PATTERNS IN NETWORK ARCHITECTURE:

MULTIHOMING AND MULTICAST

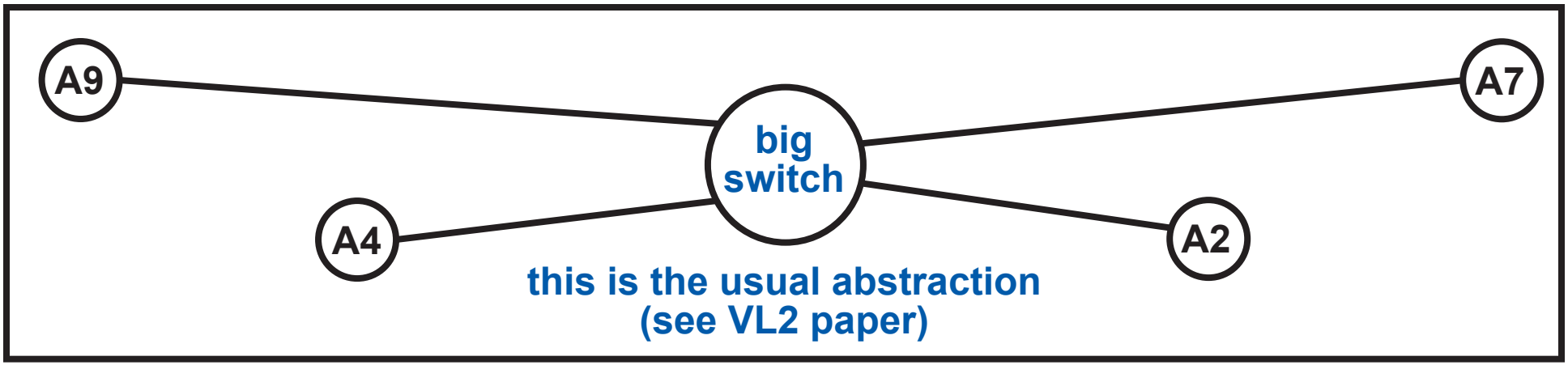
MULTIHOMING AND MULTICAST

OUTLINE

1. A short cloud topic
2. Modeling in Alloy
3. Patterns for multihoming
4. Discussion of “How hard can it be? Designing and implementing a deployable multipath TCP”
5. Discussion of “Designing distributed systems using approximate synchrony in data center networks”

TWO DIFFERENT ABSTRACTIONS

EXAMPLE: VL2



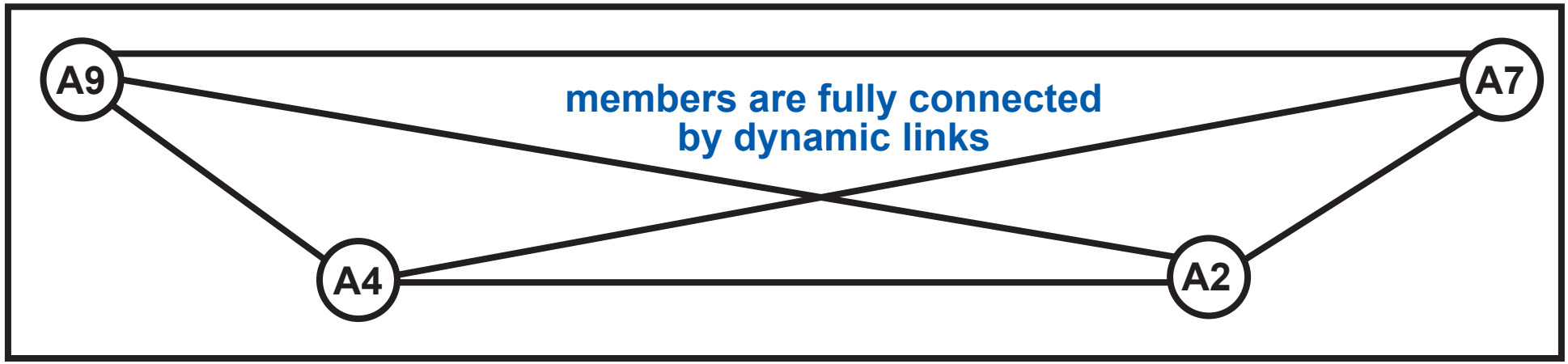
PROBLEM: it's not an abstraction,
it's a fiction

it would be necessary to prove an
implementation correct by bisimulation

THIS IS THE ABSTRACTION
WE ARE USING INSTEAD

all we have to do is show
how each link is implemented,
which is usually straightforward

dynamic links
are the unfamiliar
concept



THE MOST GENERAL PROBLEM

there is a session between two network nodes, . . .

. . . and we want it to benefit from the resources of multiple paths through the physical network

ON WHAT TIME SCALE?

- **simultaneously, to add the bandwidth of paths**

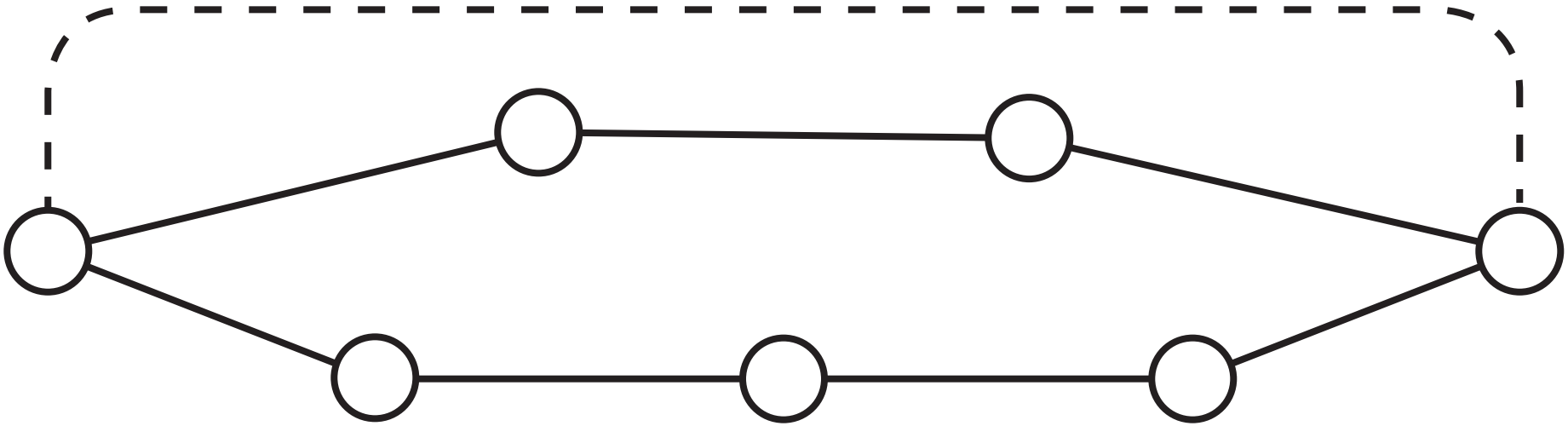
  the paths must be different even in the edge networks, so this is called "multihoming"

- **switching paths when the current one is slow or dead, for fault-tolerance, keeping all available**

- **one path goes dead before the next one is available**

  commonly called "mobility"

SOLUTION 1: MULTIPATH ROUTING

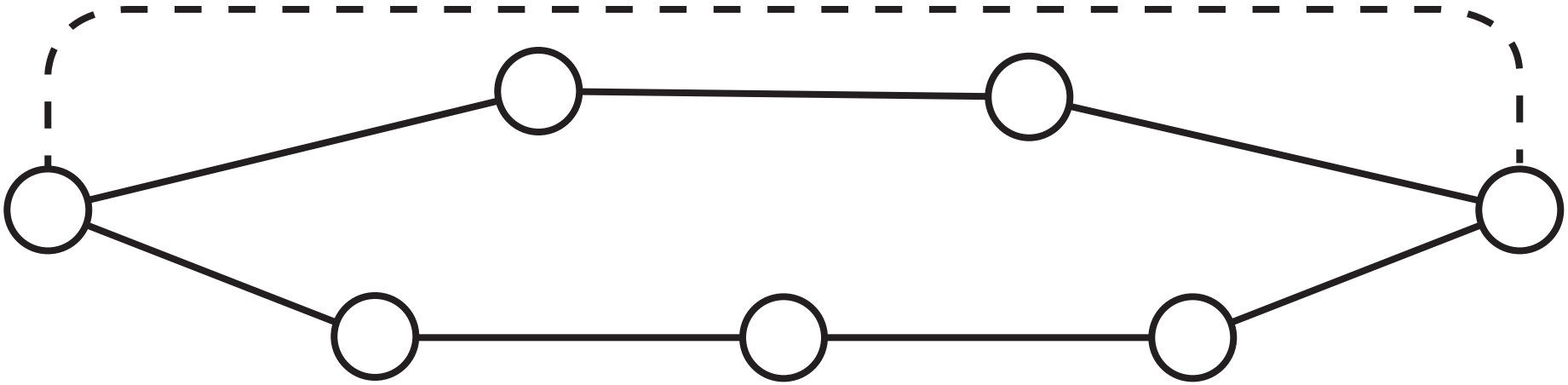


THERE IS MORE THAN ONE ROUTE BETWEEN THE ENDPOINTS

- some or all of distinct paths are implemented with different resources, but this is implicit

WHERE THE PATHS DIVERGE, THE NODE DECIDES WHICH PACKETS TO SEND ON WHICH PATH

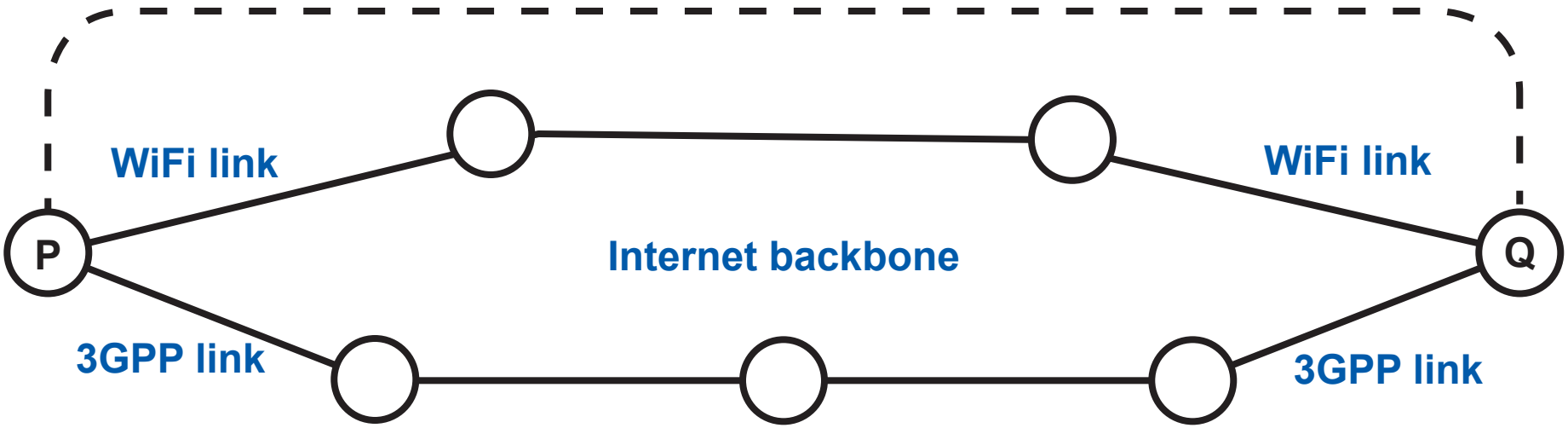
WHEN MULTIPATH ROUTING WORKS WELL



RON uses multipath routing

- used at the intermediate time scale, for fault-tolerance and enhanced performance
- the members of a RON do the multipath routing, which is easy because there are few members (and the set of possible paths is restricted!)
- because the paths are physically separated, they are known to use different physical resources

WHEN MULTIPATH ROUTING DOES NOT WORK WELL



What are P and Q?

since every access network has its own IP prefix, on some of the access networks P and Q will be anomalies

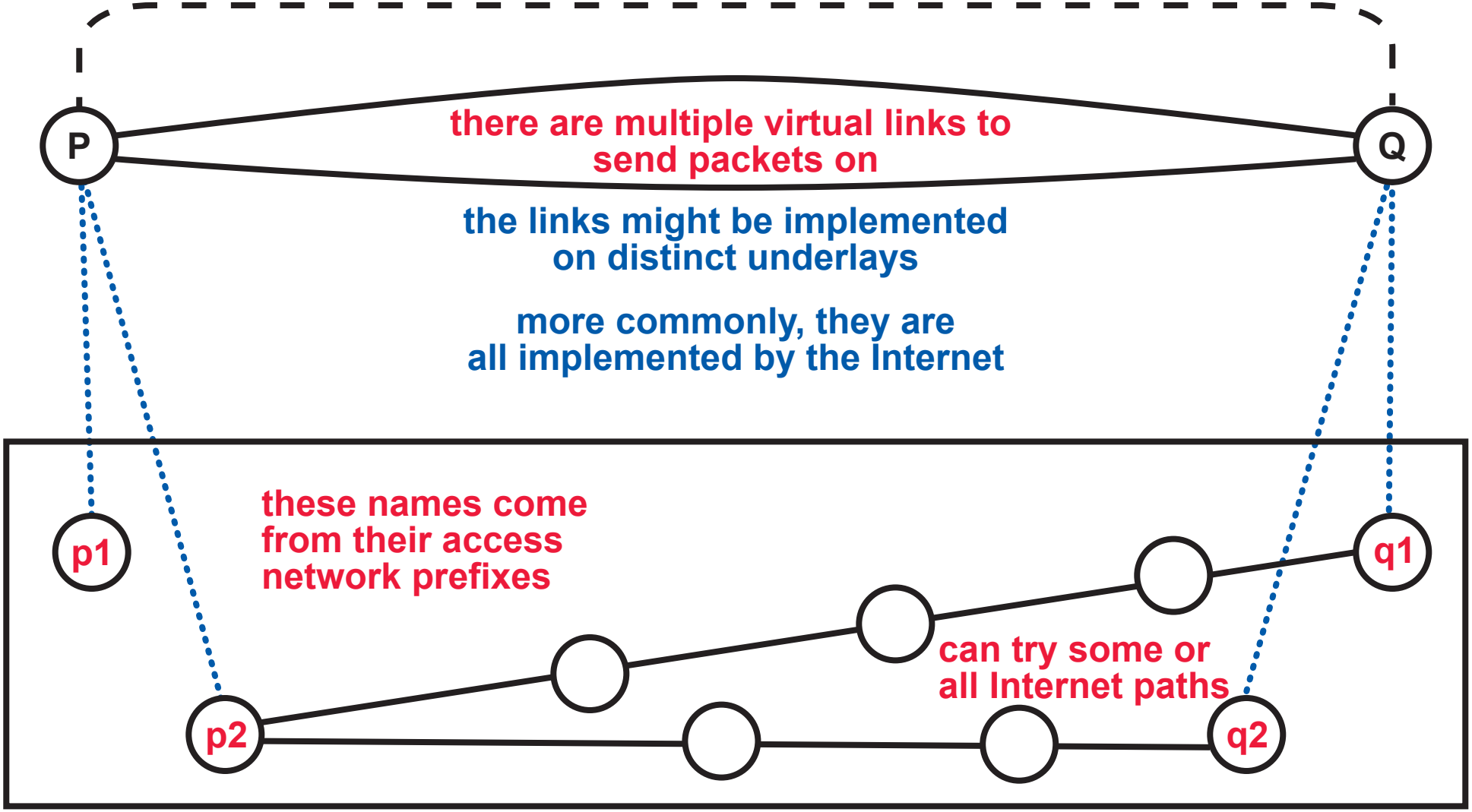
Internet routing will not find these paths, because it is based on address aggregation

EXAMPLES

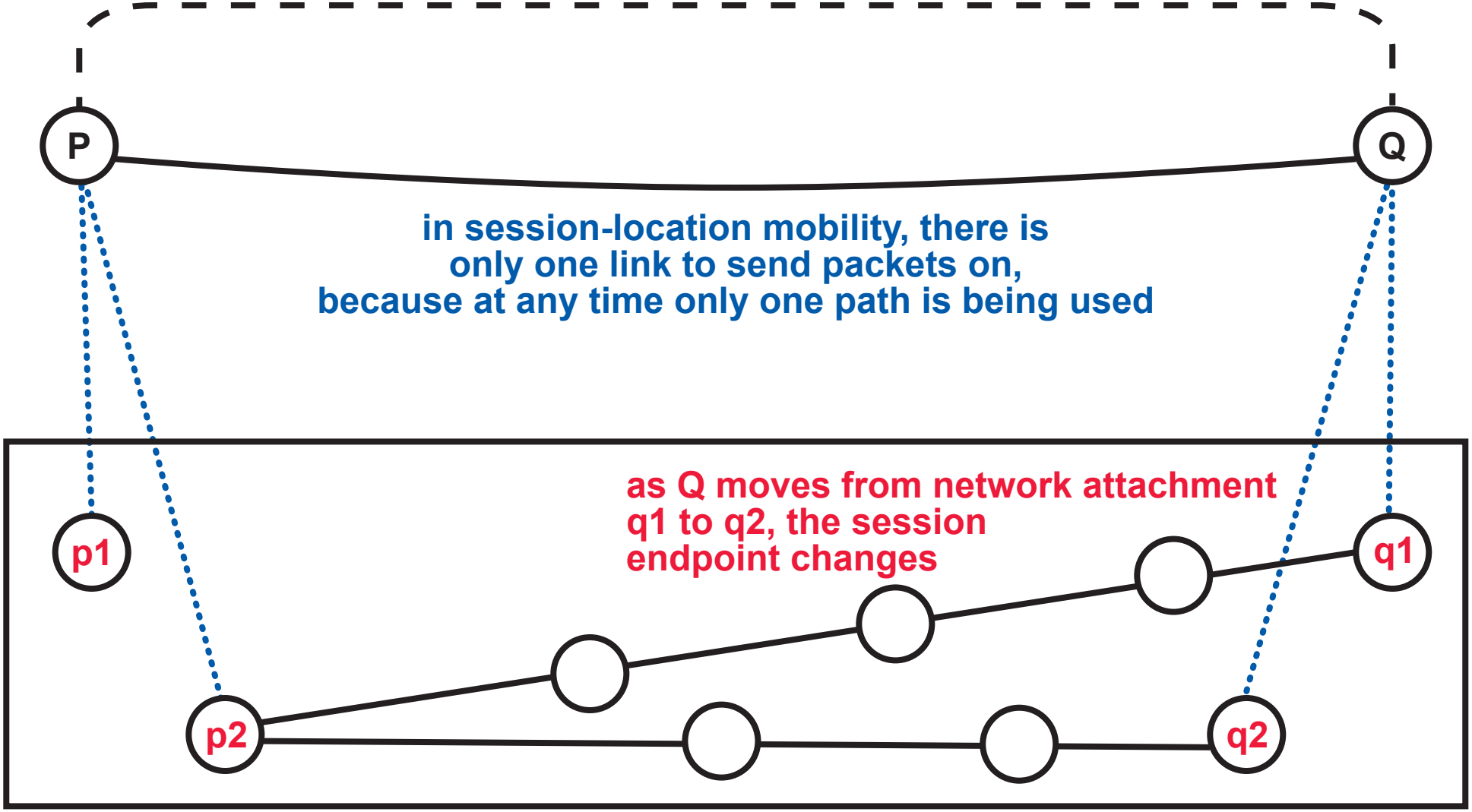
- this is what John Day recommends for multihoming in *Patterns in Network Architecture*, and we don't get it
- this also characterizes the dynamic-routing pattern for mobility

as we have seen, Mobile IP requires an escape from Internet routing to make this work

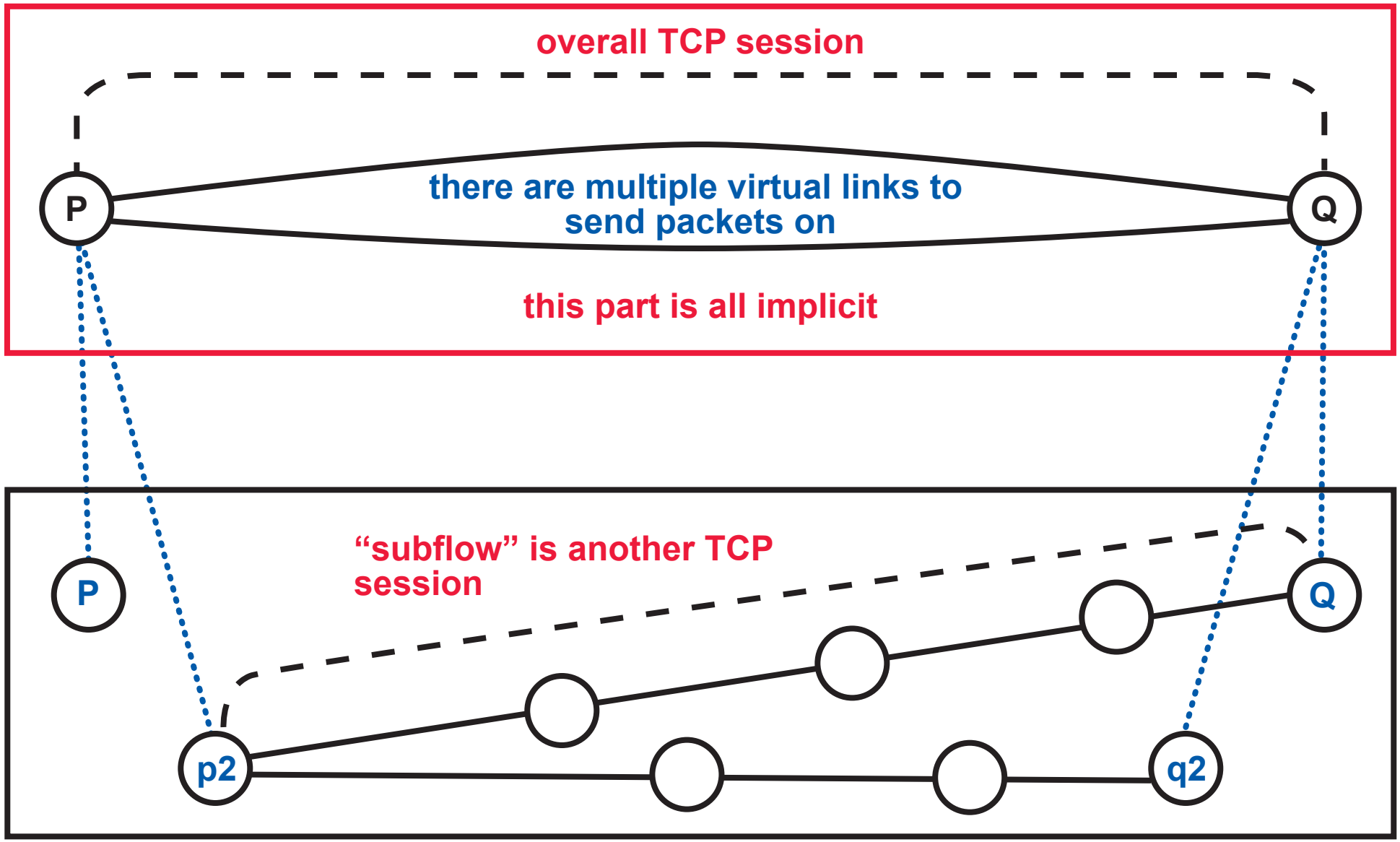
SOLUTION 2: MULTIPLE LINKS



EXAMPLE OF SOLUTION 2: SESSION-LOCATION MOBILITY



EXAMPLE OF SOLUTION 2: MULTIPATH TCP



PROBLEMS OF MPTCP: CONTROL SIGNALING

MPTCP REQUIRES MUCH MORE CONTROL SIGNALING ("METADATA") THAN TCP

- negotiate extra capabilities
- each subflow needs its own SYNs and FINs, which are distinct from those of the connection
- each subflow needs its own sequence numbers, acknowledg-ments, loss detection, and retransmission
- when you try to get clever by conflating or piggybacking information, there is always some interaction causing deadlock (this is the nature of protocols!)

THIS IS DIFFICULT BECAUSE . . .

. . . TCP does not leave much room for extra control signaling

. . . even when there is room (e.g., TCP options), on many paths the metadata is removed or altered

some alterations are broad-brush security: alter initial sequence numbers, remove TCP options

some alterations seem innocent: NICs resegment data, copying options

there is always the issue of composition: maybe some other feature needs the space!

WHAT COULD BE DONE ABOUT THESE PROBLEMS?

PROBLEMS OF MPTCP: OTHER PROTOCOL PROBLEMS

MPTCP ALLOWS SUBFLOWS TO BE SET UP IN EITHER DIRECTION, BUT THE INTERNET DOES NOT

this is the familiar NAT problem

more control signaling ("add address" option) is a reasonable solution

SOME MIDDLEBOXES CHANGE THE SIZE OF THE DATA

e.g., application-level gateways, ad insertion, compression or decompression

the sender divides the data into subflows and maps them back to the original sequence, which breaks when a subflow changes size

THE SUBFLOWS REQUIRE EXTRA BUFFER SPACE, WHICH MAY NOT BE UTILIZED WELL

WHAT COULD BE DONE ABOUT THESE PROBLEMS?

PROBLEMS OF MPTCP: MIDDLEBOXES

The paper focuses on the problem of getting subflows to pass through middleboxes, i.e., on satisfying the reachability or progress requirements.

It ignores the safety or security requirements—in particular, some middleboxes *must* see all the data of the TCP connection.

e.g., parental controls

Dysco provides enough control to get all the subflows to one middlebox, but . . .

. . . Would Dysco (which also alters TCP) work with MPTCP?

. . . How would the middlebox make sense of the subflows?

ANOTHER VIEW OF MPTCP

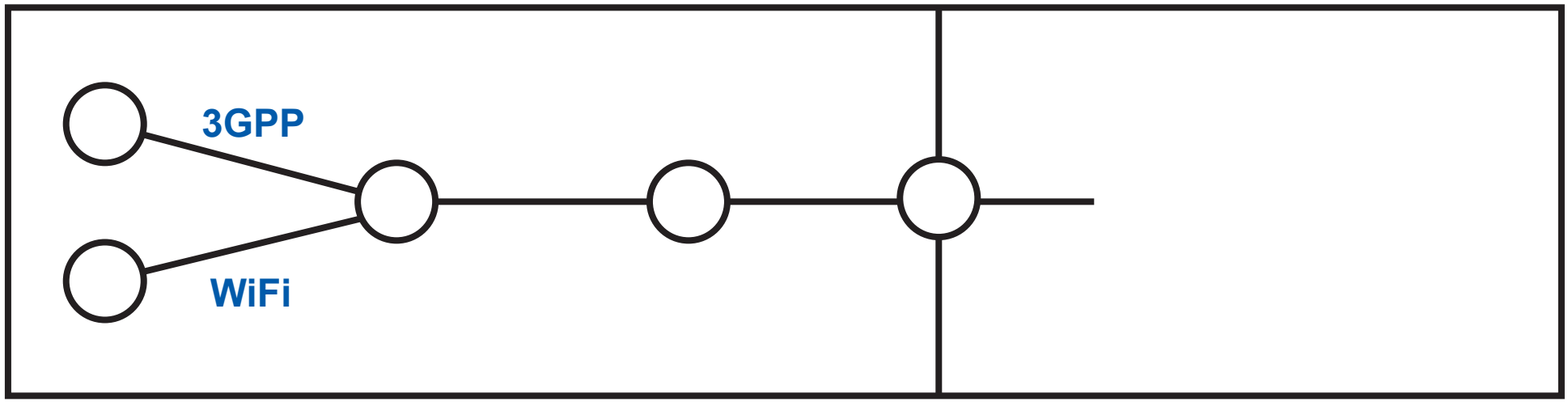
the problem of adding the bandwidths of multiple wireless networks is not end-to-end!

why should the other end know, care, or cooperate?

rather, it is a problem of bridging and interoperation

access network of multipath customer, with a proxy for merging paths

open Internet



customer's WiFi and 3GPP attachments

multipath proxy

middlebox that needs to see all the data

of course, the creators of MPTCP have no way to deploy this solution, hence current design

MOSTLY ORDERED MULTICAST

Multicast is a non-point-to-point communication service. Packets sent to a multicast name are delivered to all members of a multicast group.

TO ADD MULTICAST TO OUR MODEL, WE MUST ANSWER MANY QUESTIONS

- **can a member have a multicast and no individual name?**
- **if the service is to be added to our model, there must be both multicast links and multicast sessions—does a multicast link or session have a group of nodes that are allowed to send, or does each sender have a separate multicast link/session?**
- **what are the inter-layer mappings to show that a multicast session properly implements a multicast link?**
- **how would you model the implementation of a multicast session in Alloy, using point-to-point links?**