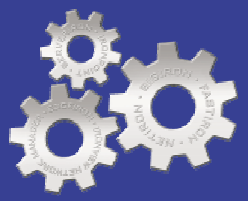
Boosting Capacity Utilization in MPLS Networks using Load-Sharing

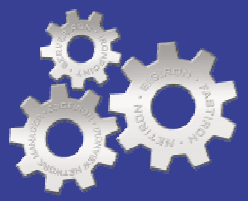
MPLS JAPAN 2007

Sanjay Khanna
Foundry Networks
skhanna@foundrynet.com



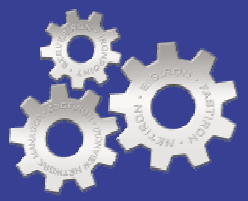
Agenda

- **Why we need Load-Sharing**
- **Methods to boost capacity**
  - **Trunks/Link Aggregation**
  - **Routing Protocols ECMP**
  - **MPLS Signaling Protocols ECMP**
  - **Mapping MPLS payloads to LSPs**
- **Methods for efficient utilization**
  - **Packet Forwarding Schemes**
  - **Enhancements to Existing Schemes**



Why Load Sharing

- Utilize investment in existing infrastructure
- Ability to add bandwidth in small increments
- Cost-effective to add bandwidth
- 1+N link protection
- End to End protection with diverse paths
- Avoid idling of backup paths
- Transport carrier not offering higher bandwidth links
- 100GbE not available yet

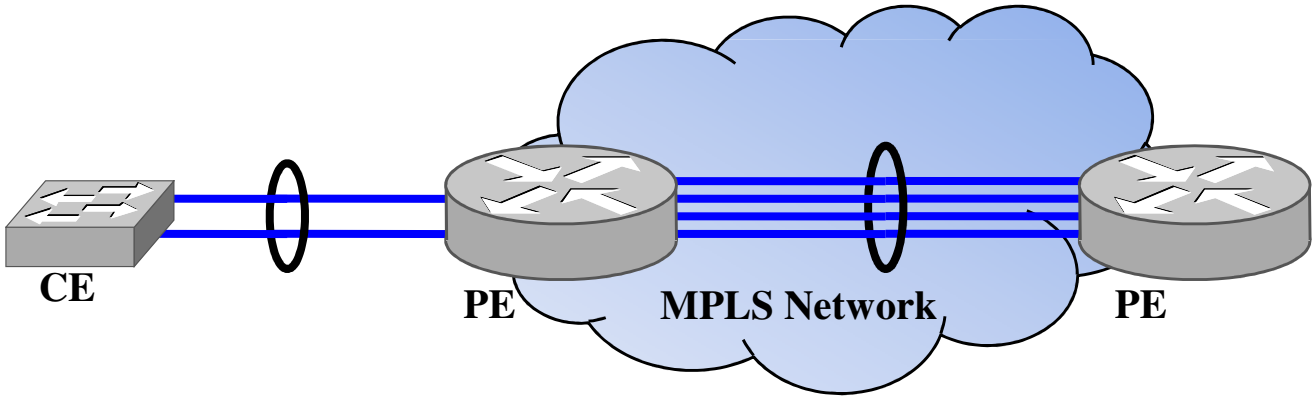


Factors affecting MPLS Load Sharing

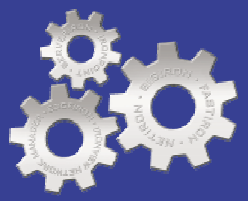
- **Trunks**: Provide multiple links for load-sharing
  - Link Aggregation/bundling
- **Protocols**: Determine multiple paths for load-sharing
  - Routing Protocols: IGP, BGP
  - MPLS Protocols: LDP, RSVP-TE
- **MPLS payload mapping:** Assign MPLS services to LSPs
  - BGP/MPLS-VPN & IPv4/v6 mapping to LSPs
  - PW mapping to LSPs
- **Data Forwarding:** Decision on how packets are load-shared
  - Per packet based
  - Flow based

Link Aggregation/Trunking

- A very popular method of bundling multiple physical links between 2 devices
- Used on both local side and network side

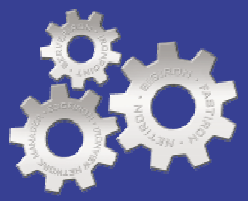


- Typically, higher layer protocols unaware of the link bundling



Routing Protocols ECMP

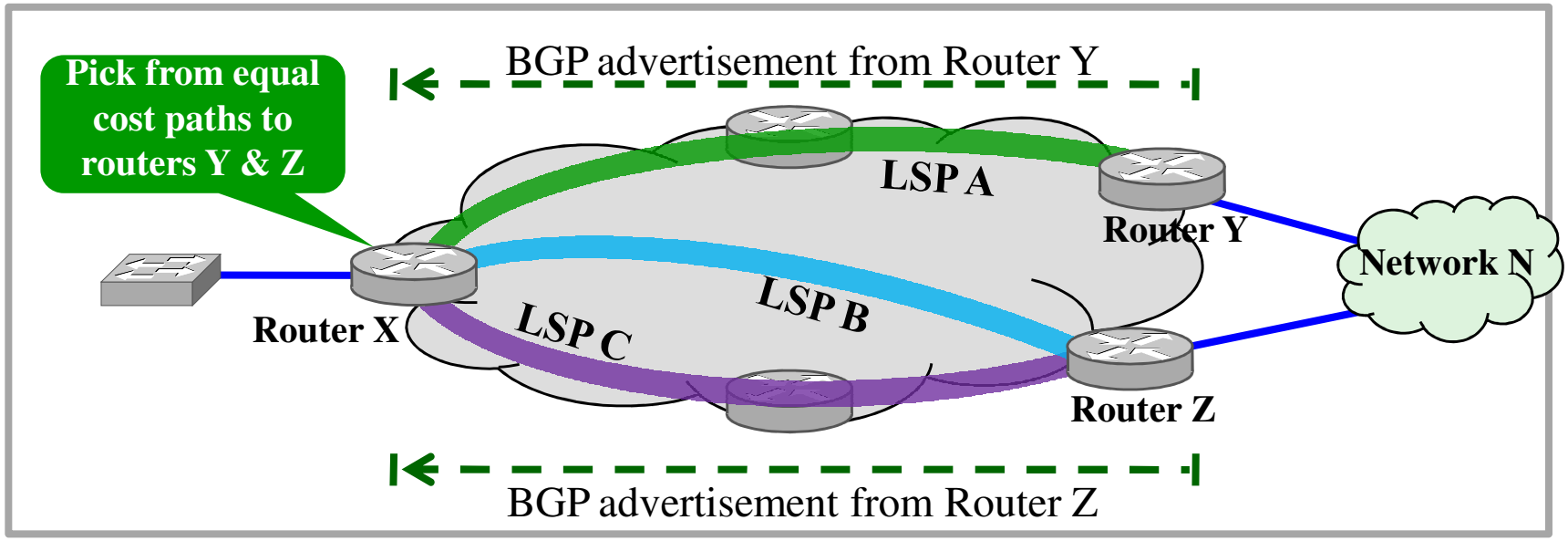
- IGP (ISIS/OSPF): Determine multiple equal cost paths to a destination
  - Affects paths taken by MPLS LSPs
    - IGP affects ECMP determination for LDP and non-CSPF RSVP-TE LSPs
    - IGP-TE affects CSPF RSVP-TE ECMP determination
  - Import LSPs as IGP Shortcuts into IGP
    - Allows router to use multiple equal cost LSP paths
- BGP: Determine multiple equal cost paths to a destination
  - Multiple equal cost paths to different next-hops reachable by diverse LSPs
  - Multiple LSP paths to a BGP next-hop
  - BGP can calculate multiple equal cost label paths for Inter-AS



MPLS Signaling Protocols ECMP

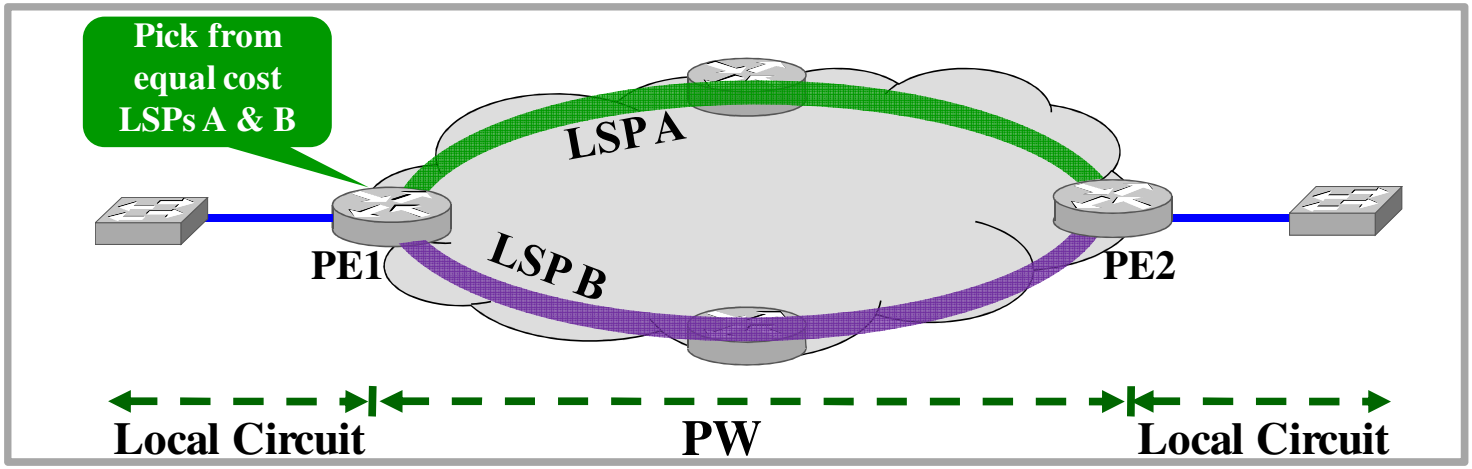
- MPLS signaling allows multiple LSPs to the same destination
- RSVP-TE: Selects a path for a LSP from multiple equal cost paths that satisfy the LSP constraints, as determined through CSPF
  - Typical criteria used:
    - Hops: Pick the path with least number of hops
      - Less probability of failure
    - Least-fill: Pick the path with highest available bandwidth
      - Even spread of traffic
    - Most-fill: Pick the path with lowest available bandwidth
      - Leave room for higher bandwidth LSPs
- LDP: Allows a prefix to be reachable through multiple equal cost label paths

IP Mapping to LSPs:
For IPv4/v6 Routing and BGP/MPLS-VPNs

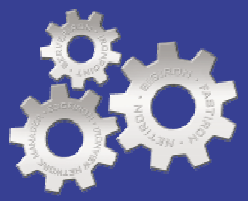


- Typical mapping criteria used:
  - Assign a prefix to single LSP
    - Better predictability
  - Map prefixes within a VRF to single LSP
    - Better operator control
  - Load-share on per flow basis
    - Better traffic distribution

PW Mapping to LSPs:
For VPWS and VPLS



- Typical mapping criteria used:
  - Bind PW to least used LSP (LSP with lowest number of PWs)
    - Good distribution of traffic
  - Bind PW to LSP with most available bandwidth or same class of service
    - Useful for services with dedicated bandwidth requirements
  - Explicitly bind PW to LSP
    - Better operator control
  - PW traffic split across multiple LSPs
    - Difficult to monitor and guarantee SLAs, potential packet reordering



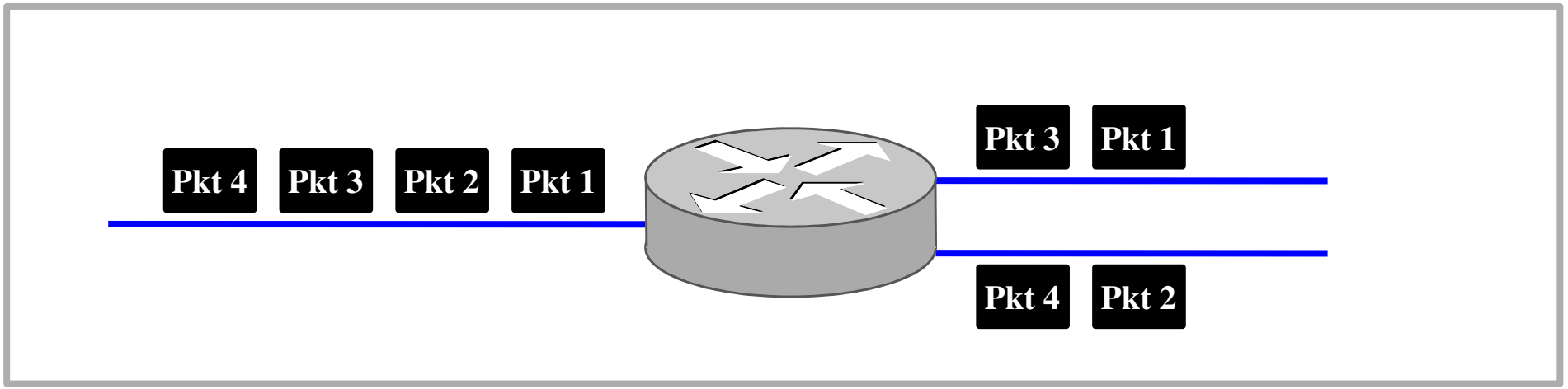
Load-Sharing in the Forwarding Plane

- Routing Protocols ECMP, MPLS Signaling Protocols ECMP and Link Aggregation provide multiple paths for a MPLS router to forward traffic

- 2 common schemes to load-share traffic over these equal cost paths/links
  - Packet based
  - Hash/Flow based

Packet based Forwarding

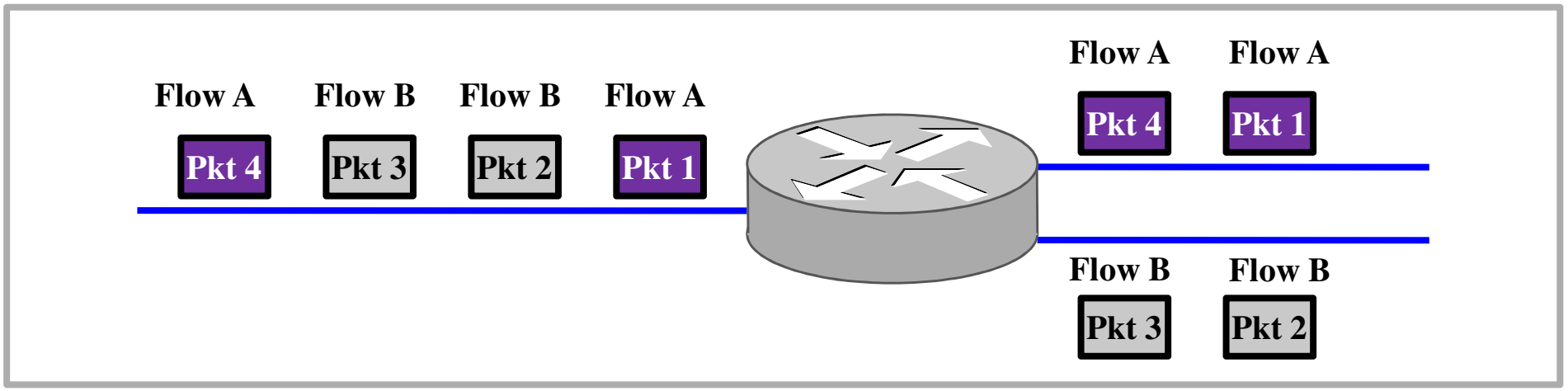
- Each packet in turn is sent on the next link



- Perfect load balancing
- Potential packet reordering issues
- Possible increase in latency and jitter for some flows

Flow/Hash based Forwarding

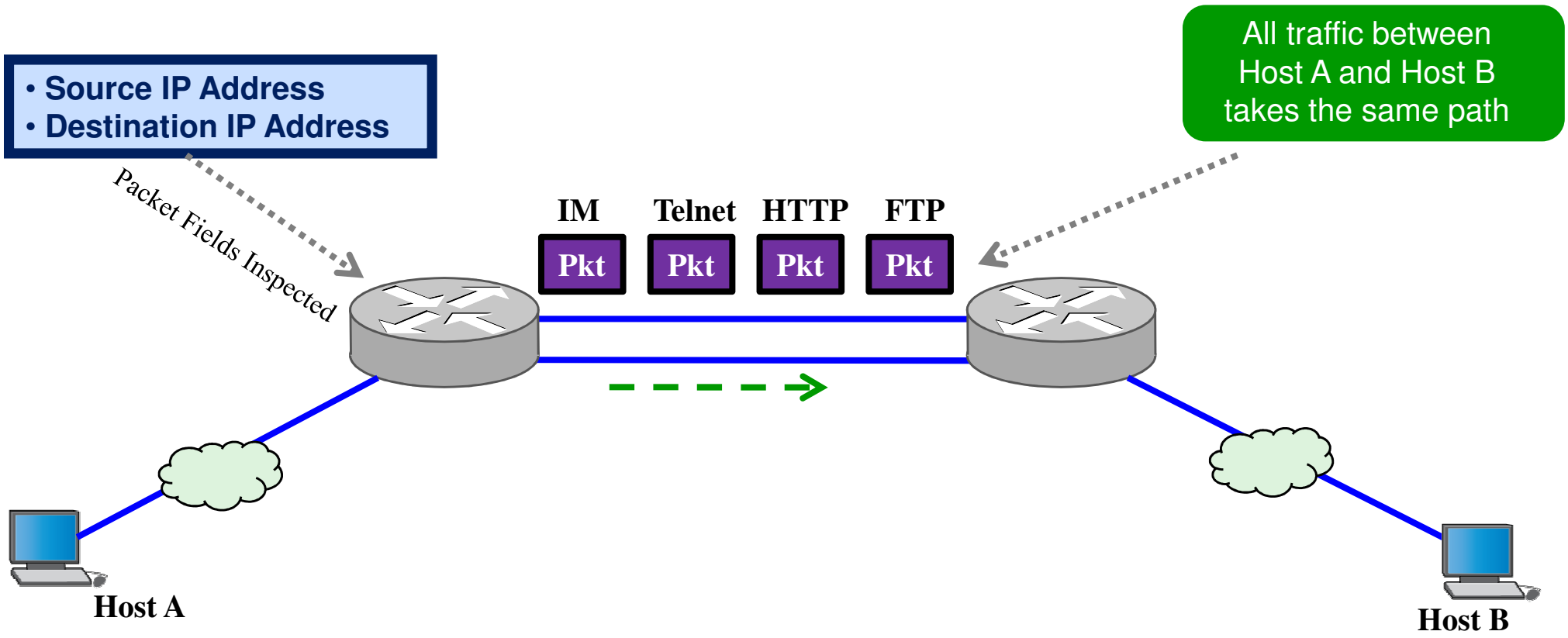
- Identifies packets as flows
  - Based on packet content such as IP header
- Keeps flows on the same path
  - Maintains packet ordering



- Hashing is the most popular scheme for flow based forwarding

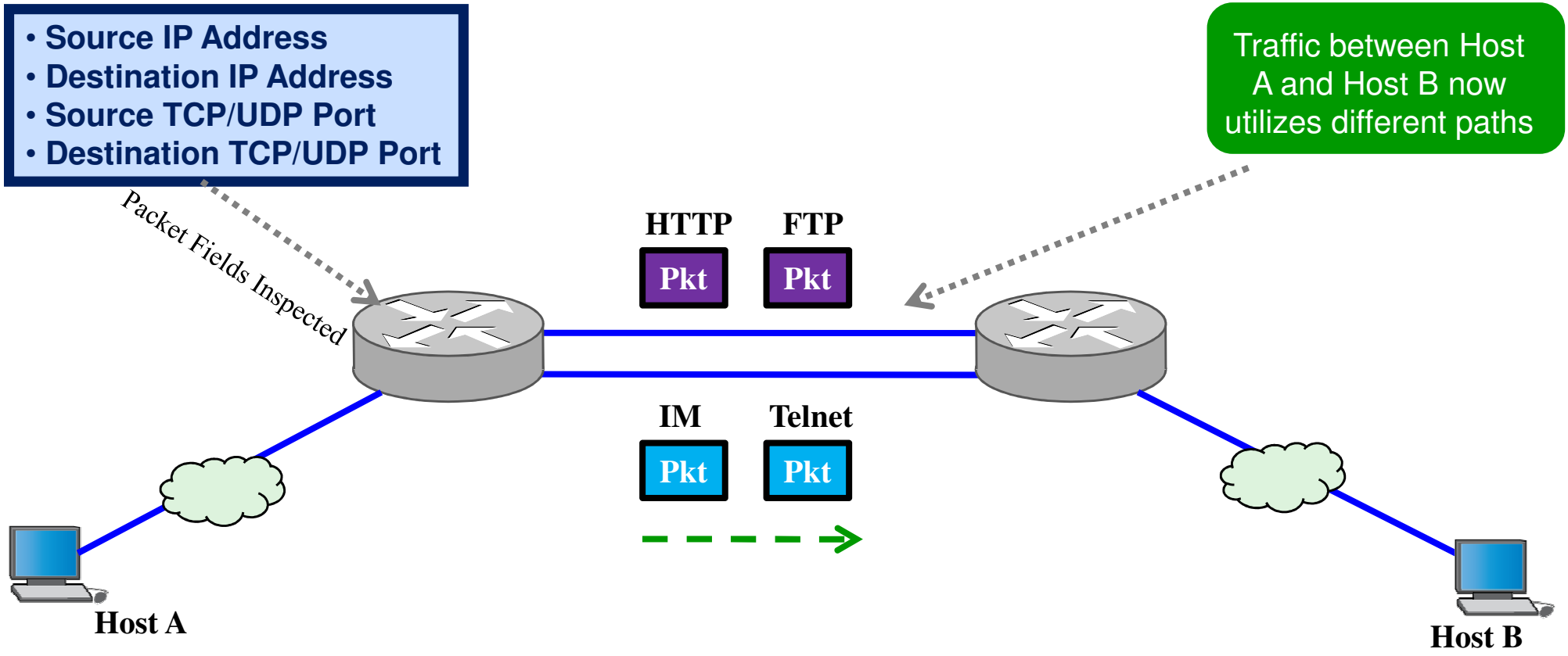
Hash Based Forwarding for L3 Flows (1)

- Flows based on Source IP & Destination IP addresses
  - All traffic between 2 hosts relegated to one path
    - Can lead to over-utilization of one path



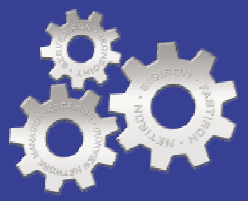
Hash Based Forwarding for L3 Flows (2)

- Flows based on L3 and L4 information
  - Better traffic distribution for applications between 2 hosts



- Applicable to IPv4 and IPv6 packets

FOUNDRY NETWORKS

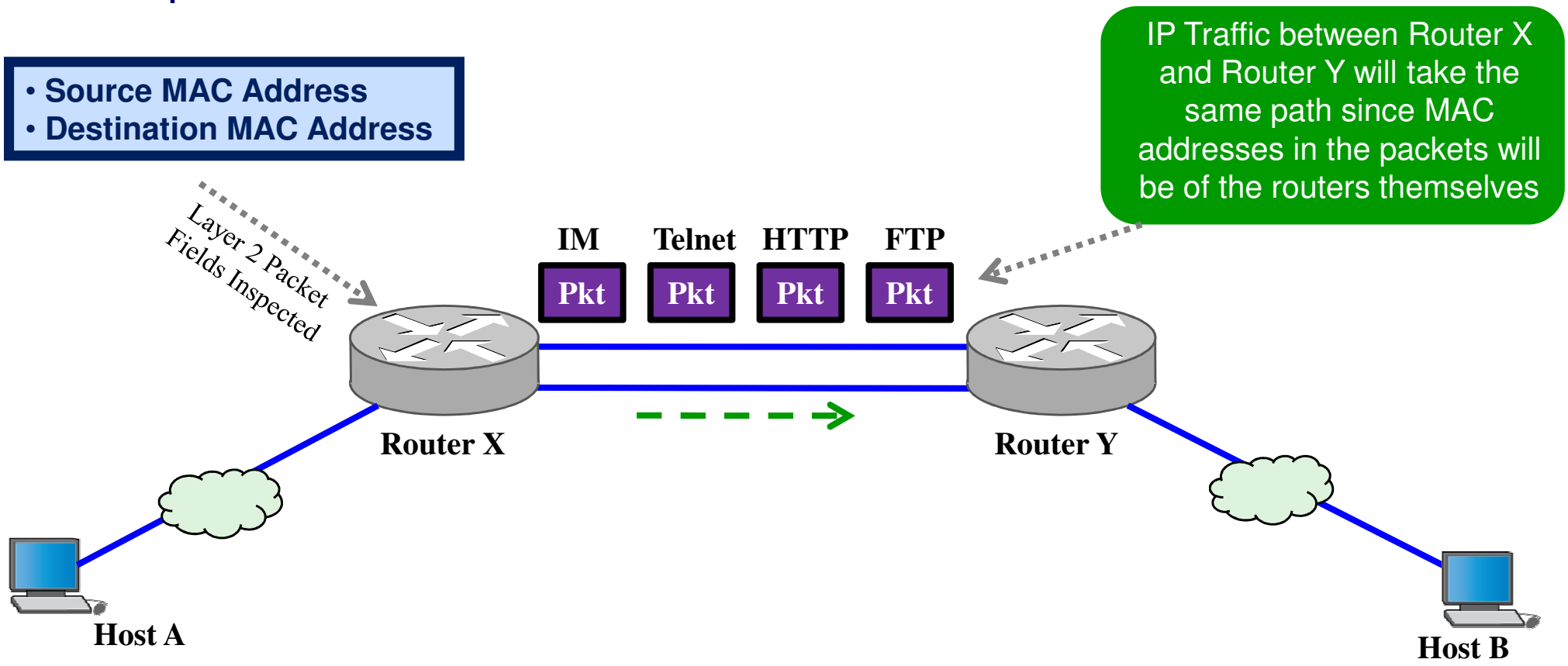


Hash Based Forwarding for L3 Flows (3)

- Note:
  - Fragmentation Issue:
    - Example: If payload is fragmented in IP packets, only the first IP packet carries the L4 information
- Solution:
  - Set TCP segment size lower than the IP MTU to avoid fragmentation
  - Load balance fragmented packets using L3 information only

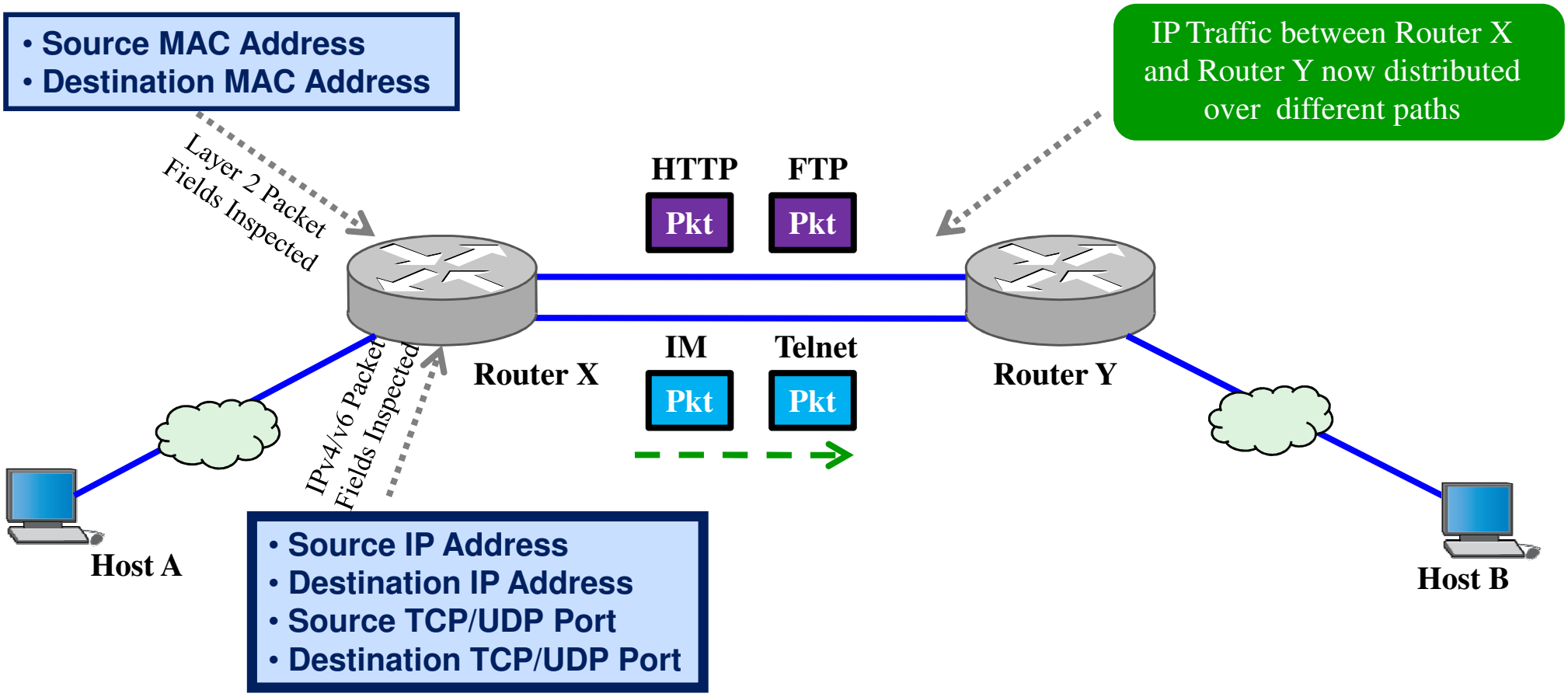
Hash based Forwarding for L2 flows (1)

- Flows based on Layer 2 information in packet
  - However, IP packets between 2 routers will always take the same path



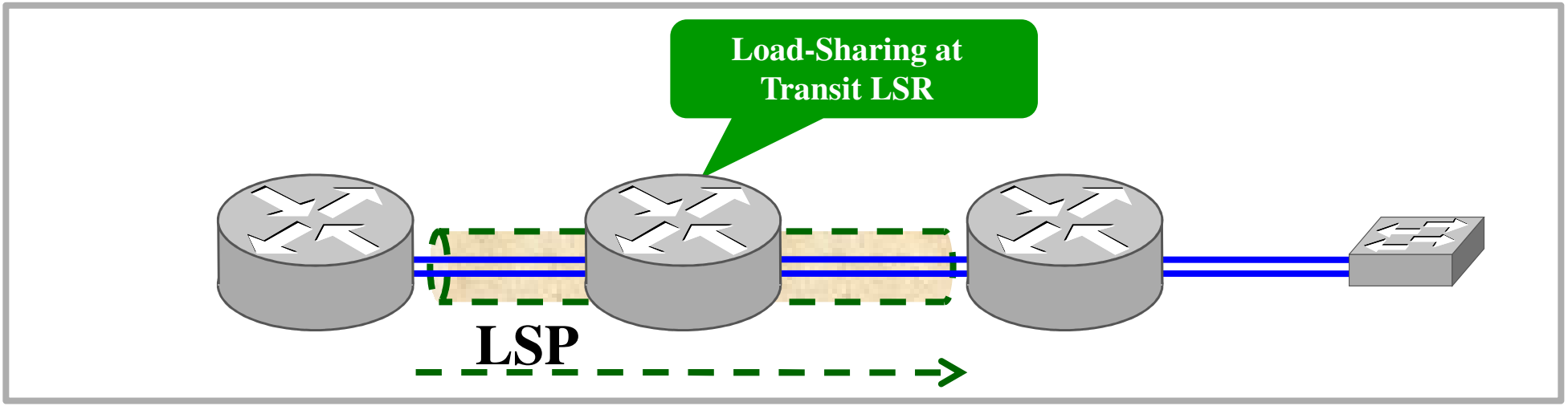
Hash based Forwarding for L2 flows (2)

- Determine IPv4/v6 packets in L2 flows:
  - Flows based on Layer3/4 (and optionally Layer2) information for IPv4/v6 packets
  - Flows based on Layer 2 information for other packets



FOUNDRY NETWORKS

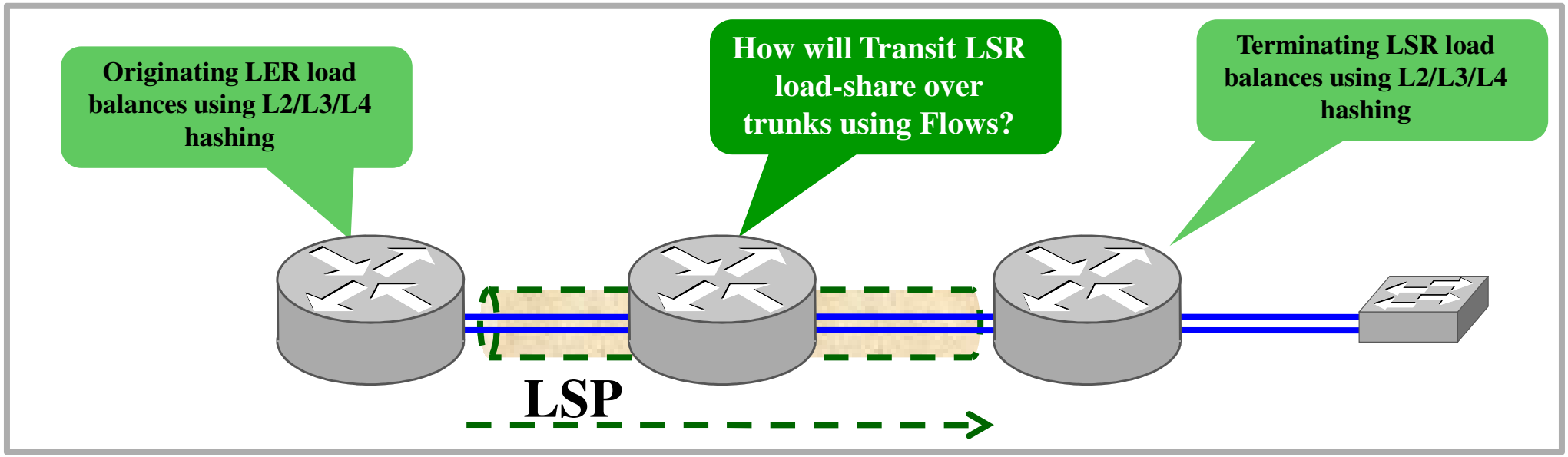
Forwarding on LSRs



- Transit LSR can load-share over multiple links between 2 LSRs
- Typical criteria used:
  - Per packet: Potential packet reordering issue
  - LSP assigned to one link: High usage LSPs will over-utilize one link
  - Per Inner/Outer Labels: High usage PWs/VPN labels will over-utilize one link
  - Per flow: Better distribution of traffic
    - Requires *Packet Speculation*

Forwarding on LSRs: *Packet Speculation*

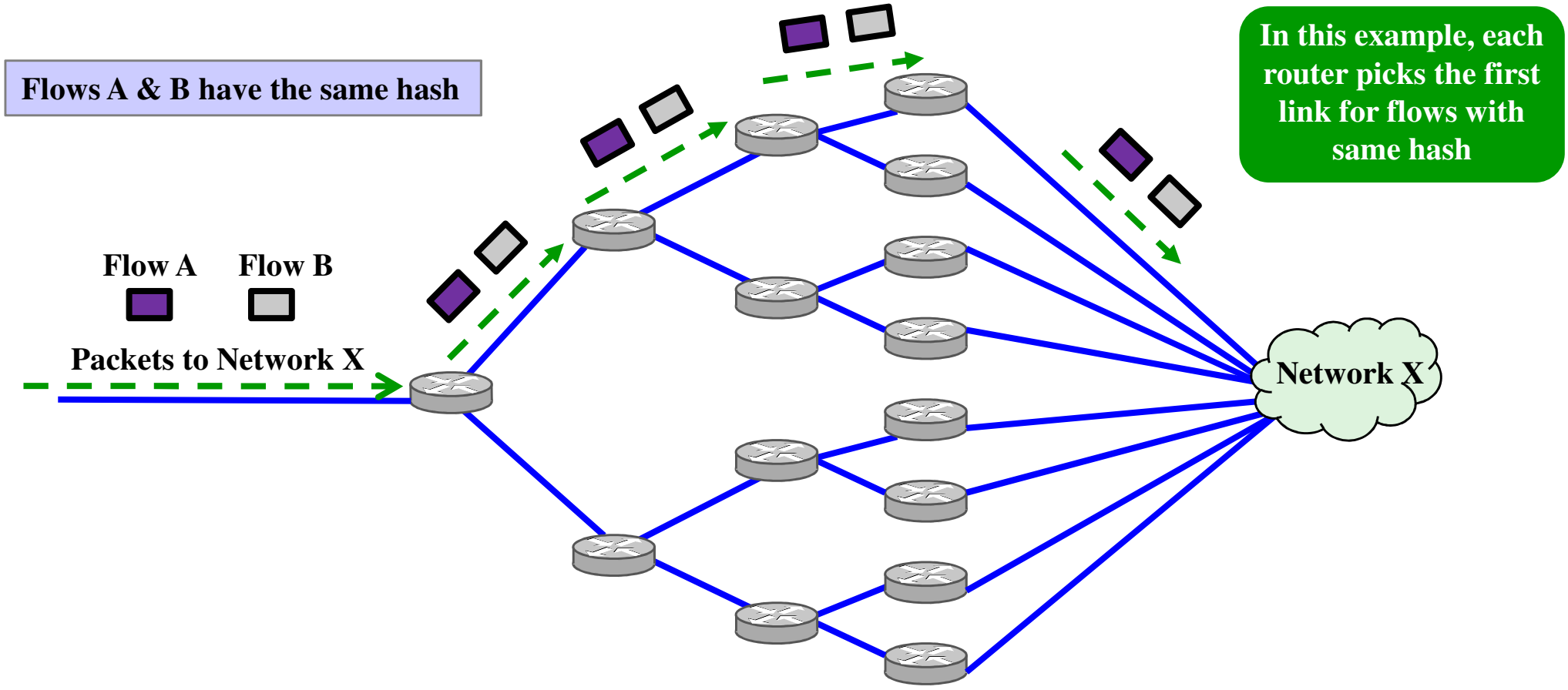
- Transit LSRs have no information on packet payload



- Transit LSR speculates on the packet type
  - Checks first nibble after bottommost label
    - If 4/6, speculates on packet as IPv4/IPv6
    - Else (optionally) speculates on packet as Ethernet
  - Can now load-share using "LSP Label/VC label/L2/L3/L4 headers"

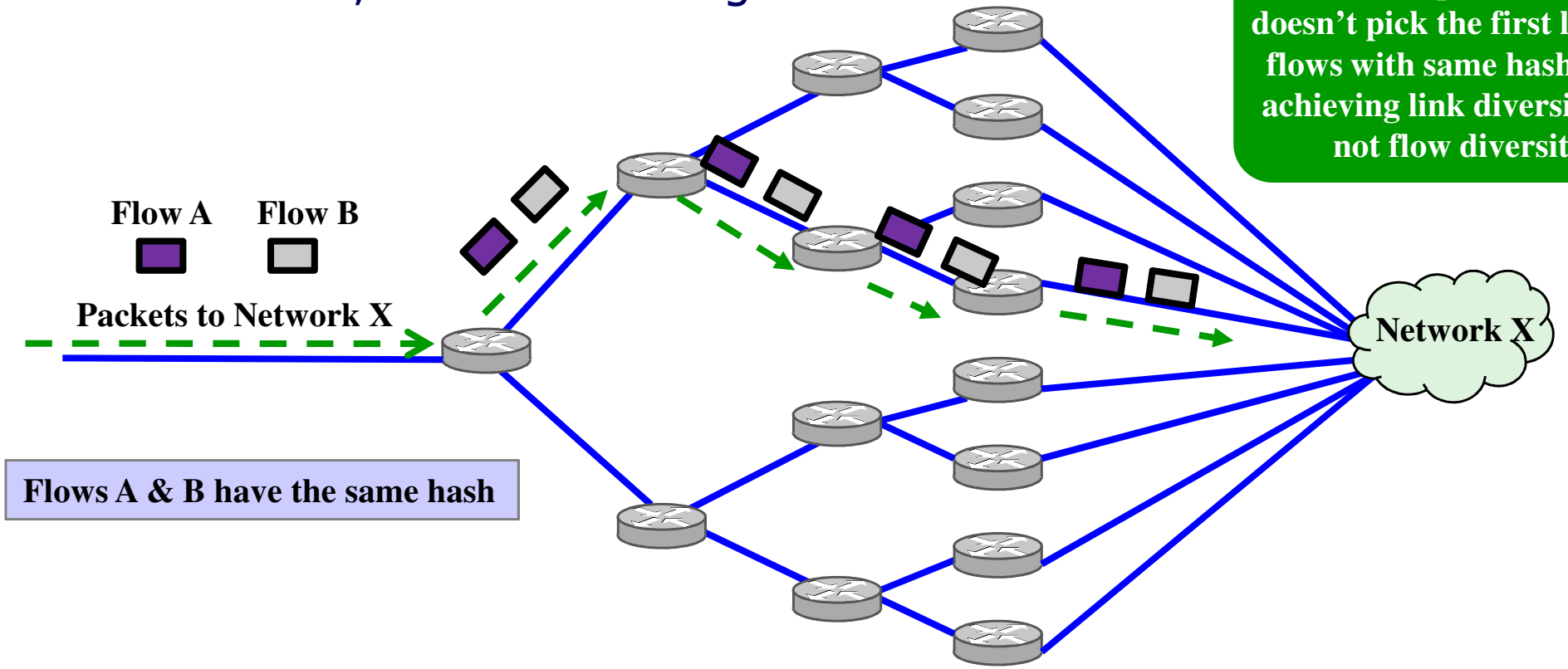
Hash based forwarding issues and solutions:
Polarization Effect

- Similar routers pick the same path for flows with identical hash
  - Leads to over-utilization of some parts of the network



Hash based forwarding issues and solutions:
Hash Diversification (Neutralizes Polarization Effect)

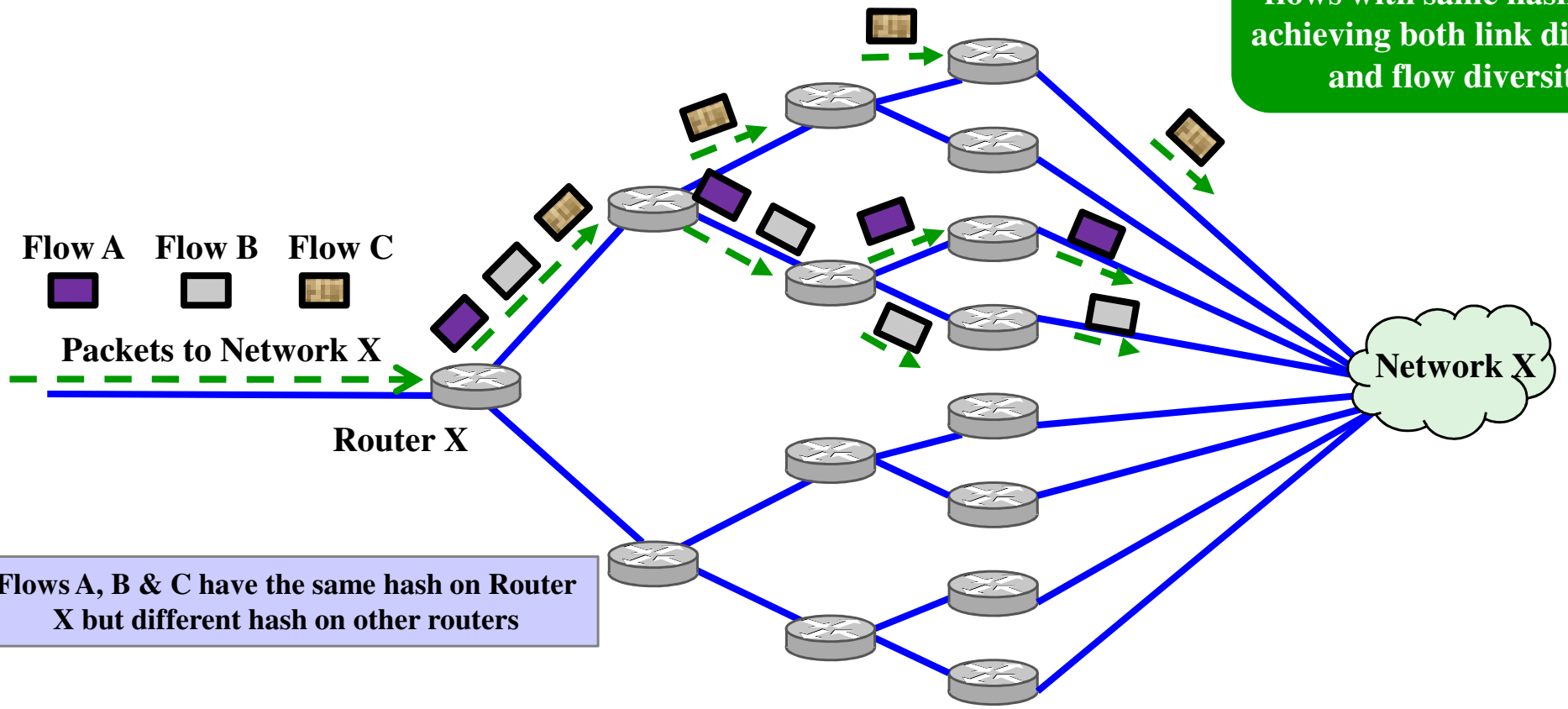
- Each router uses a unique-id per router in hash calculations
  - Alternatively, hashing using Source and Destination MACs may give comparable results in most scenarios
  - Similar routers now pick different links
  - However, flows are still together on same links

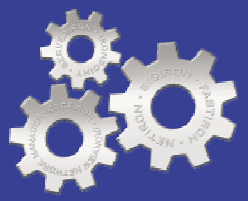


FOUNDRY NETWORKS

Hash based forwarding issues and solutions:
Hash Variation (Neutralizes Polarization Effect)

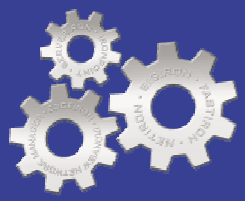
- Each router uses a different variant of the hash algorithm
  - Flows can now be distributed





Summary

- Load-Sharing is a cost-effective technique to improve network utilization
  - Works over multiple paths and links
- Multiple methods to boost capacity at various layers
  - Can effectively increase throughput beyond the current limits of physical link capacity
- Flow/Hash based forwarding offers many advantages for efficient utilization of the increased capacity
  - Watch out for polarization effects and neutralize them
- Not a one size fits all approach
  - Choose optimal schemes based on traffic types and operator policy



Questions?