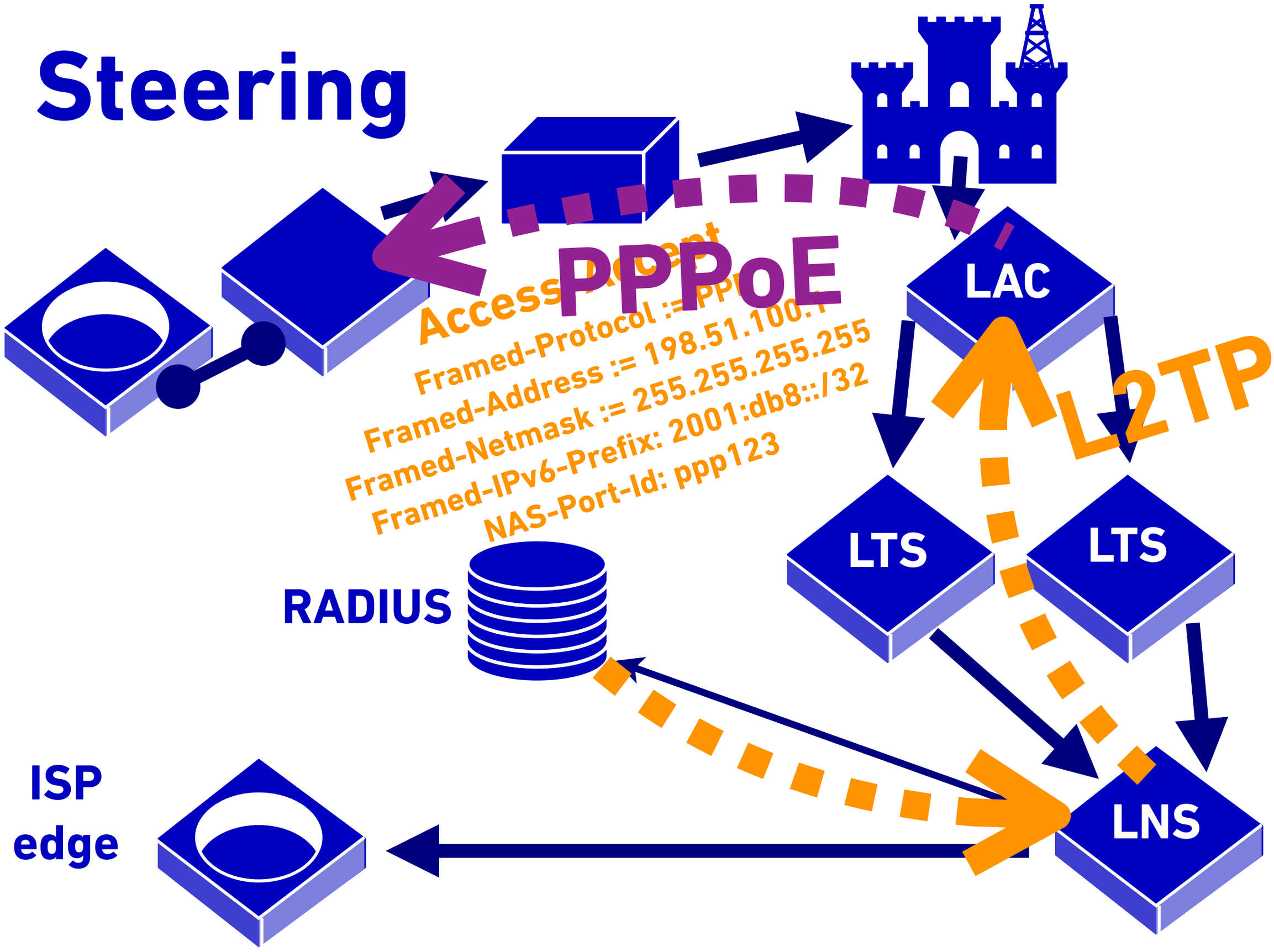
Steering


Access-Accept
Framed-Protocol := PPP
Framed-Address := 198.51.100.1
Framed-Netmask := 255.255.255.255
Framed-IPv6-Prefix: 2001:db8::/32
NAS-Port-Id: ppp123
LAC
LTS
LTS
RADIUS
ISP
edge
LNS

Steering


Access-Accept
Framed-Protocol := PPP
Framed-Address := 198.51.100.1
Framed-Netmask := 255.255.255.255
Framed-IPv6-Prefix: 2001:db8::/32
NAS-Port-Id: ppp123
LAC
L2TP
LTS
LTS
RADIUS
ISP
edge
LNS

Steering


PPPoE
Access-Accept
Framed-Protocol := PPP
Framed-Address := 198.51.100.1
Framed-Netmask := 255.255.255.255
Framed-IPv6-Prefix: 2001:db8::/32
NAS-Port-Id: ppp123
LAC
L2TP
LTS
LTS
RADIUS
ISP
edge
LNS

Steering

PPP

LAC

LTS

LTS

LNS

RADIUS

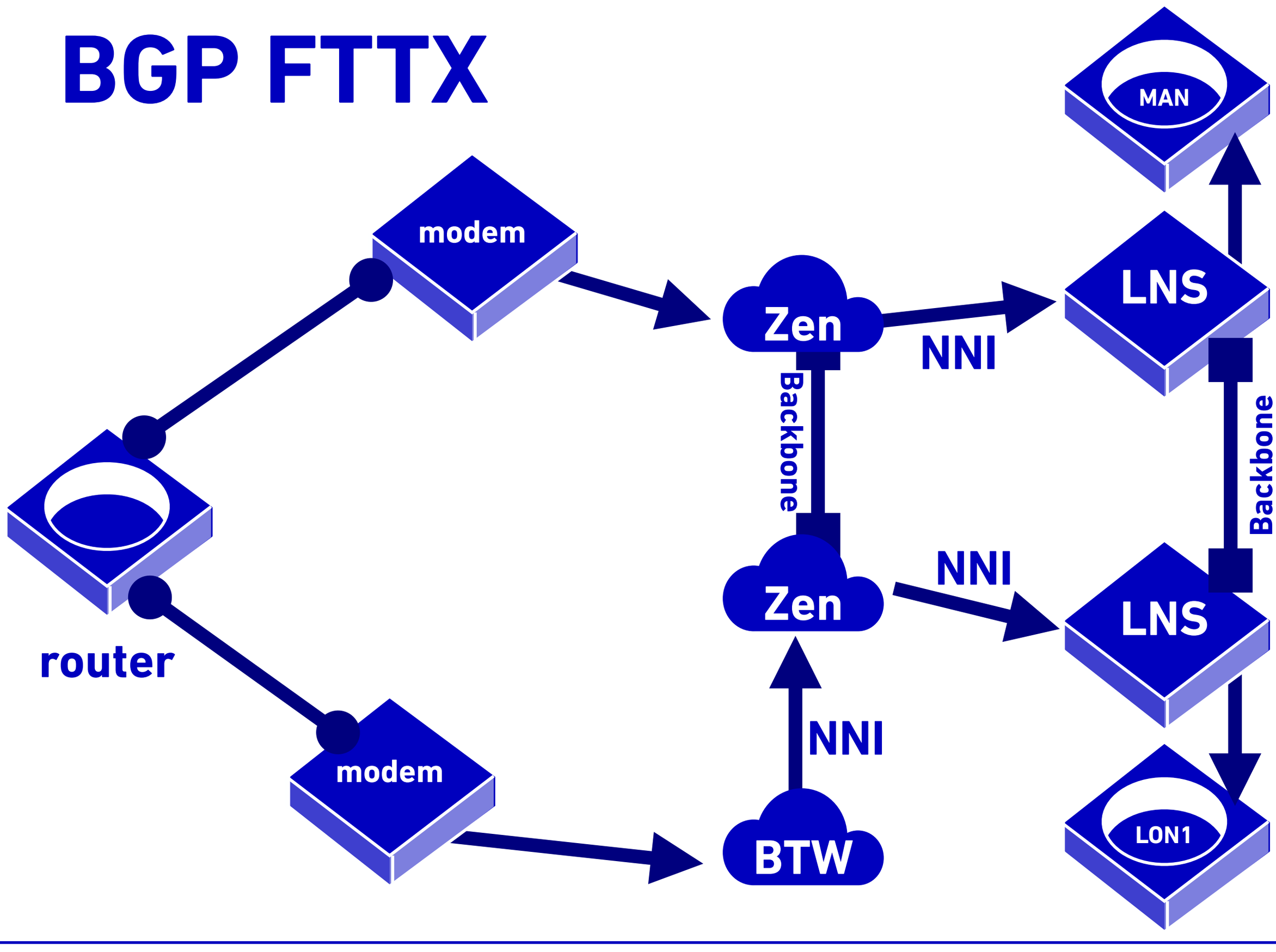
ISP
edge

CPE-Directed Steering

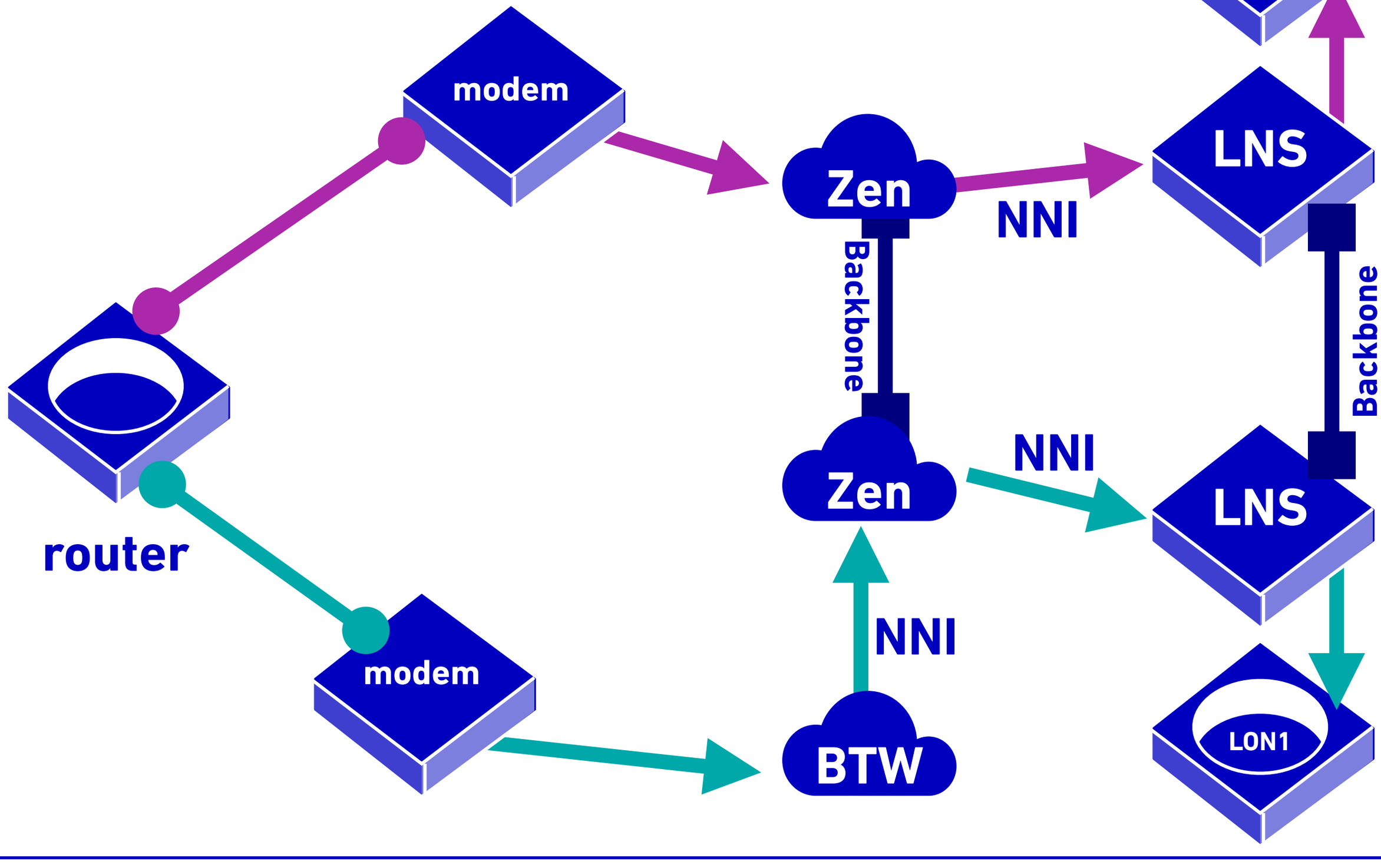
- Order circuits on both BTW and LLU networks.
- Choose which region's LNSs they establish to.

- We added a feature to our RADIUS+L2TP setup: **user+steer@realm**
  - Rewritten to **user@steer.realm** for L2TP steering.
  - But RADIUS behaves likes **user@realm** for LNS.

BGP FTTX

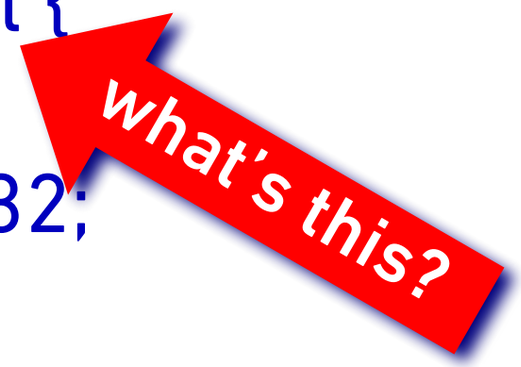

MAN
modem
LNS
Zen
NNI
Backbone
Backbone
router
NNI
Zen
NNI
LNS
modem
BTW
LON1

BGP FTTX

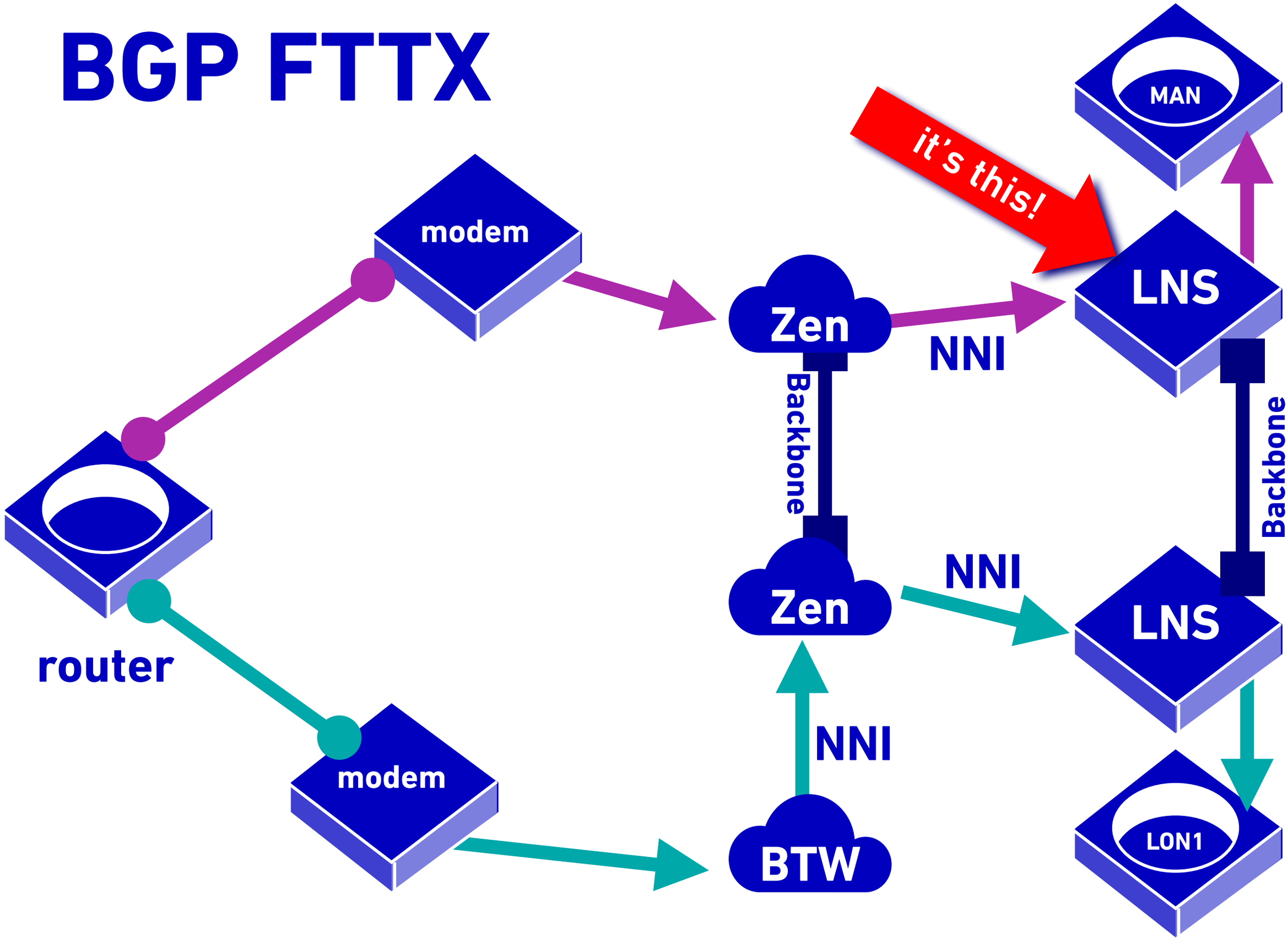

modem
MAN
Zen
NNI
LNS
Backbone
router
modem
Zen
NNI
LNS
NNI
Backbone
BTW
LON1

BIRD: CPE via eBGP

- **protocol bgp** node200222_cgnv6dsl {
  local as 41495;
  neighbor 46.227.202.104 **as** 65432;
  source address 46.227.201.12;
  multihop;
  igp table myself;
  gateway recursive;
  next hop keep;
  import none;
  export filter golden_cdn;
}



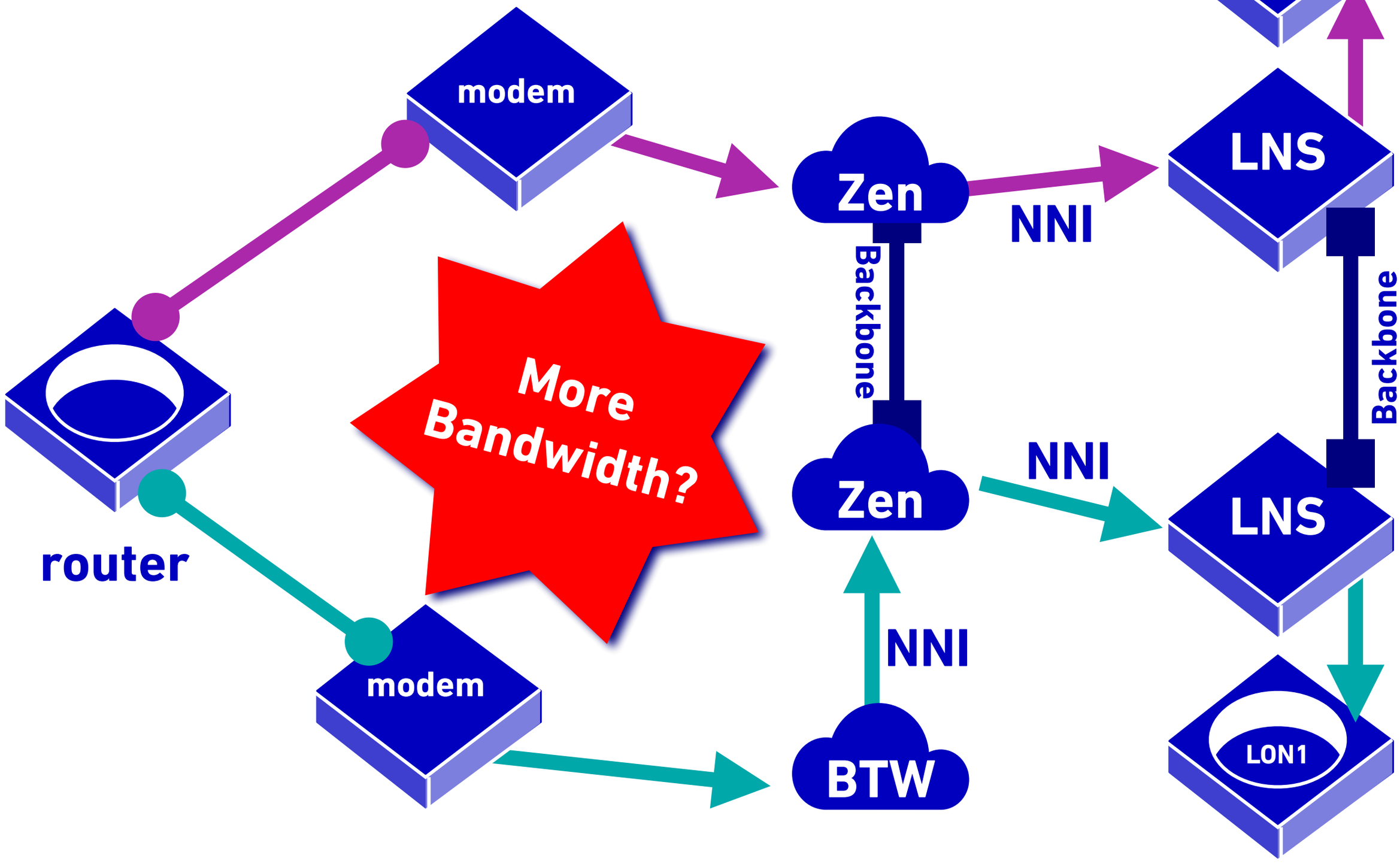
BGP FTTX


MAN
it's this!
modem
LNS
Zen
NNI
Backbone
Backbone
router
NNI
Zen
LNS
modem
NNI
BTW
LON1

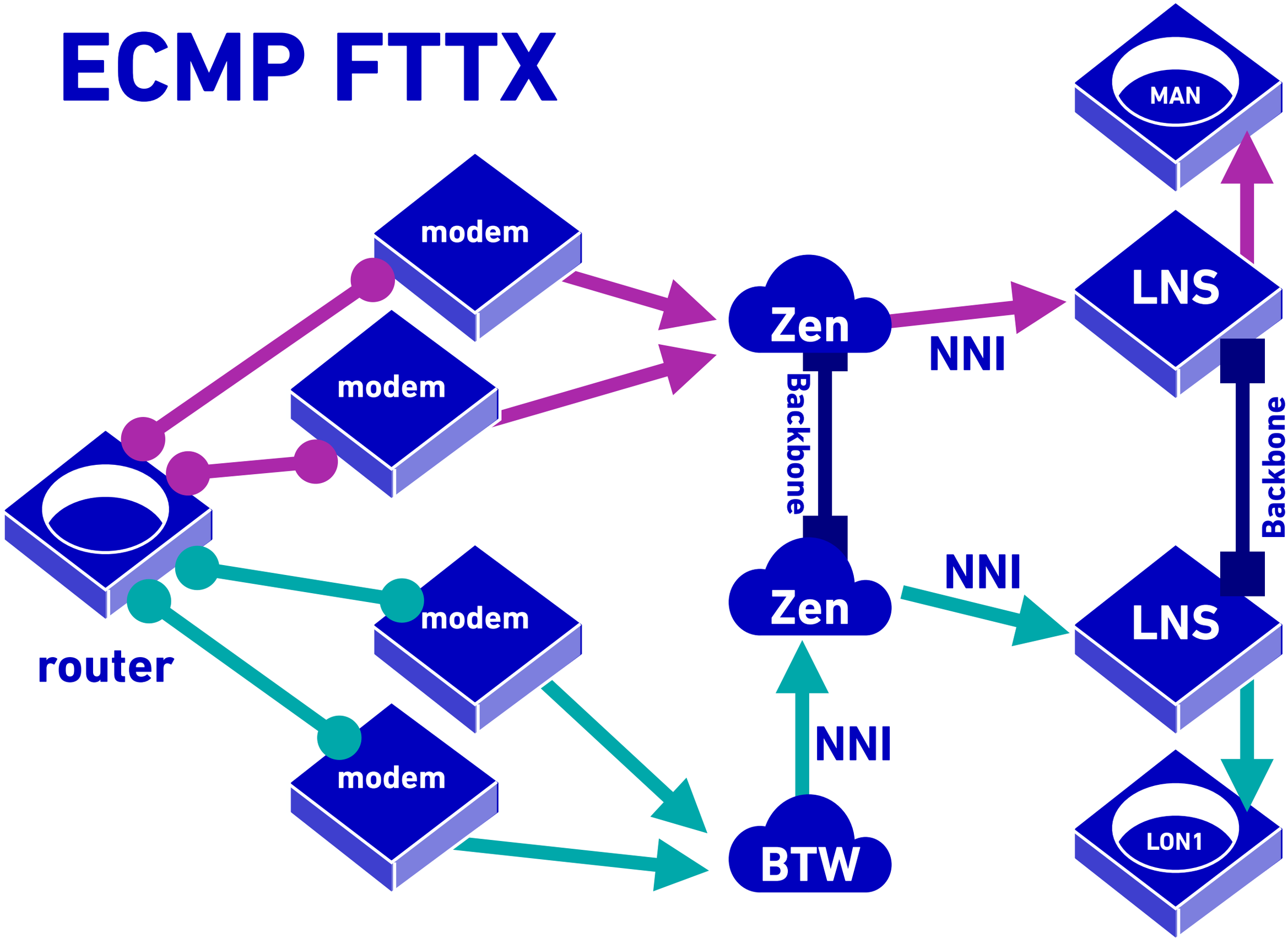
BIRD: CPE via eBGP

```
protocol bgp node200222_cgnv6dsl {
    local as 41495;
    neighbor 46.227.202.104 as 65432;
    source address 46.227.201.12;
    multihop;
    igp table myself;
    gateway recursive;
    next hop keep;
    import none;
    export filter golden_cdn;
}
```

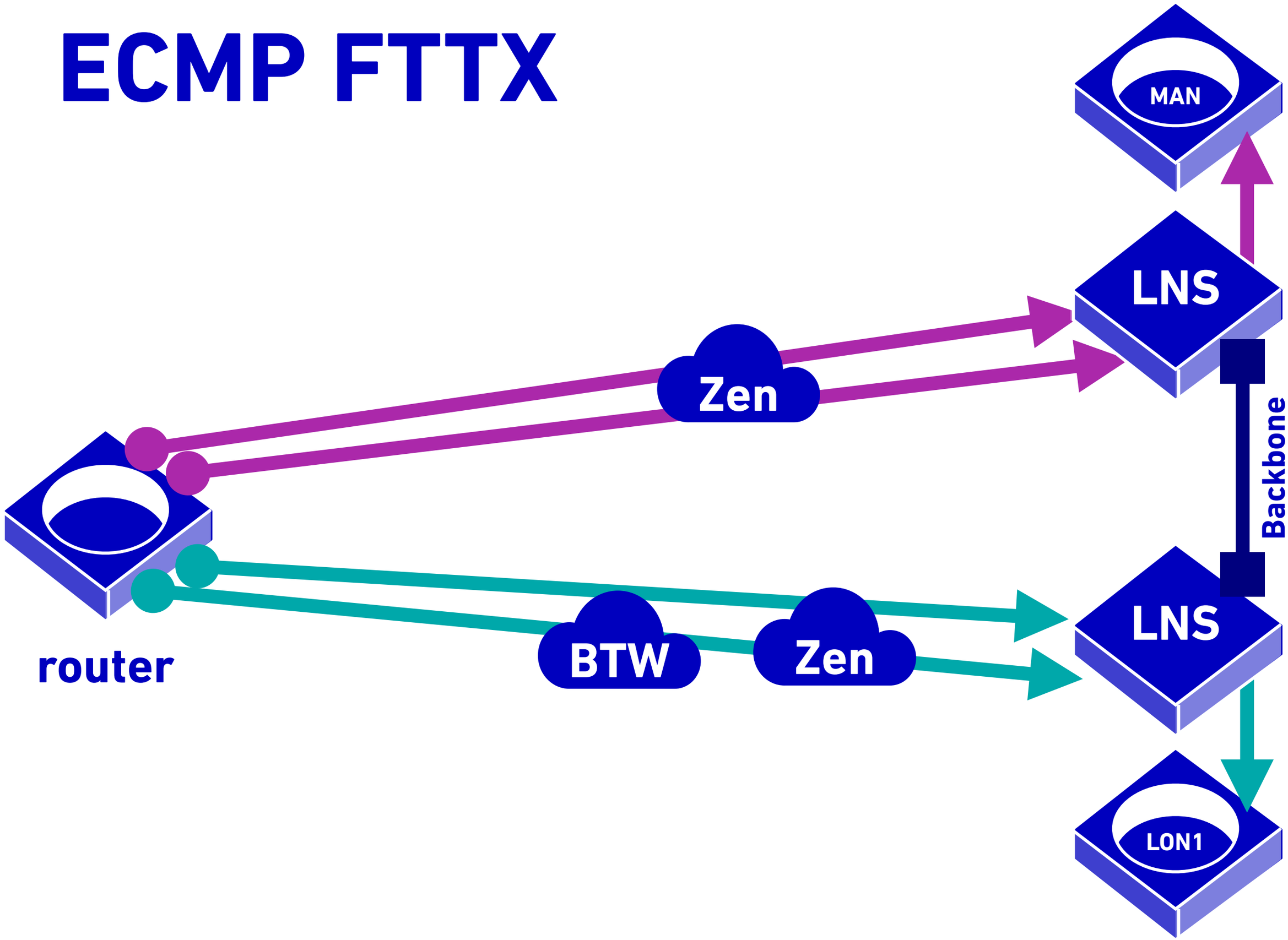
BGP FTTX


MAN
modem
Zen
NNI
LNS
Backbone
Backbone
More Bandwidth?
router
modem
Zen
NNI
NNI
LNS
BTW
LON1

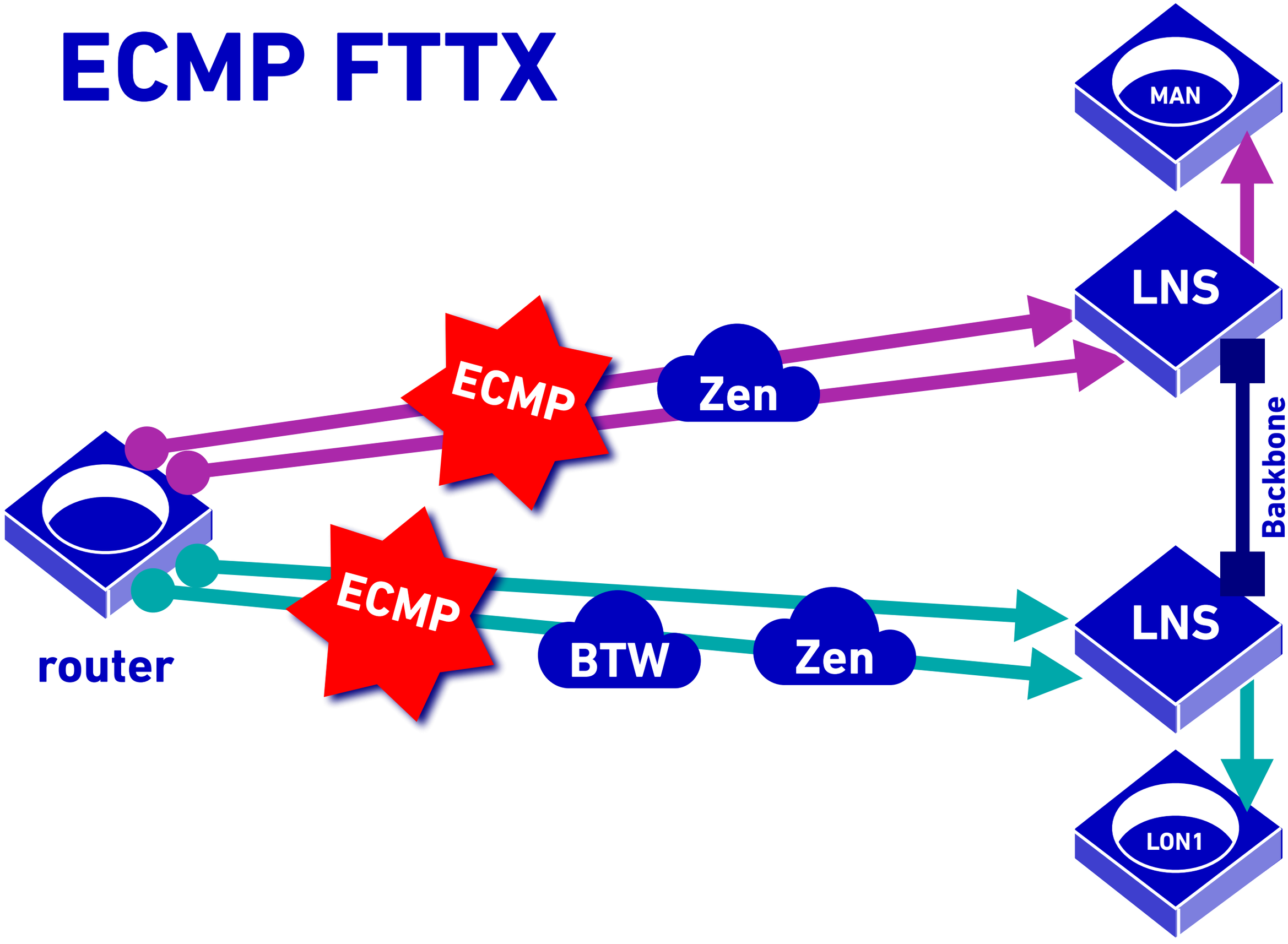
ECMP FTTX


MAN
modem
modem
Zen
NNI
LNS
Backbone
Backbone
router
modem
modem
NNI
NNI
Zen
LNS
BTW
LON1

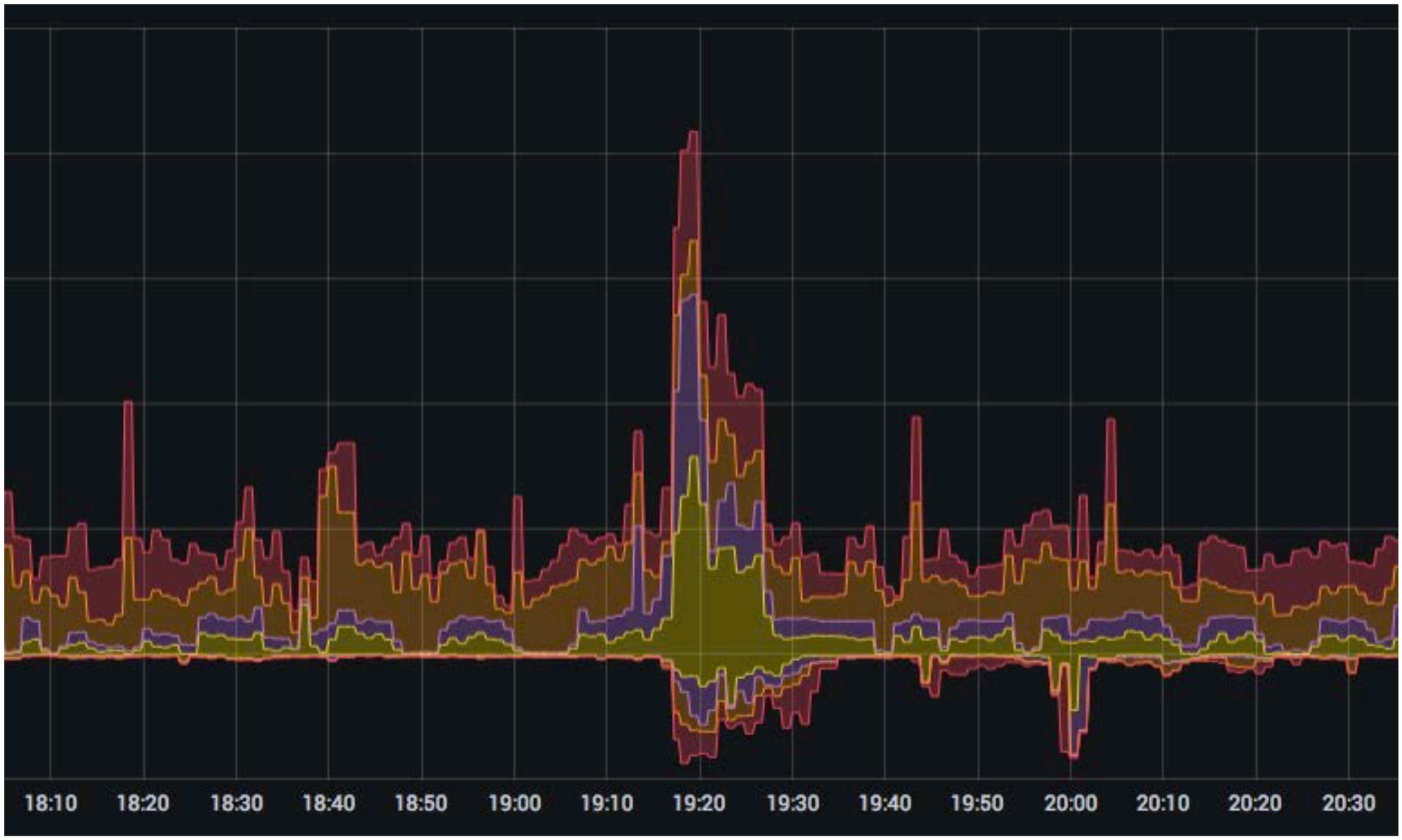
ECMP FTTX



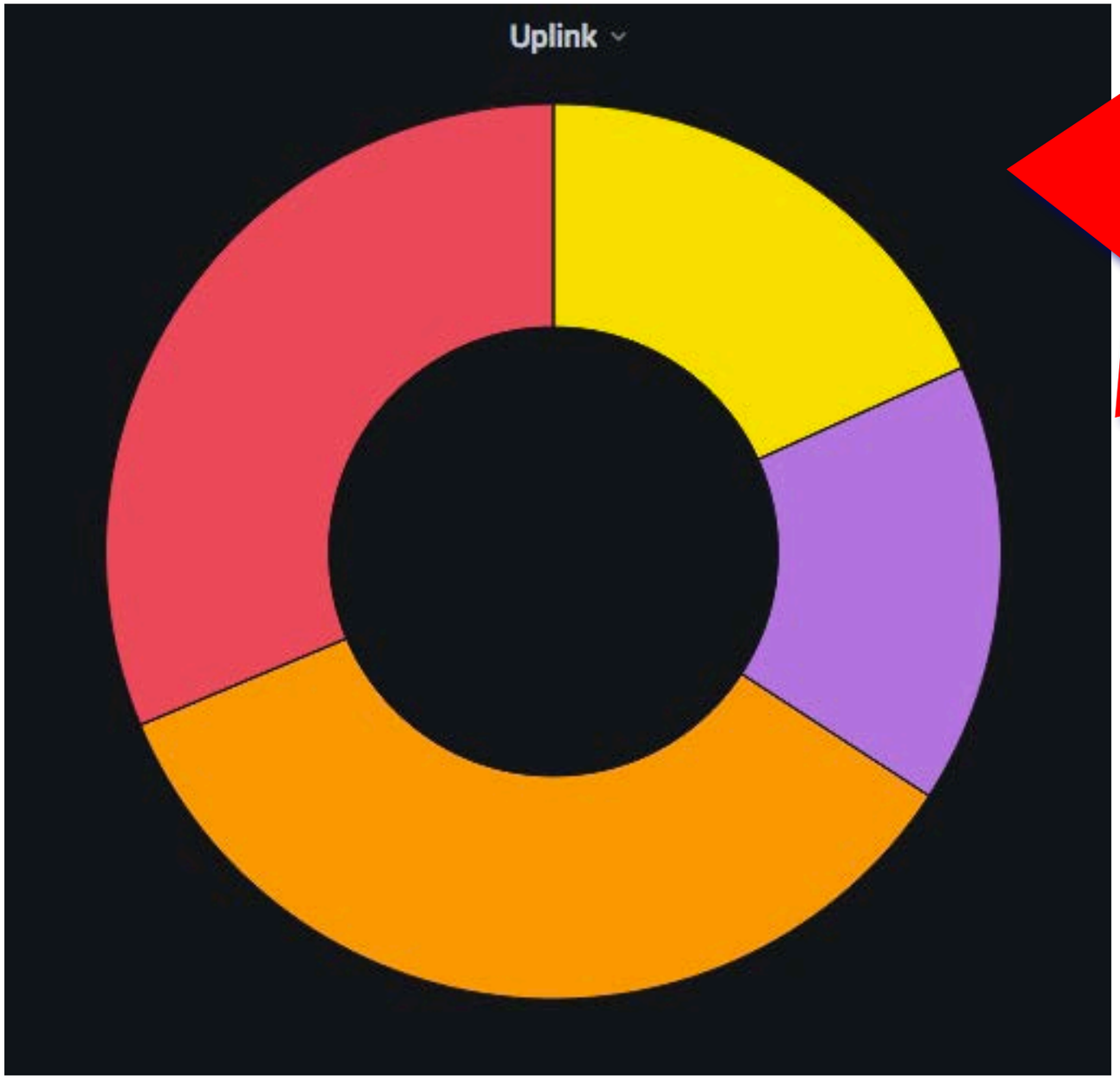
ECMP FTTX



ECMP BGP FTTX In Action

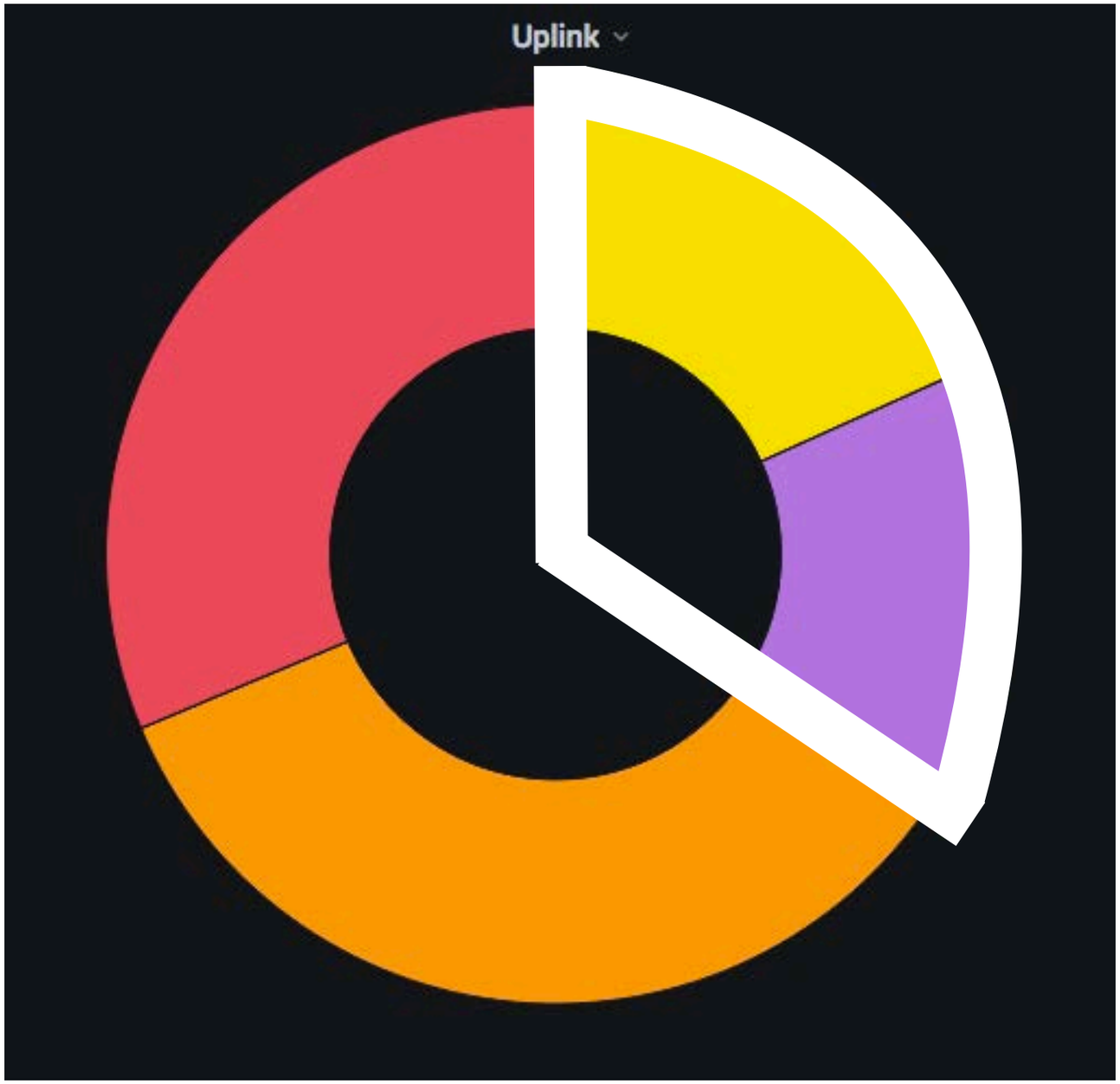


ECMP BGP FTTX In Action

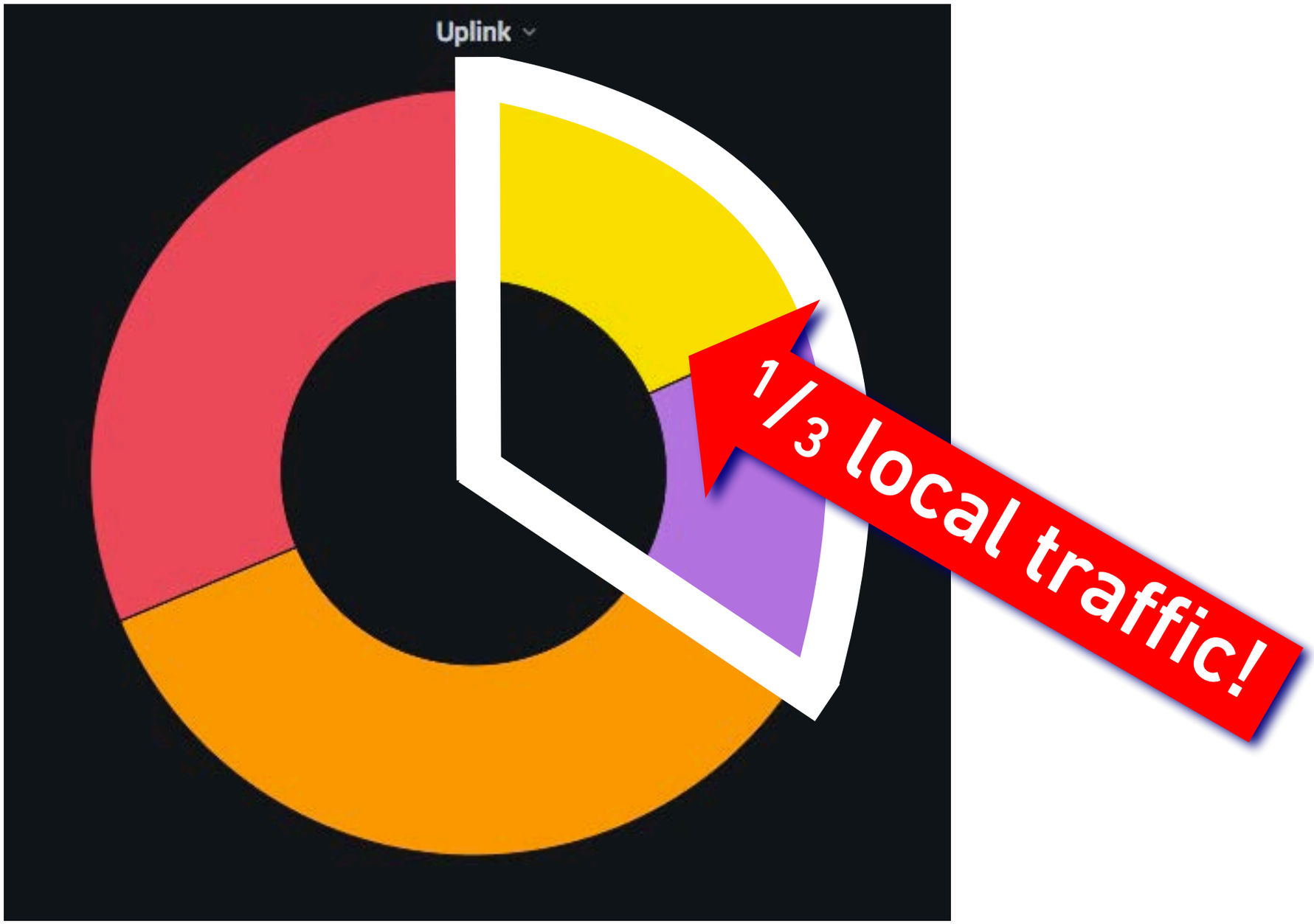


7 Days of Data

ECMP BGP FTTX In Action



ECMP BGP FTTX In Action



CONCLUSIONS

Summary

- Analysing how DNS resolution affects which POPs a CDN sends you traffic is the beginning.
- Using multiple physical or logical connections to a customer site can add resilience, but also can be beneficial for traffic routing.
- Applicable to both ethernet and broadband services.

Why Bother?

- Keep traffic local:
  - Lower latency, fewer hops.
  - Less traffic tromboning backbones.
- Use local peering:
  - Regional IX tend to have lower port costs.
  - Spreads load for you and for the CDNs.
- Improve resilience:
  - Survive single-POP and single-IX problems.

ASK ME ABOUT BGP

E: marek @ faelix . net
T: @maznu
T: @faelix
W: https://faelix.net/

https://faelix.link/linxman2021