Carnegie Mellon
CyLab
CONFIDENCE FOR A NETWORKED WORLD

Carnegie Mellon CyLab
4616 HENRY STREET
PITTSBURGH, PA 15213
PH: (412) 268-7195
FX: (412) 268-7196
www.cylab.cmu.edu

SCION:

Scalability, Control and Isolation On Next-Generation Networks

Current main team: Soo Bum Lee, Hsu-Chun Hsiao, Hyun Jin Kim, Yue-Hsun Lin, Sangjae Yoo, Adrian Perrig, Virgil Gligor

Previous members: Xin Zhang, Geoff Hasker, Haowen Chan, David Andersen

Fundamental (S-)BGP Limitations

- Lack of routing isolation
  - A failure/attack can have global effects
  - Global visibility of paths is not scalable
- Source / destination lack path control
- Slow convergence / route oscillation
- Route inconsistencies
  - Forwarding state may be different from announced state
- Large routing tables
  - Multi-homing / flat namespaces prevent aggregation
- Lack of route freshness
  - Current (S-)BGP cannot prevent replay of old paths

Note that these issues are fundamental to (S-)BGP, they cannot be easily fixed by small changes!

S-BGP Limitations

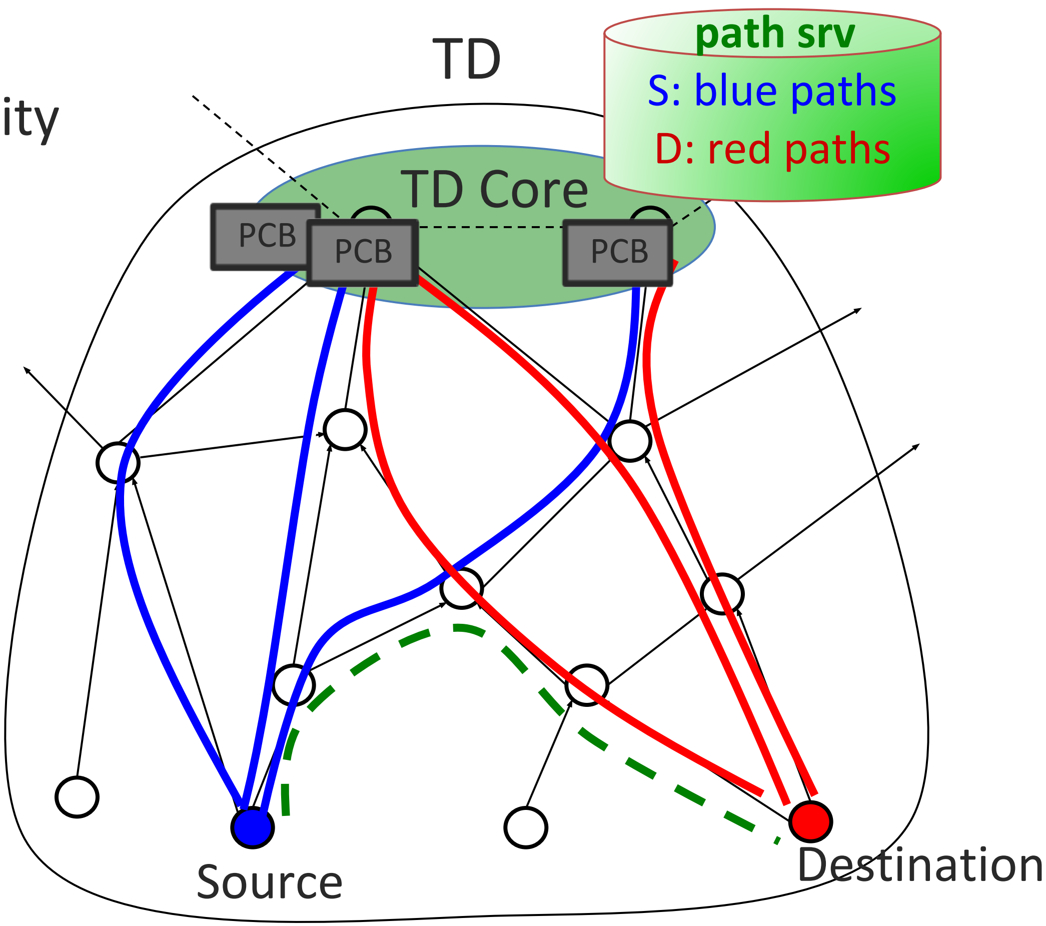
- Slow convergence
- Router outage causes high overhead
- Circular dependency between UPDATE message and connectivity with RPKI server
  - Route Origin Authentication (ROA), prefix certificate and BGPSEC router certificate needs to be downloaded to validate UPDATE message!
  - Rebooting Internet would be very slow as initial UPDATE messages cannot be validated
- Route flap dampening can be misused
  - Ensure an AS's updates are ignored
  - Prevent updates to fix a path
  - Potential to create a loop that persists

SCION Architectural Goals

- High availability, even for networks with malicious parties
  - Communication should be available if attacker-free path exists
- Explicit trust for network operations
- Minimal TCB: minimize trusted entities for any operation
  - Strong isolation from untrusted parties
- Operate with mutually distrusting entities
  - No single root of trust
- Balanced route control for ISPs, receivers, senders
- No circular dependencies during setup: enable rebootability
- Simplicity, efficiency, flexibility, and scalability

SCION Architecture Overview

- Trust domain (TD)s
  - Isolation and scalability
  - Enforceable accountability
- Path construction
  - Path construction beacons (PCBs)
- Path resolution
  - Control
  - Explicit trust
- Route joining (shortcuts)
  - Efficiency, flexibility



Trust Domain Decomposition

- Global set of TD (Trust Domains)
  - Map to geographic, political, legal boundaries
  - Usually corresponds to a jurisdiction
  - Provide enforceable accountability
- TD Core: set of top-tier ISPs that manage TD
  - Route to other TDs
  - Initiate path construction beacons
  - Manage Address and Path Translation Servers
  - Handle TD membership
  - Root of trust for TD: manage root key and certificates
- AD: Autonomous Domain
  - Transit AD or endpoint AD

Part 1: Implementation

SCION Components

- Certificate Server
  - Certificate, Policy, Topology, Key management
- Beacon Server
  - Path Construction (PCB propagation, Path selection/registration (req), Path distribution)
- Path Server
  - Path registration/resolution
- Border Router
  - Opaque Field verification, packet forwarding
- Switch
  - Abstract intra-domain routing
- Gateway
  - Backward compatibility

Root Of Trust File

```
<ROT>
	<header>
		<policyNumber> “Policy Number” </policyNumber>
		<TDID> “Trust Domain ID” </TDID>
		<policyThreshold> “Policy Threshold” </policyThreshold>
		<certificateThreshold> “Certificate Threshold” </certificateThreshold>
	</header>
	<coreADs>
		<coreAD>
			<AID>”AID”</AID>
			<publicKey>”Public Key”</publicKey>
		</coreAD>
		     :
		     :
		<coreAD>
			<AID>”AID”</AID>
			<publicKey>”Public Key”</publicKey>
		</coreAD>
	</coreADs>
	<signatures>
		<coreAD>
			<AID>”AID”</AID>
			<sign>”Signature”</ sign>
		</coreAD>
		     :
		     :
		<coreAD>
			<AID>”AID”</AID>
			<sign>”Signature”</sign>
```

Topology File

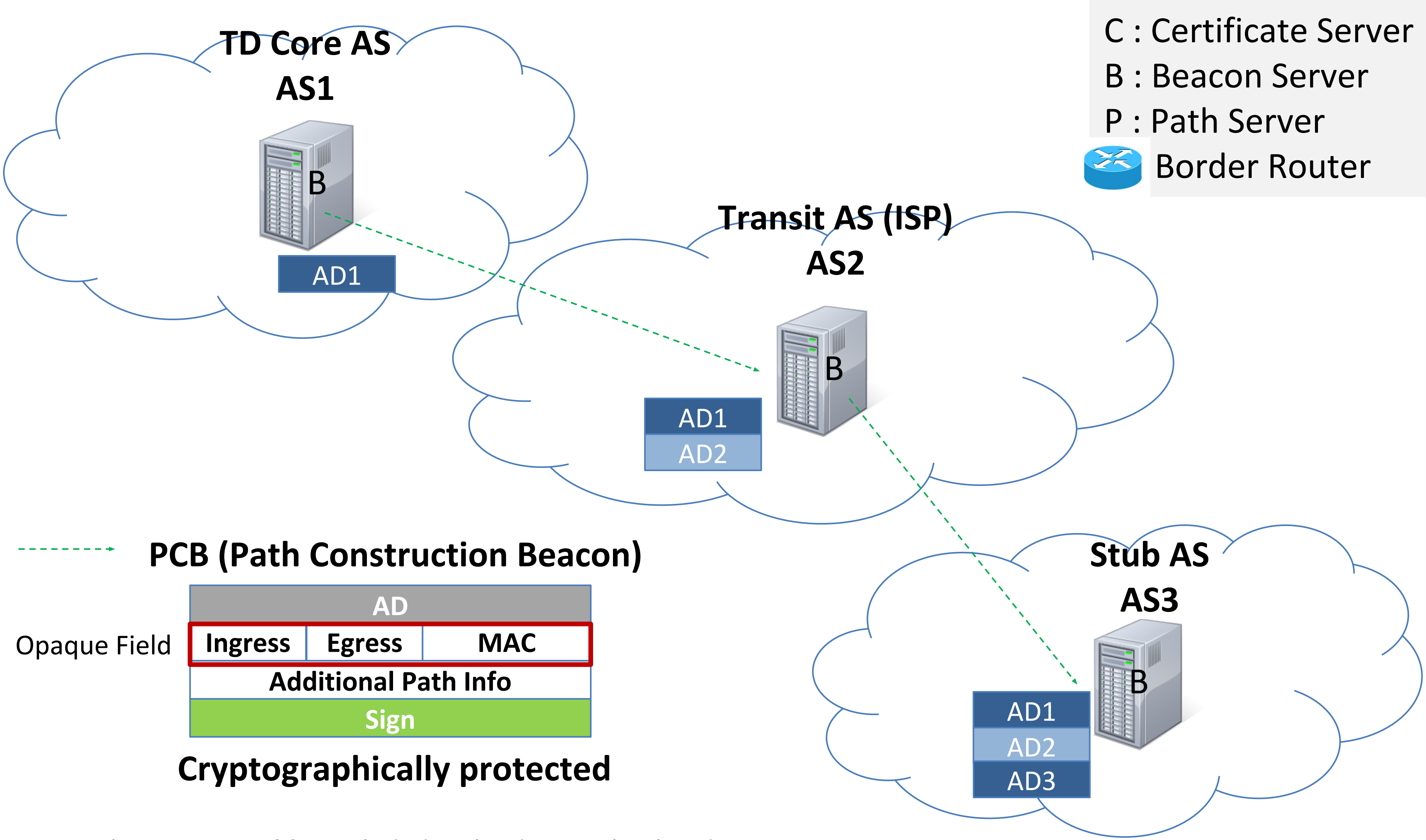
```
<TDID> “TD ID” </TDID>
<ADID> “AD ID” </ADID>
<Topology>
        <Servers>
                <BeaconServer> “AID” </BeaconServer>
                <PathServer> “AID” </PathServer>
                <CertificateServer> “AID” </CertificateServer>
        </Servers>
        <BorderRouters>
                <Router>
                        <AID>”AID”</AID>
                        <Interface>
                                <IFID>”Interface ID”</IFID>
                                <NeighborAD>”AD ID”</NeighborAD>
                                <NeighborType>”Neighbor Type”</NeighborType>
                        </Interface>
                        <Interface>
                                <IFID>”Interface ID”</IFID>
                                <NeighborAD>”AD ID”</NeighborAD>
                                <NeighborType>”Neighbor Type”</NeighborType>
                        </Interface>
                            :
                </Router>
                    :
                    :
        </BorderRouters>
        <Gateways>
                <Gateway>
                        <AID>”AID”</AID>
                        <Protocol>”Protocol Type”</Protocol>
```

Path Construction

Goal: each endpoint learns multiple verifiable paths to its core

- Discovering paths via Path Construction Beacons (PCBs)
  - TD Core periodically initiates PCBs
  - ADs asynchronously propagate PCBs
- ADs perform the following operations
  - Collect PCBs
  - For each customer/peer AD, select which *k* PCBs to forward
  - Update cryptographic information in PCBs
- Endpoint AD receives at least *k* PCBs from each provider AD, selects *k* down-paths to advertise

SCION Component: Beacon Server



PCB (Path Construction Beacon)

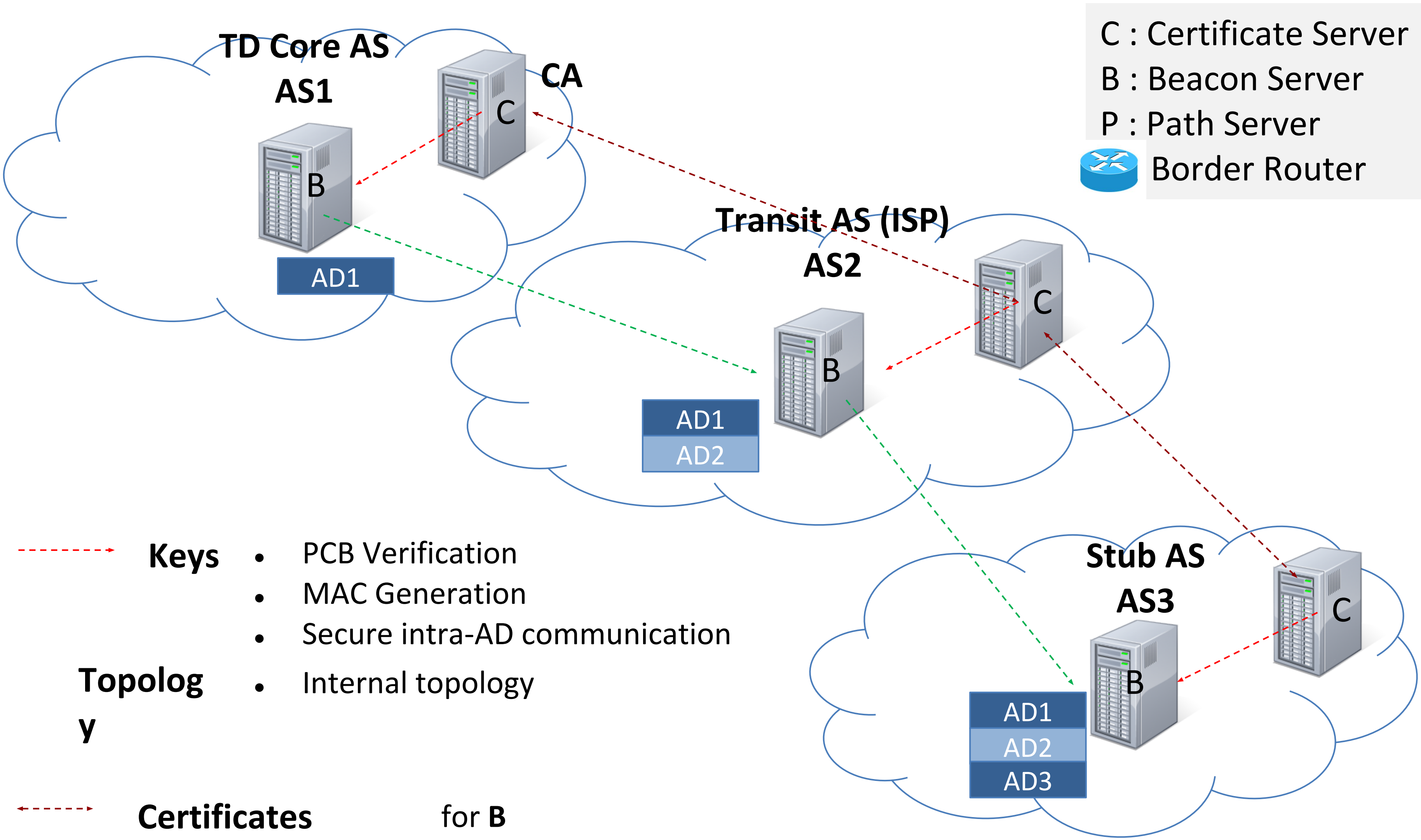
	AD		
Opaque Field	Ingress	Egress	MAC
	Additional Path Info		
	Sign		

Cryptographically protected

Only **Opaque Field** is included in the data packet header

* **Additional Path Info**: bandwidth, policy, pricing…

SCION Component: Cert. Server

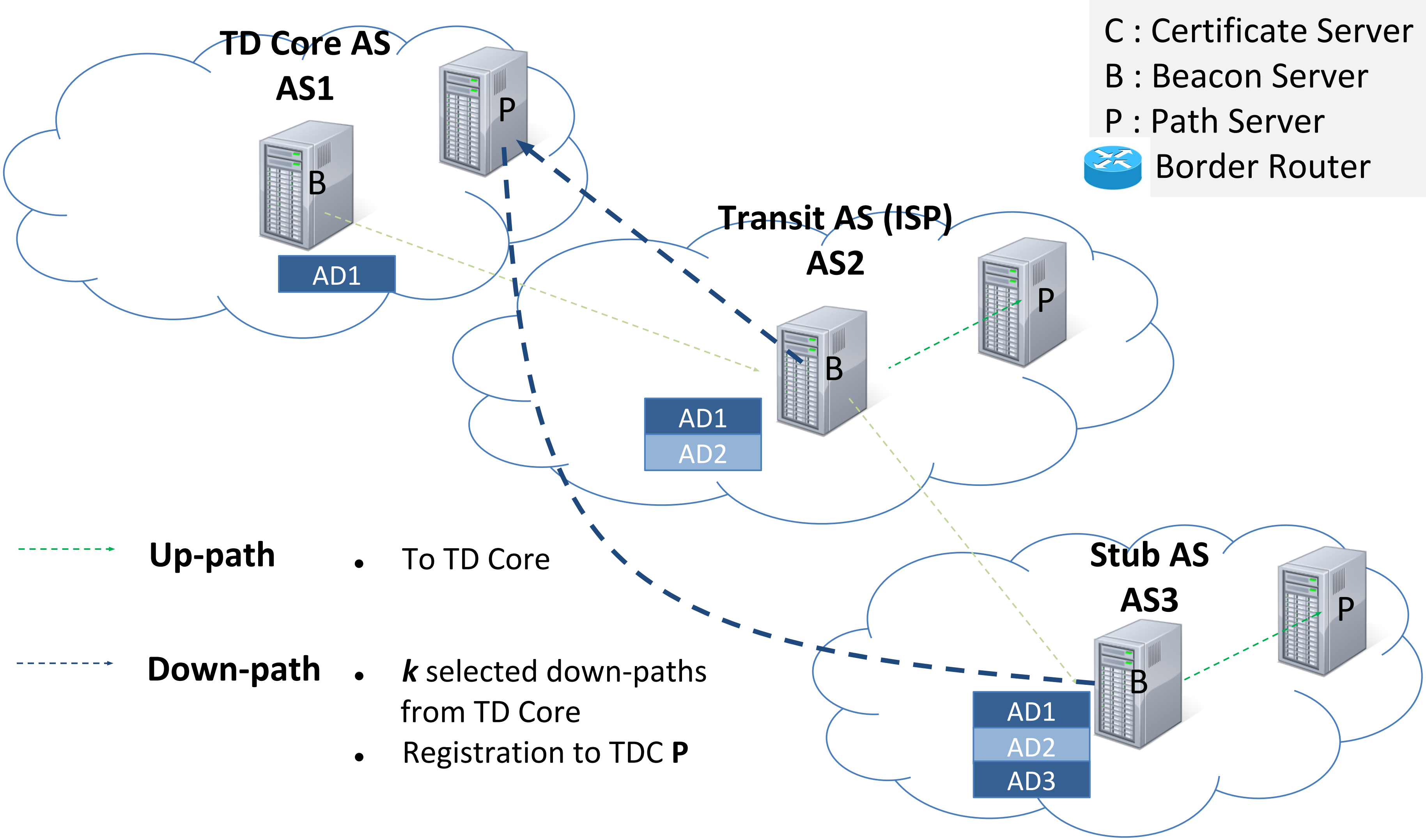


Keys
- PCB Verification
- MAC Generation
- Secure intra-AD communication

Topolog y
- Internal topology

Certificates for **B**
Root of Trust File

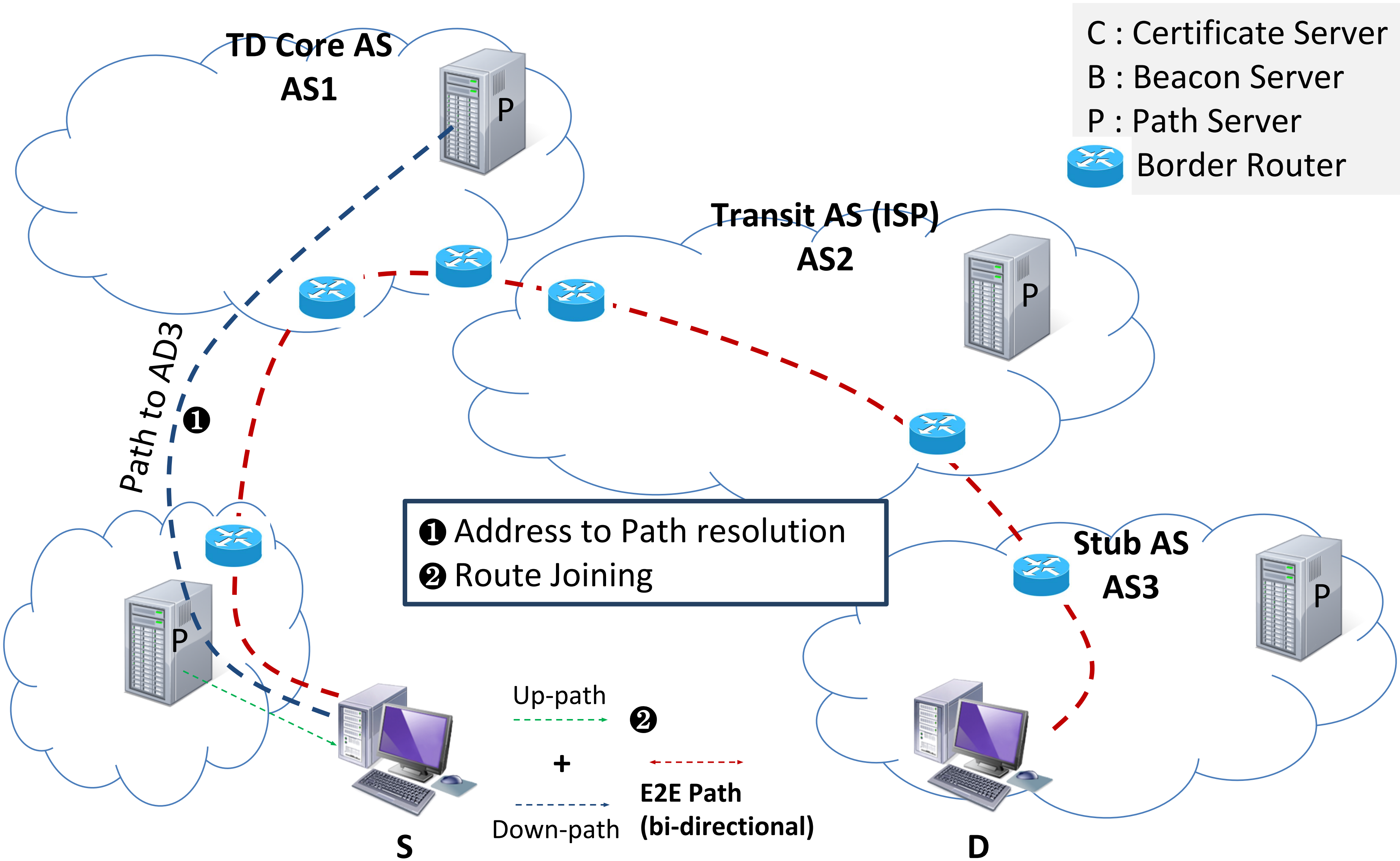
SCION Component: Path Server



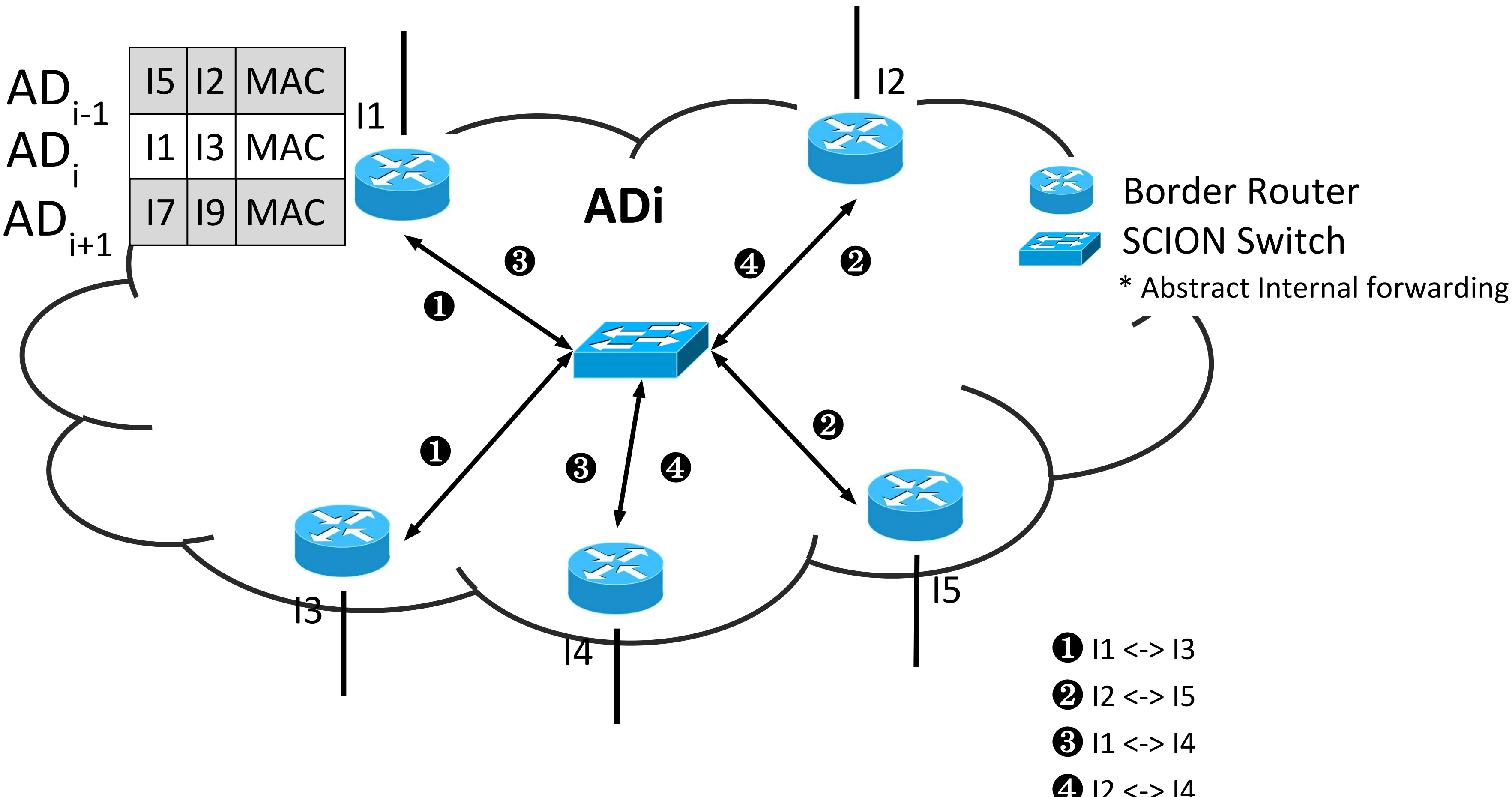
- To TD Core

- *k* selected down-paths from TD Core
- Registration to TDC **P**

End-to-End Path Construction



SCION Component: Border Router



In real deployment,
SCION Switch would be replaced by IP / MPLS

SCION Packet Format (PCB)

	0-7	8-15	16-23	24-31	32-39	40-47	48-55	56-63	
Special Opaque Field	0x80	Timestamp				TDID		#AD hops	
AD [1]	0xFF	ROT Version				IFID		reserved	
	AD AID 1								
Opaque Field (× k’)	Certificate ID (4B index)				LSign		Block Size		× k”
OF support Info	Type	Ingress IF	Egress IF	EXP	MAC				× k”
	TDID (inter-TD peer)		BW Alloc		CLS(1bit)|Reserved				× k”
Signature	**RSA Signature** (LSign Bytes)								
AD [2]	AD AID 2								
⋮									
AD [n]	AD AID n								

AD AID: authenticated ID of AD; e.g., ASN; if Egress IF is 0, this field indicates peer AD
Ingress IF (14bits): ingress interface id (internal use)
Egress IF (14bits): egress interface id (internal use) or egress interface id of peering link if the OF is for the peering link (for peering link, Egress IF is same as that of previous OF because they are marked by the same AD)
EXP: lifetime, current assignment: 0x00-6HR, 0x01-12HR, 0x02-18HR, 0x03-24HR
LSign: Signature length (e.g., 1024 bits, 2048 bits)
Block Size: total marking block size of AD[i], which includes all peering links
MAC: Massage Authentication Code, MAC(i) = AES-CBC-MAC_{K_i}(Ingress IF||Egress IF||OF(i-1)||AIDi+1)
TD ID: Trusted Domain ID, only for Inter-TD peering link
BW Alloc: bandwidth allocation for STRIDE
Certificate ID: ID of the certificate used for signature generation; used for informing public key change to downstream ADs
Signature: RSA Signature signed by AD[i] over its marking (including chaining)
CLS: Bandwidth Class (Static, Dynamic, BE)

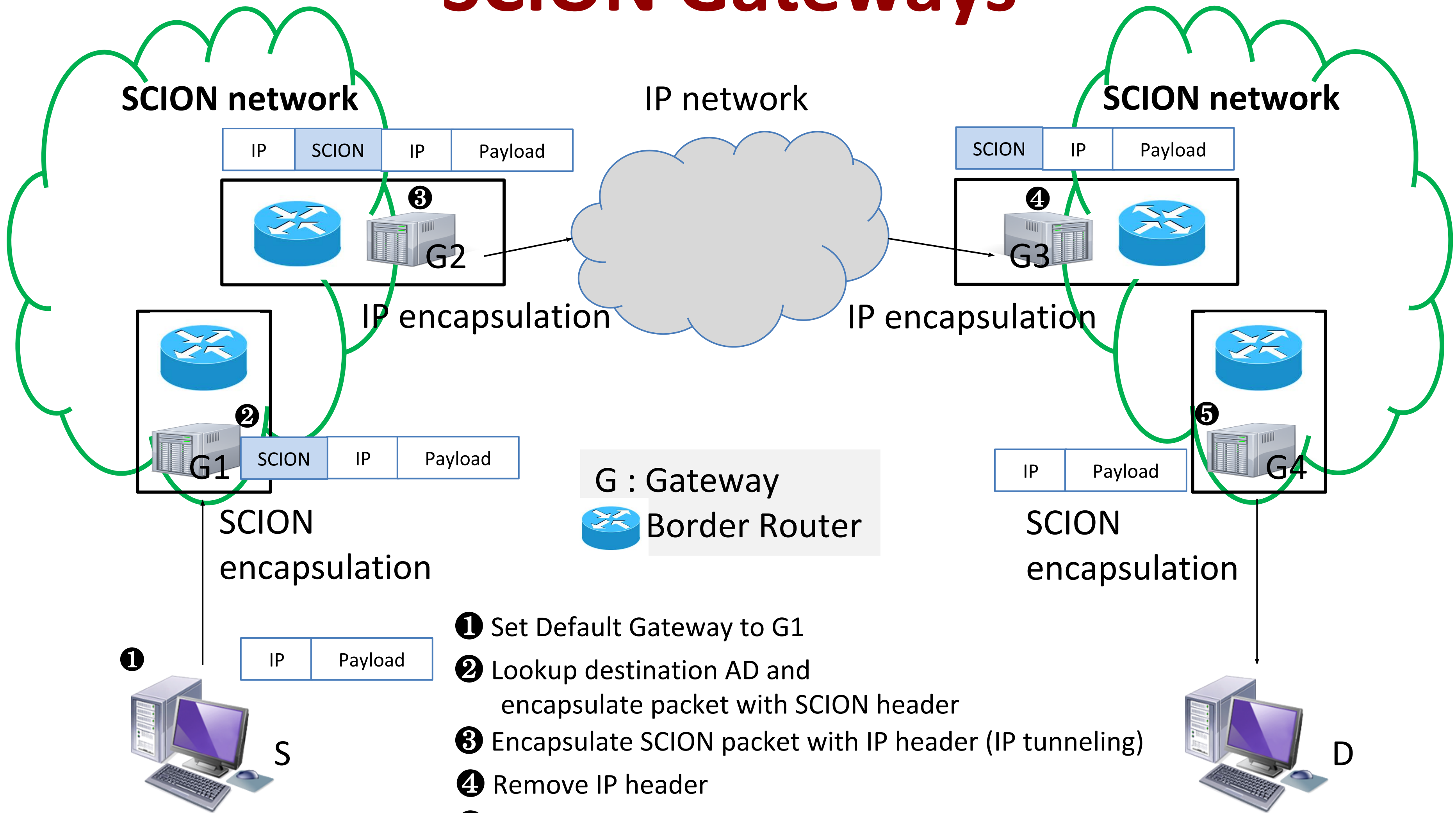
SCION Packet Format (Data)

	0-7	8-15	16-23	24-31	32-39	40-47	48-55	56-63
Common Header	Type	HDR Len	Total Len		TS*	Src Len	Dst Len	Flag
	Curr OF*	# OF	L4 Proto	nRetCAP	CAP Req Info	New CAP*	Path Val*	Src Auth*
	Source Address (variable size)							
	Destination Address (variable size)							
Special OF	Info	Timestamp				TDID		reserved
Regular OFs *for **up-path** forwarding*	Opaque Field (0)							
	...							
Special OF	Info	Timestamp				TDID		reserved
Regular OFs *for **down-path** forwarding*	Opaque Field (0)							
	...							
Return Capabilities	Timestamp				CAP*	Ret CAP		
	Source Validation (variable size)							
	Path Validation (variable size)							
New Capabilities	Timestamp				CAP*	New CAP		

Incremental Deployment

- Current ISP topologies consistent with SCION TDs
- Minor changes for ISPs
  - SCION edge router deployment
  - Beacon / certificate / path server deployment (inexpensive commodity hardware)
  - Regular MPLS forwarding internally
  - IP tunnels connect SCION edge routers in different ADs
- Minor changes in end-domains
  - IP routing used for basic connectivity
  - SCION gateway enables legacy end hosts to benefit from SCION network

SCION Gateways



➊ Set Default Gateway to G1

➋ Lookup destination AD and encapsulate packet with SCION header

➌ Encapsulate SCION packet with IP header (IP tunneling)

➍ Remove IP header

➎ Remove SCION header

* Destination sets the default gateway to G4 in order to use SCION network

Part 2: STRIDE

Wishlist for DDoS Resilience

- Multi-path routing and re-route-ability
  - SCION provides k^2 end-to-end path diversity
- Real-time congestion information
  - SCION symmetric paths support the use of capabilities
  - Capabilities can carry real-time congestion information
- Precise bandwidth guarantees
  - SCION top-down topology discovery enables tree-based bandwidth allocation

STRIDE

Sanctuary Trail: Rescue from Internet DDoS Entrapment

- Precise Bandwidth Guarantees
  - Connection setup guarantees
  - Flow bandwidth guarantees
- Flexible Route Control
  - Inbound/outbound path control
  - Selective path disclosure; e.g., public/private paths
- Robustness and Efficiency
  - Separation of control and data plane
  - Efficient router operation; e.g., packet forwarding

STRIDE Bandwidth Spectrum

- Bandwidth Classes
  - Static class
    - Long-term bandwidth guarantees
    - Connection setup in capability protocols
    - 10-15% of link bandwidth
  - Dynamic class
    - Short-term bandwidth guarantees
    - Per-flow bandwidth allocation and guarantees
    - 60% of link bandwidth
  - Best-effort class
    - No bandwidth guarantees
    - Basic class given correct SCION path
    - 25-30% of link bandwidth

STRIDE

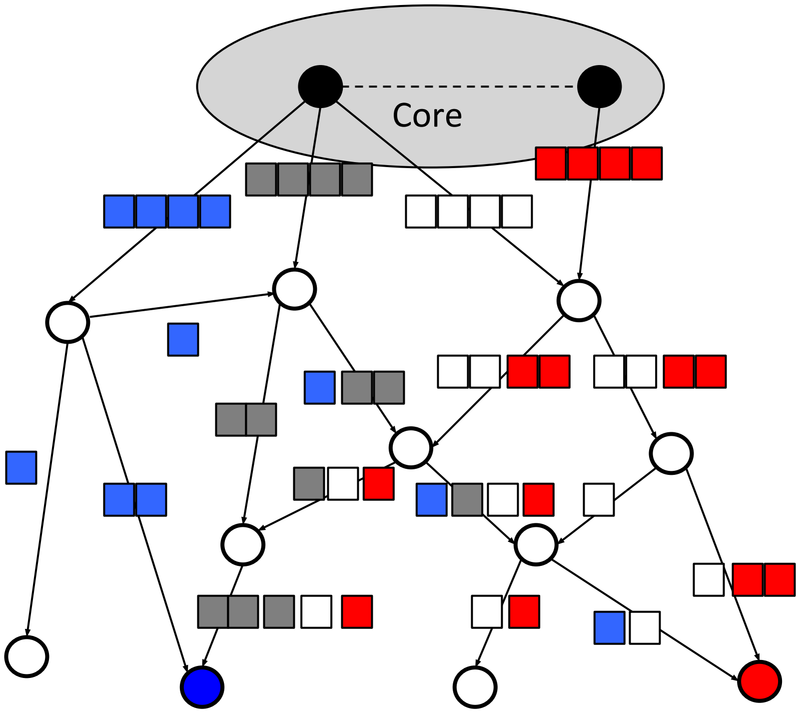
- Bandwidth guarantees for half-paths
  - When up- and down-paths are constructed, SCION provides bandwidth guarantees for these paths
  - Use a three-way handshake approach
  - Bandwidth guaranteed half-path on **Static Class**
- End-to-end bandwidth guaranteed paths
  - Destinations control which paths to disclose to which sources
  - Source selects which paths to send traffic
  - **Static Channel**: bandwidth guaranteed end-to-end path on **Static Class**
- End-to-end flow bandwidth guarantees
  - STRIDE allocates bandwidth to capabilities to protect end-to-end communication
  - **Dynamic Channel:** bandwidth guaranteed end-to-end flow on **Dynamic Class**

Bandwidth Guarantees for Half Paths

- Step 1: Path **announcement** annotated with ***bandwidth***
  - Provider splits upstream bandwidth to customers
  - Bandwidth overbooking for maximal bandwidth utilization
  - Endpoint ADs obtain bandwidth guarantees for half paths
- Step 2: **Activation** and allocation
  - Endpoint ADs activate selected paths
  - Providers handle overbooking
- Step 3: **Confirmation**
  - Static half-path

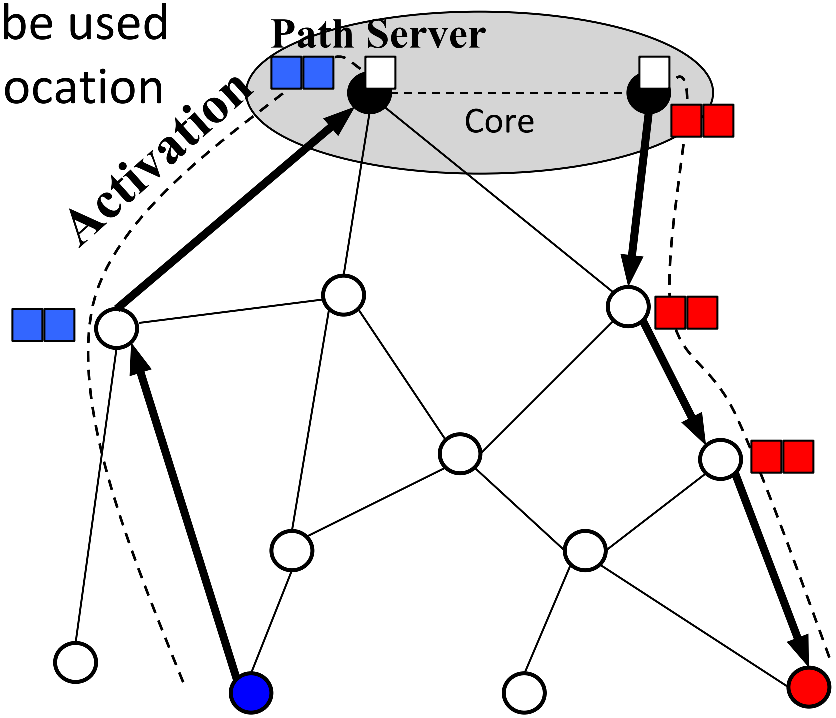
Step 1: Bandwidth Allocation

- Provider splits bandwidth along PCBs
  - Lower-bound guarantee
  - PCBs contain temporary opaque fields
    - Switched to long-term ones after activation



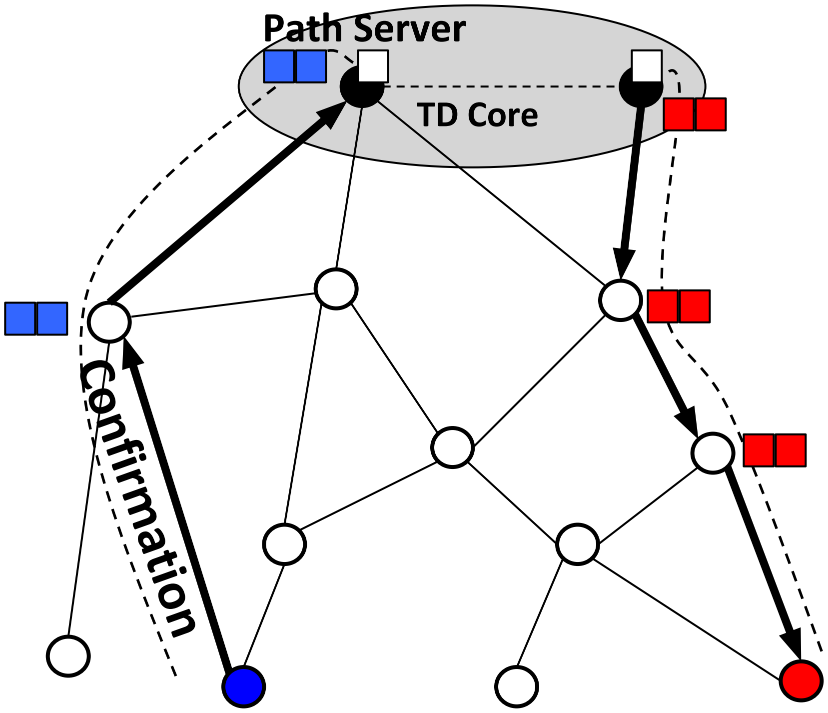
Step 2: Path Activation/Bandwidth Allocation

- Increase utilization / prevent path misuse
  Provider ADs
  - Overbook bandwidth
    - Not all announced paths would be used
    - Increase per-path bandwidth allocation
  - Prevent bandwidth overuse
    - Malicious nodes cannot use an excessive number of paths



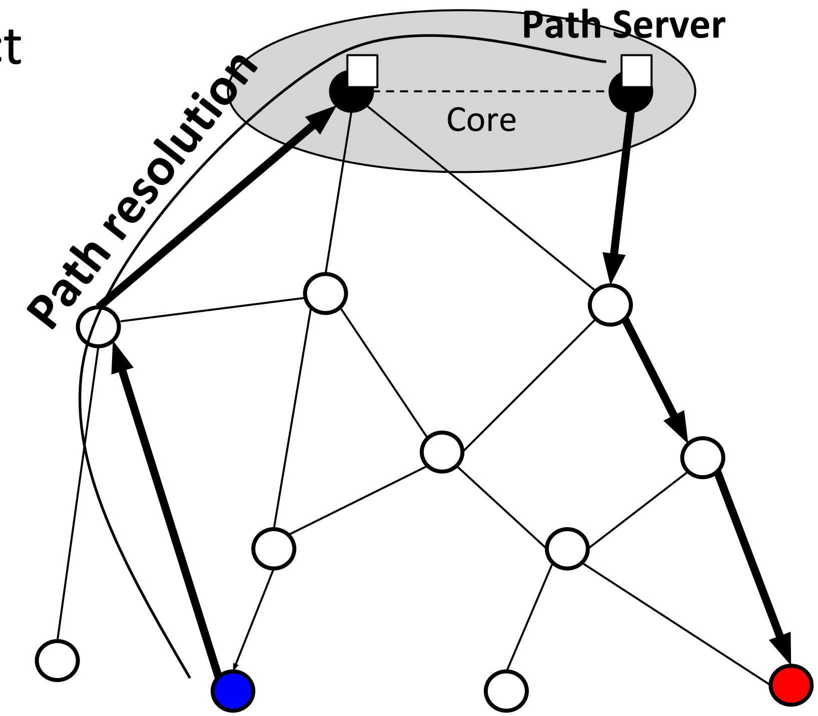
Step 3: Confirmation

- If path activation is accepted:
  - TD Core issues a confirmation message along that path
  - ADs in the path embed the long-term, activated opaque fields
- Path activation may be rejected if:
  - Endpoint AD activated more than k paths
  - All announced paths are activated at a provider AD
  - Path is not compliant with a provider AD's routing policies



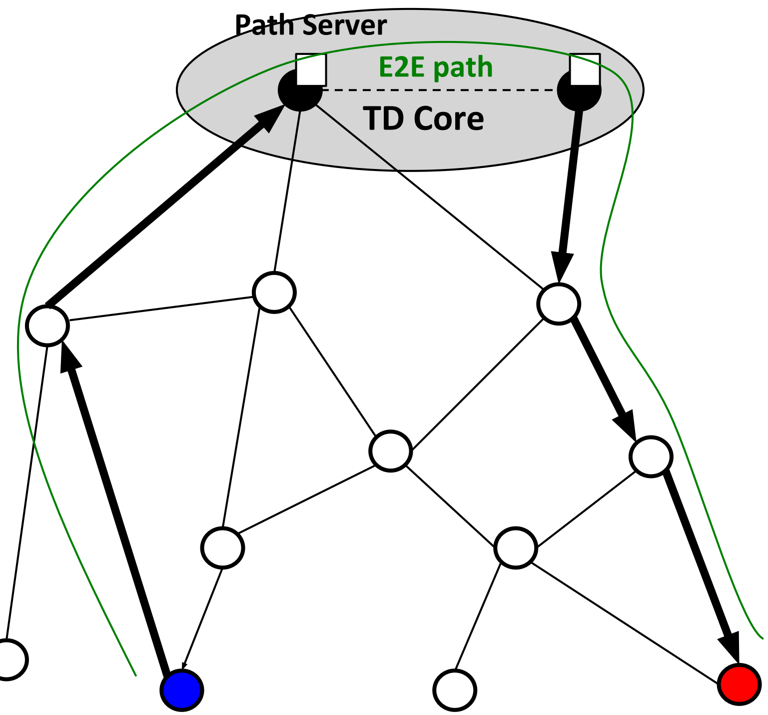
End-to-End Static Channel

- AD to Path Resolution
  - Source nodes obtain the down-paths annotated with bandwidth guarantees
  - Source follows preferences to select down-path



End-to-end Static Channel

- Source AD combines static up-path and static down-path
  - Resulting path turns into **Static Channel**
- Initial packet (capability request)
  - Forwarded through a **Static or Best-Effort Channel**
  - Transit ADs provide real-time bandwidth allocation to capabilities
  - Transit ADs set congestion bit to capabilities
  - **Static or BE Ch → Dynamic Channel**
- Following packets
  - Capability-carrying packets use **Dynamic Channel**
  - Capabilities are periodically renewed (i.e., short-term bandwidth guarantees)
  - Transit ADs perform stateless accounting



End-to-end Dynamic Channel

- End-to-end dynamic channel can be derived from:
  - Static Channel
  - Static up-path + Best-effort down-path
  - Best-effort up-path + Static down-path
  - Best-effort channel

- Private path can be used to circumvent congested paths

  Options
  - Set path disclosure policy: provide paths only to designated source nodes
  - Register encrypted paths and disclose keys to selected source nodes
  - Use unregistered Best-Effort private channel

Enforcement

- Each bandwidth class has defined bw subclasses
  - E.g., 512kbps, 1Mbps, 2Mbps, …
- Each capability is associated with a given subclass
- Each AD on path and receiver approve bw requested, or indicate max bw available
- Each AD uses “Elephant detector” within each bw subclass
  - Fixed state required for each Elephant detector
  - No per-flow state!
  - Blacklist with flows that overstepped their allocation determines preferential packet drop
- Per-flow state only required for flow admission, initial capability

Traffic Priority of Capability Requests

- Priority of Capability Request packet on best-effort downpath
  - If static downpath available, packet arrival is guaranteed

Priority	Up-path	Bits set	Notes
1	Static	-	Uppath used within allocated bw
2	Best-effort	-	Best-effort is preferred if it is sent over uncongested links
3	Static	Overuse	Uppath used beyond allocated bw
4	Best-effort	Congestion	Experienced congestion on uppath
5	Outside TD	-	Outside TD is de-prioritized

Part 3: SCION over XIA (XION)
Future Work

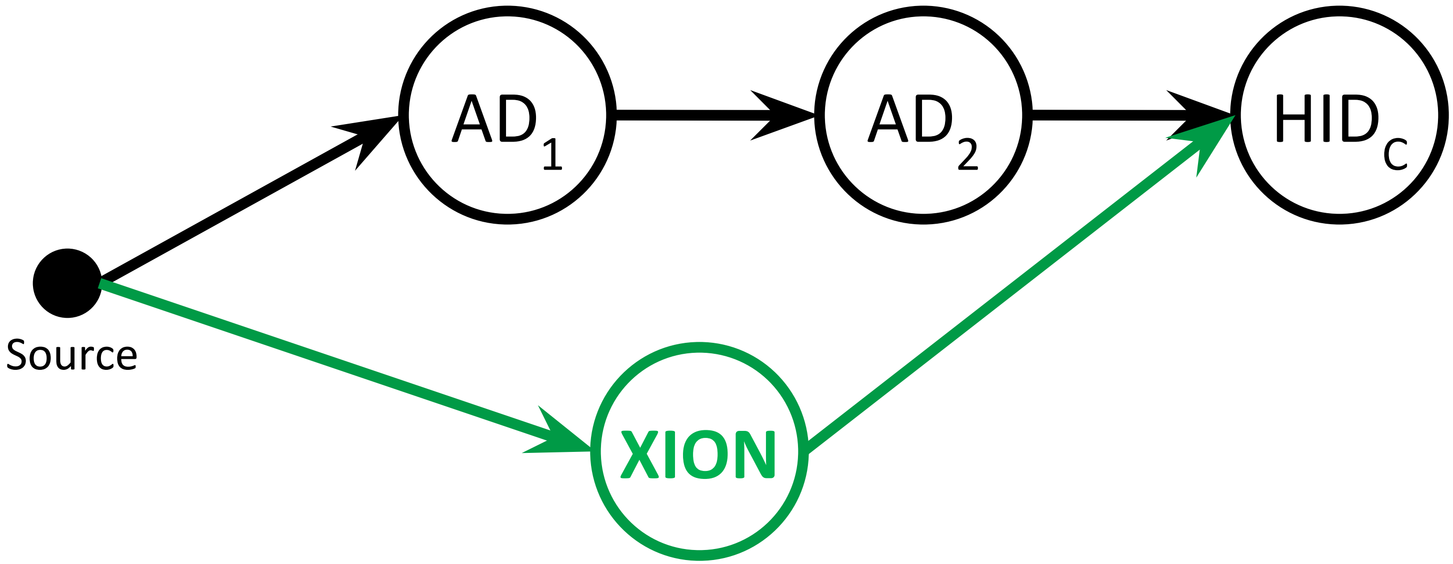
XIA+SCION=XION

- SCION offers interdomain routing and forwarding for XIA $\longrightarrow$ XION
- XION is a new principal type, which offers forwarding along a path segment in a DAG

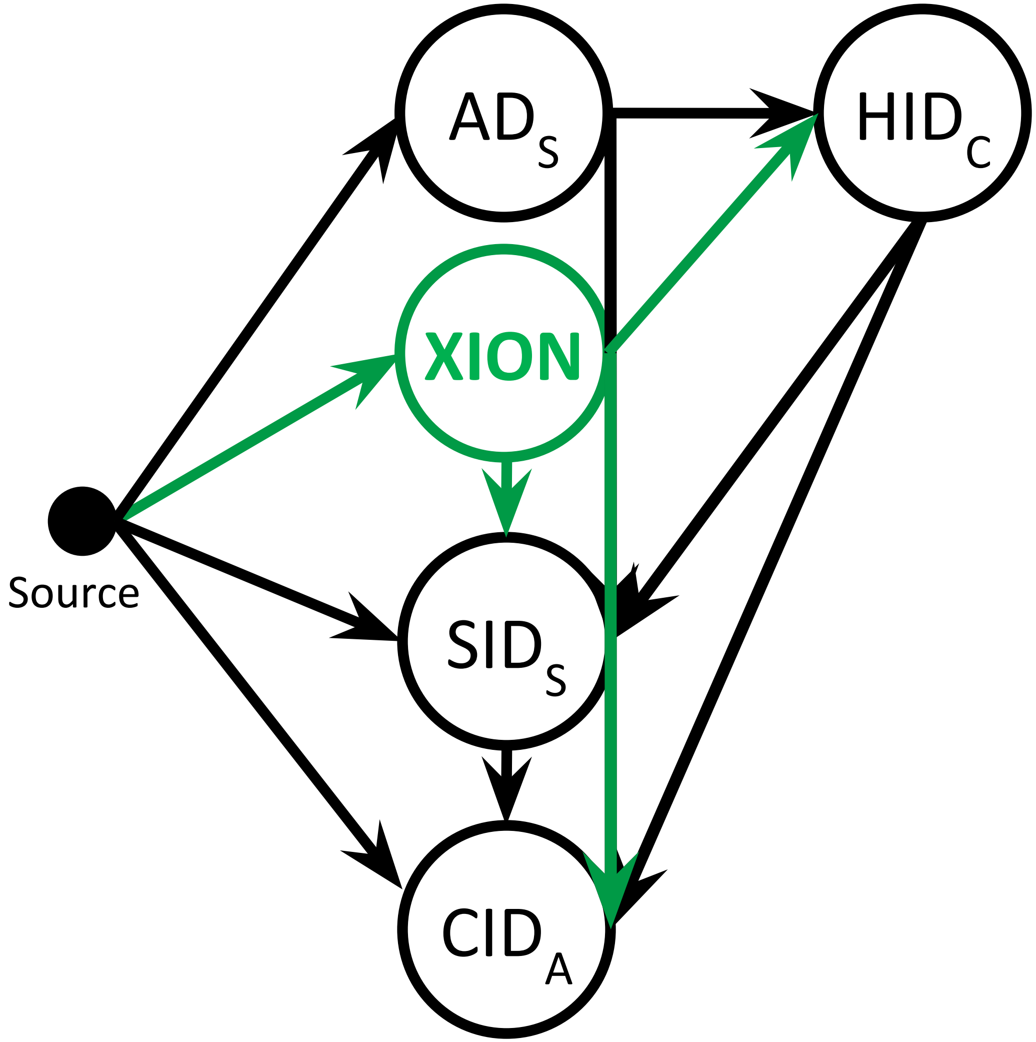
XION as Alternate Path

If Source needs ***Explicit Trust* on the Path (Network)** to HID

SCOPING in XIA terminology?

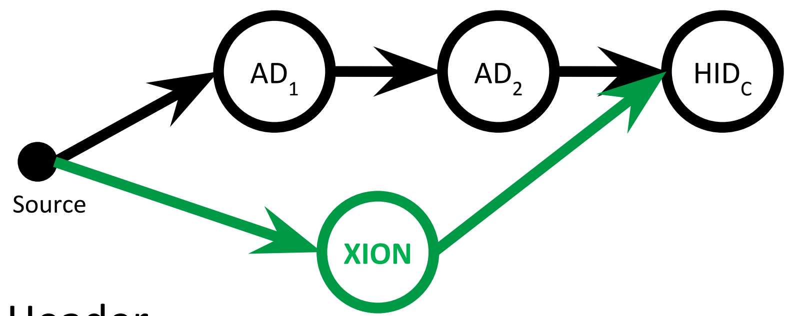


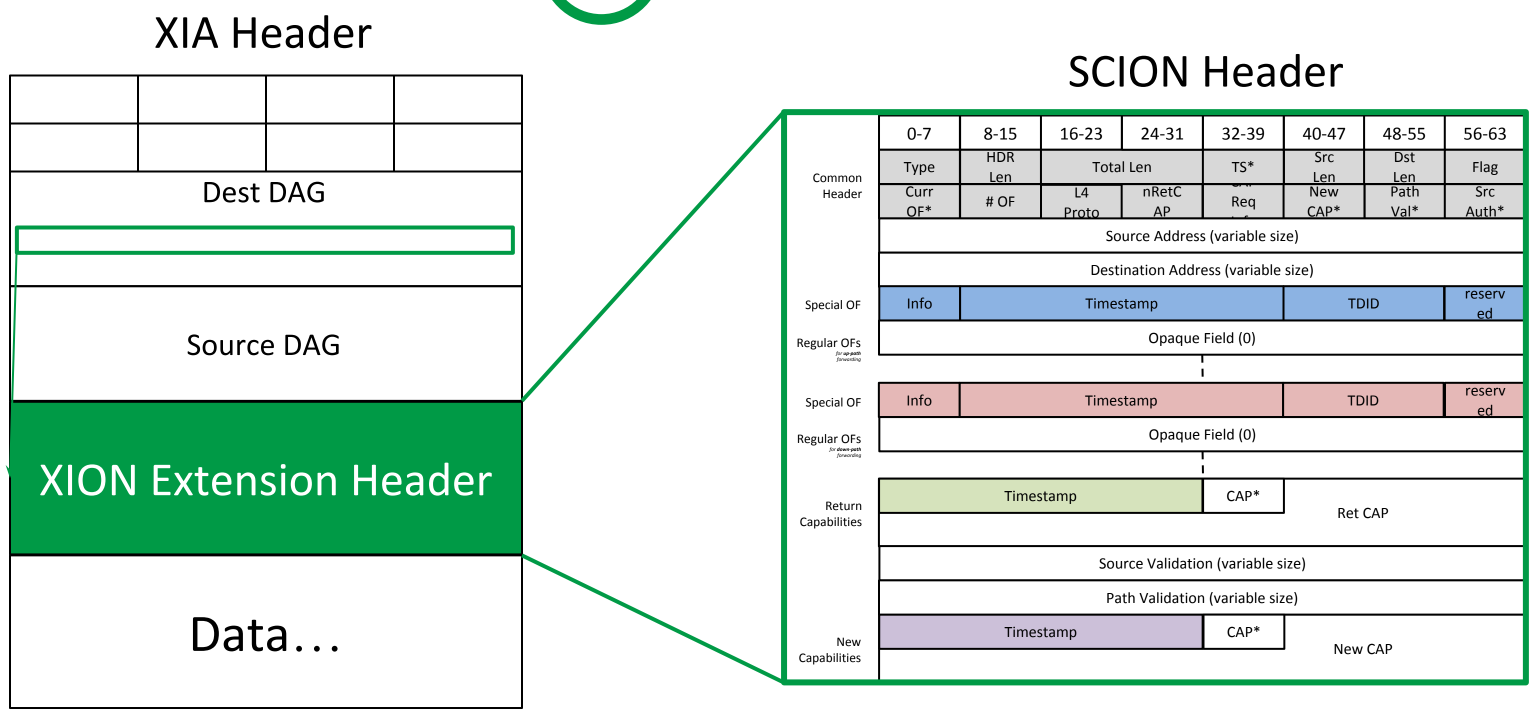
XION as Alternate Path



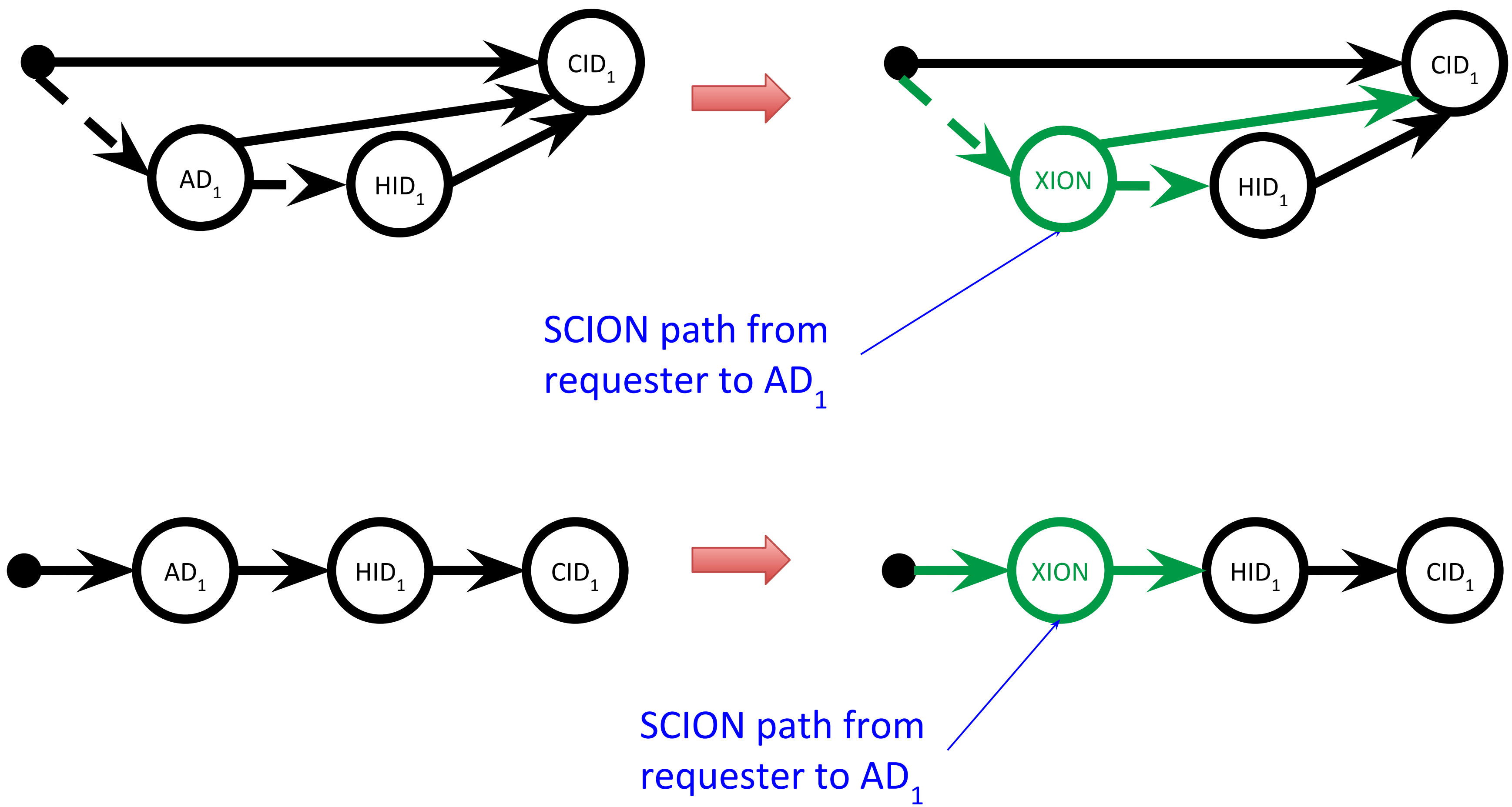
SCION + XIA Breakout Notes

XION Principal Type:
Point to SCION Header in XIA Extension Hdr

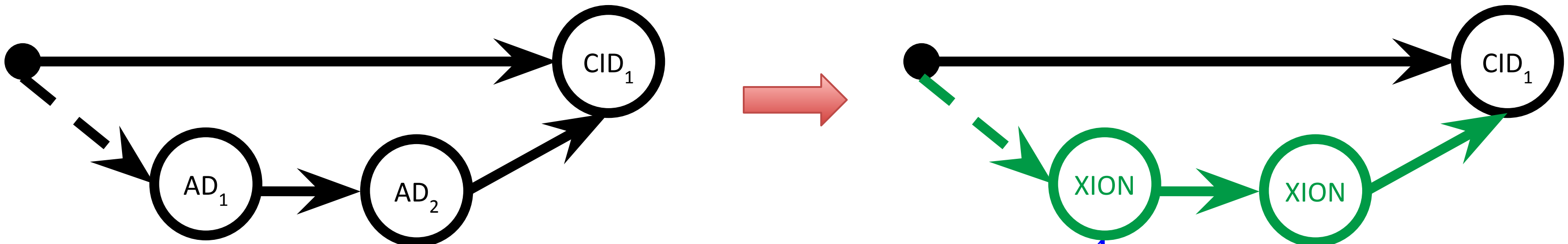




DAGs using XION paths can match pure XIA DAG semantics, e.g.:



Complex DAGs might require multiple XION nodes



AD$_1$ and intermediate ADs from AD$_1$ to AD$_2$ may not serve the content, but ADs before AD$_1$ may.

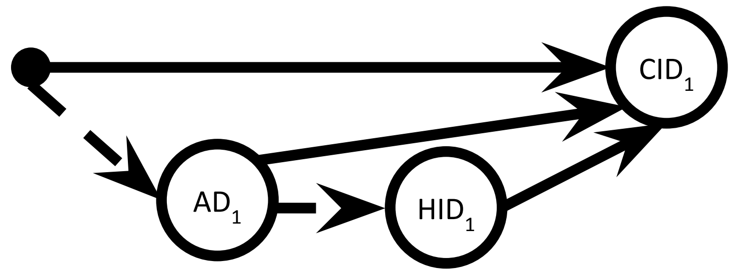
SCION path from requester to AD$_1$

SCION path from AD$_1$ to AD$_2$

Multiple SCION headers in one packet means lots of overhead!
Hopefully this is rare.

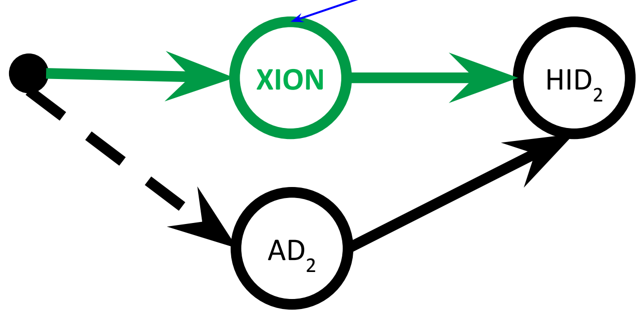
Fill in XION Path for Fast Forwarding on Return Path

Destination



Initially empty; filled in by SCION-aware routers as request packet moves toward content.

Source



Summary

SCION Advantages

- Security
  - Isolation of data plane from control plane
    - Data plane still usable even if control plane disrupted
  - Cryptographic validation of forwarding information
  - Trust agility: local & selectable roots of trust (no global root of trust)
  - Avoidance of BGP / IP attacks (blackhole, wormhole, etc.)
  - No single point of failure
  - Explicit trust for packet forwarding, small Trusted Computing Base (TCB)
- Reliability
  - Isolation between mutually untrusted network domains
  - Multi-path forwarding, dozens of potential paths available
  - ISP / sender / receiver controllable paths
  - Instant convergence of routing protocol
  - No route-flap dampening necessary
- Efficiency
  - Scalability: routing overhead independent on # of destinations

SCION Disadvantages

- New protocols, new equipment
- Packet header larger than IP
- Static path binding
  - No automated route failure recovery

SCION Security Benefits

	S-BGP + DNSSec	SCION
Isolation	**No** collusion/wormhole attacks poor path freshness path replay attacks single root of trust	**Yes** no cross-TD attacks path freshness scalability no single root of trust
TCB	**The whole Internet**	**TD Core and on-path ADs**
Path Control	**Too little (dst) or too much (src),** empowering DDoS attacks	**Balanced control** enabling DDoS defenses

SCION Stakeholder Pros and Cons

- Manufacturers
  - Sale of additional equipment
  - Commoditization: routers become simple and inexpensive
- ISPs
  - New revenue streams through service differentiation
  - High-availability service offerings, powerful DDoS defenses
  - Resilient to attacks and configuration errors
  - Incremental update, only new edge routers needed, inexpensive routers
  - New equipment, new protocols
- Consumers
  - High reliability and availability
  - Differentiated services, path choice, trading off quality and price
  - Trust agility
  - Software / HW upgrade
- Government

SCION Summary

- Basic architecture design for a next-generation network that emphasizes isolation, control and explicit trust
- Highly efficient, scalable, available architecture
- Enables numerous additional security mechanisms, e.g., network capabilities, DoS defenses