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

Xin Zhang, Hsu-Chun Hsiao, Geoff Hasker,
Haowen Chan, Adrian Perrig, David Andersen

Reasons for Clean-Slate Design

- Someone may just want to deploy a new Internet ☺
  - ✓ Possible for specialized high-reliability networks, e.g., smart grid
  - ✓ We need to have a design ready
- Even if we want to evolve current Internet, we need to have a goal, know how good a network could be

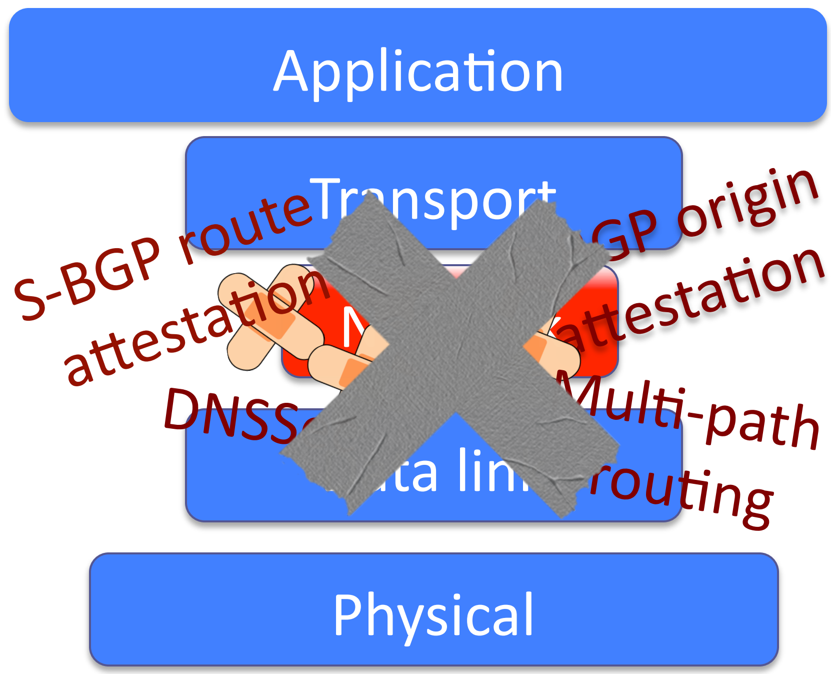
The question is not: why deploy a new Internet?
But: why are we still putting up with the current Internet?

After years of patching, the Internet is still neither Reliable nor Secure!

Feb 2008: Pakistani ISP hijacks YouTube prefix

Apr 2010: A Chinese ISP inserts fake routes affecting thousands of US networks.

Nov 2010: 10% of Internet traffic 'hijacked' to Chinese servers due to DNS Tampering.

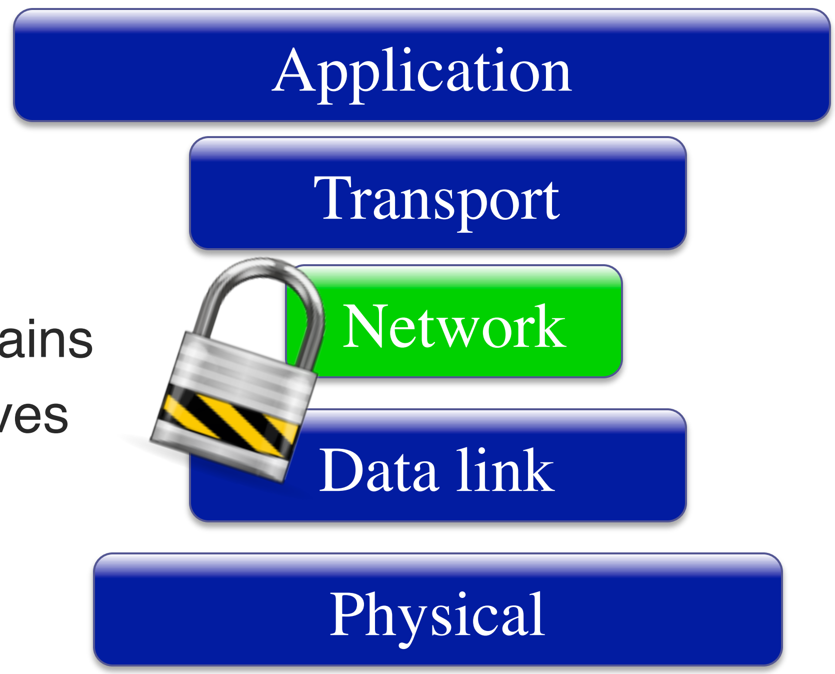


- Fixes to date – ad hoc, patches
- Inconvenient truths
  - S-BGP: delayed convergence
  - Global PKI: single root of trust

Reliable Network Layer Wishlist

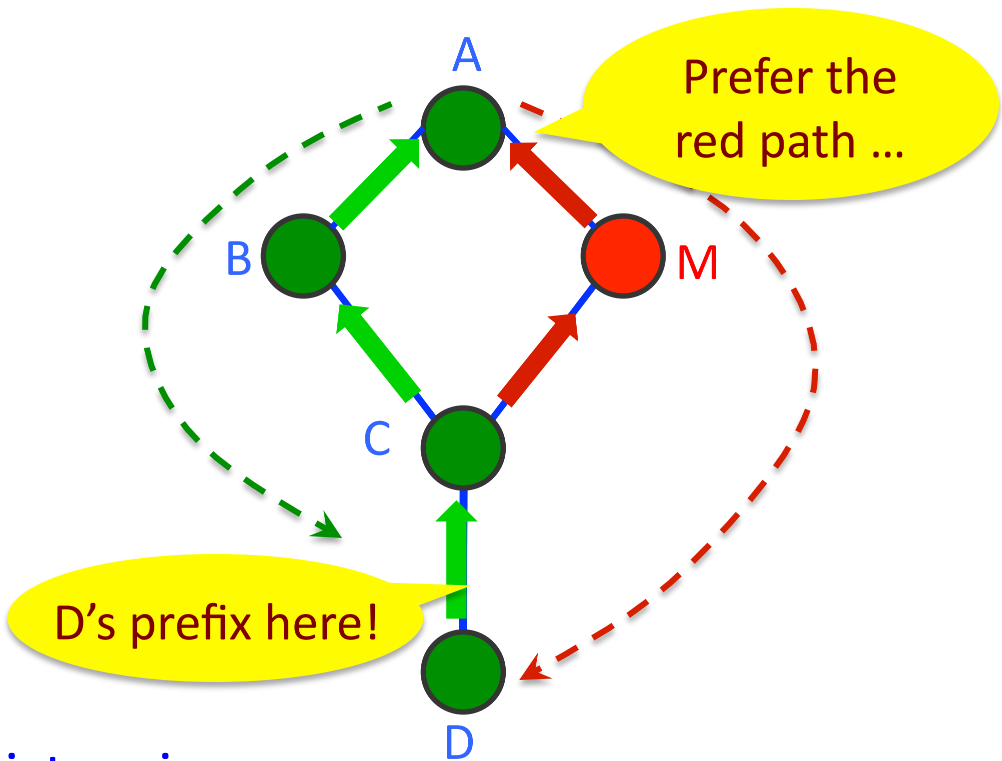
Imagine if we had:

- Explicit understanding whom you must trust for network operations
- All relevant parties have balanced control over path selection
- The architecture isolates attacks to domains with common laws and economic incentives
- No single global root of trust
- High efficiency, scalability, flexibility



Limitations of the Current Internet

- Destination or ISP have no control over inbound paths



- Route inconsistencies
  - Forwarding state may be different from announced state

Limitations of the Current Internet (cont'd)

- Lack of routing isolation
  - A failure/attack can have global effects
  - Global visibility of paths is not scalable
- Slow convergence / route oscillation
- Large routing tables
  - Multi-homing / flat namespaces prevent aggregation
- Lack of route freshness
  - Current (S-)BGP cannot prevent replay of old paths

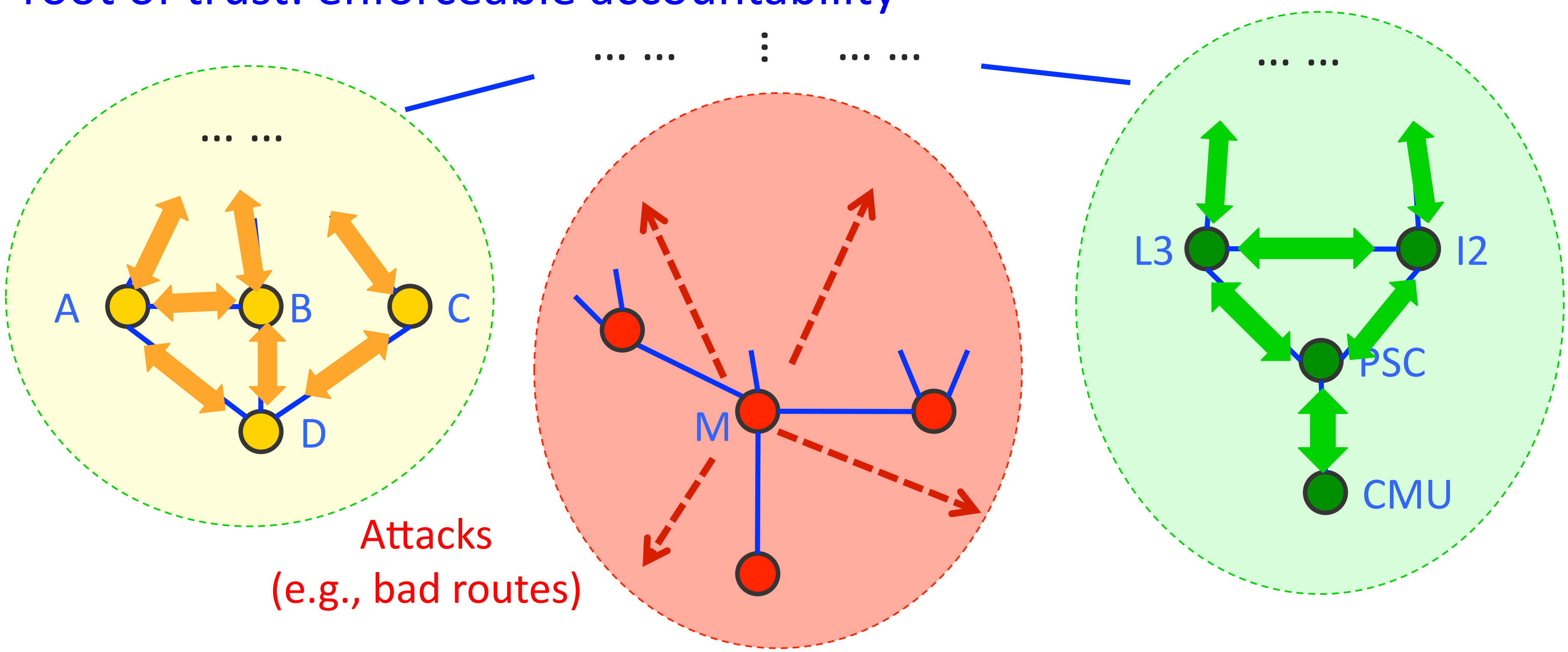
Note that these issues are fundamental to (S)-BGP!

Wish List (1): Isolation

❖ Isolation of attacks

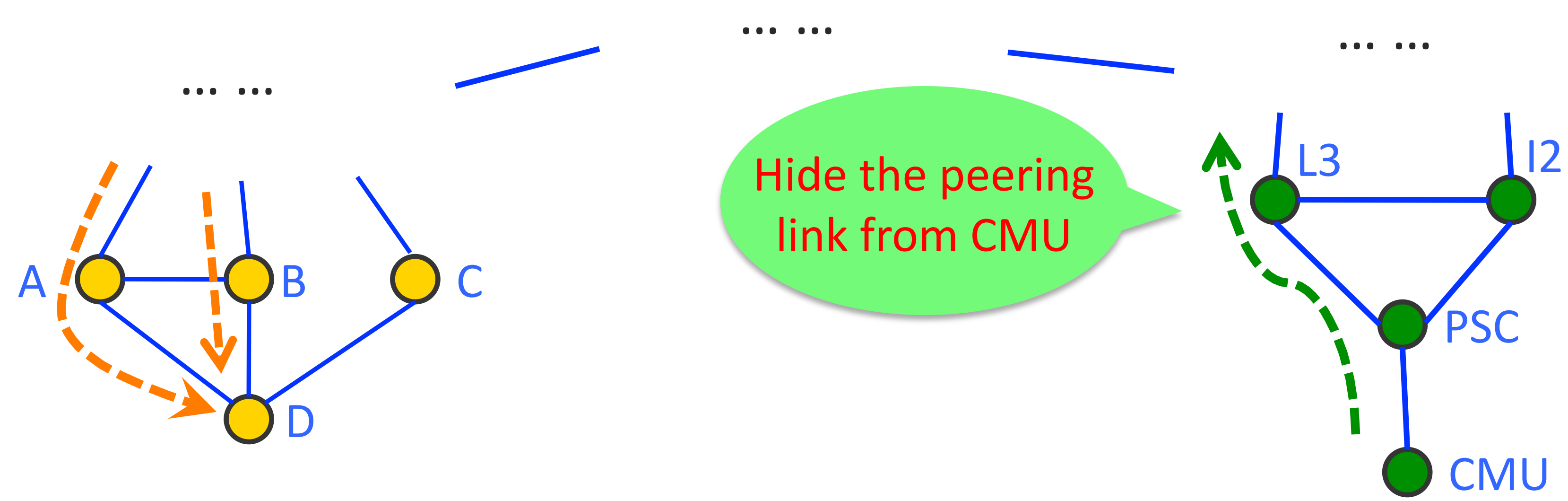
❖ Scalable and reliable routing updates

❖ Operate with mutually distrusting entities without a global single root of trust: enforceable accountability



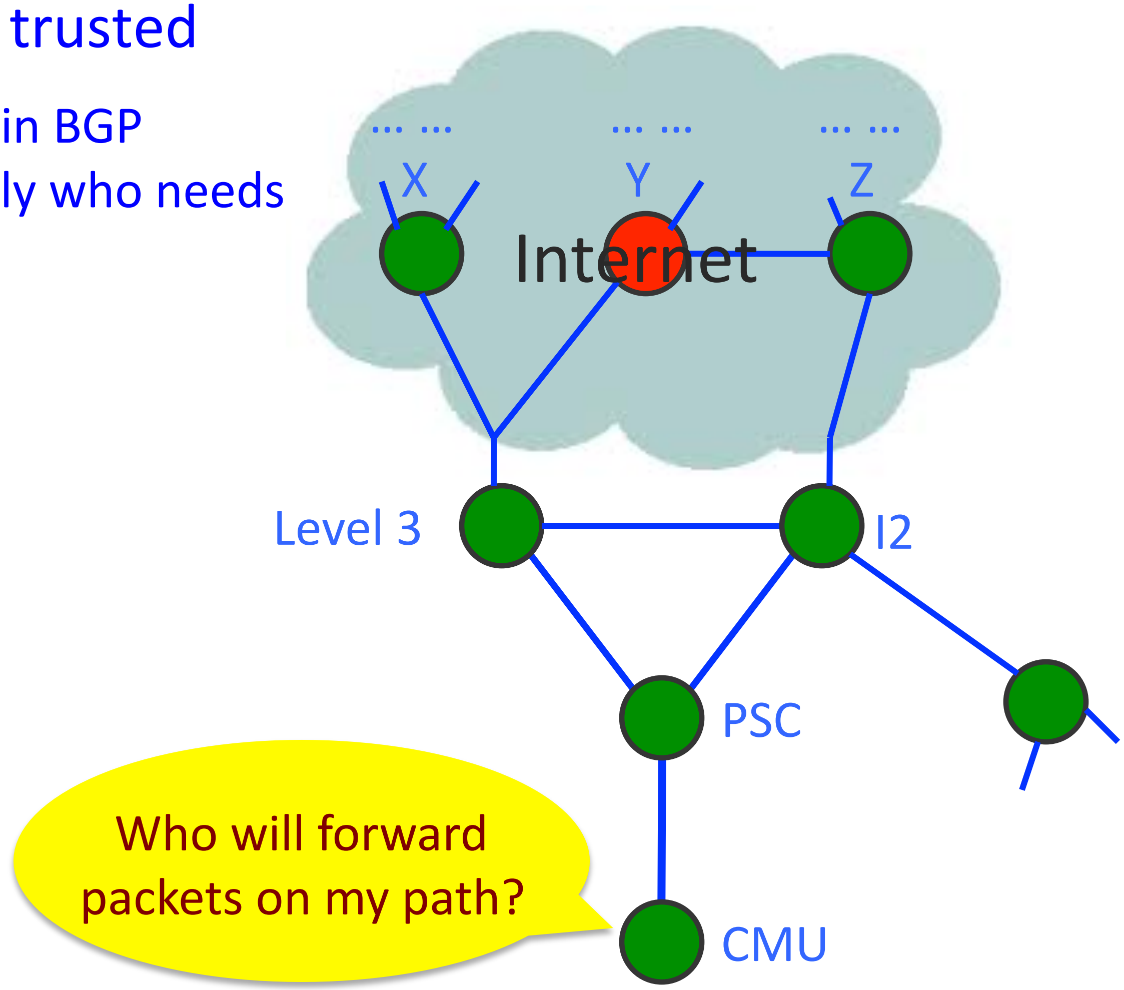
Wish List (2): Balanced Control

❖ Transit ISPs, source and destination all need path control



Wish List (3): Explicit Trust

- Know who needs to be trusted
  - Absence of consistency in BGP prevents knowing exactly who needs to be trusted

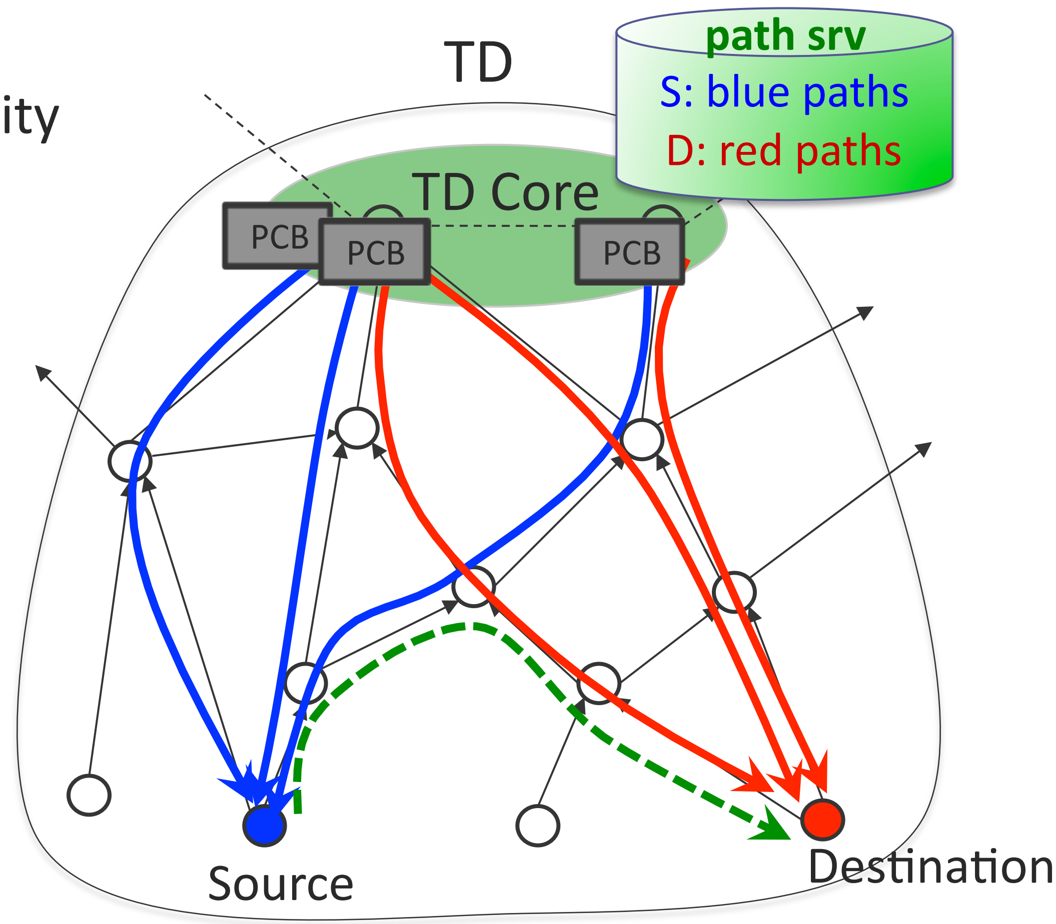


SCION Architectural Goals

- High availability, even for networks with malicious parties
- Explicit trust for network operations
- Minimal TCB: limit number of entities that need to be trusted for any operation
  - Strong isolation from untrusted parties
- Operate with mutually distrusting entities
  - No single root of trust
- Enable route control for ISPs, receivers, senders
- Simplicity, efficiency, flexibility, and scalability

SCION Architecture Overview

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

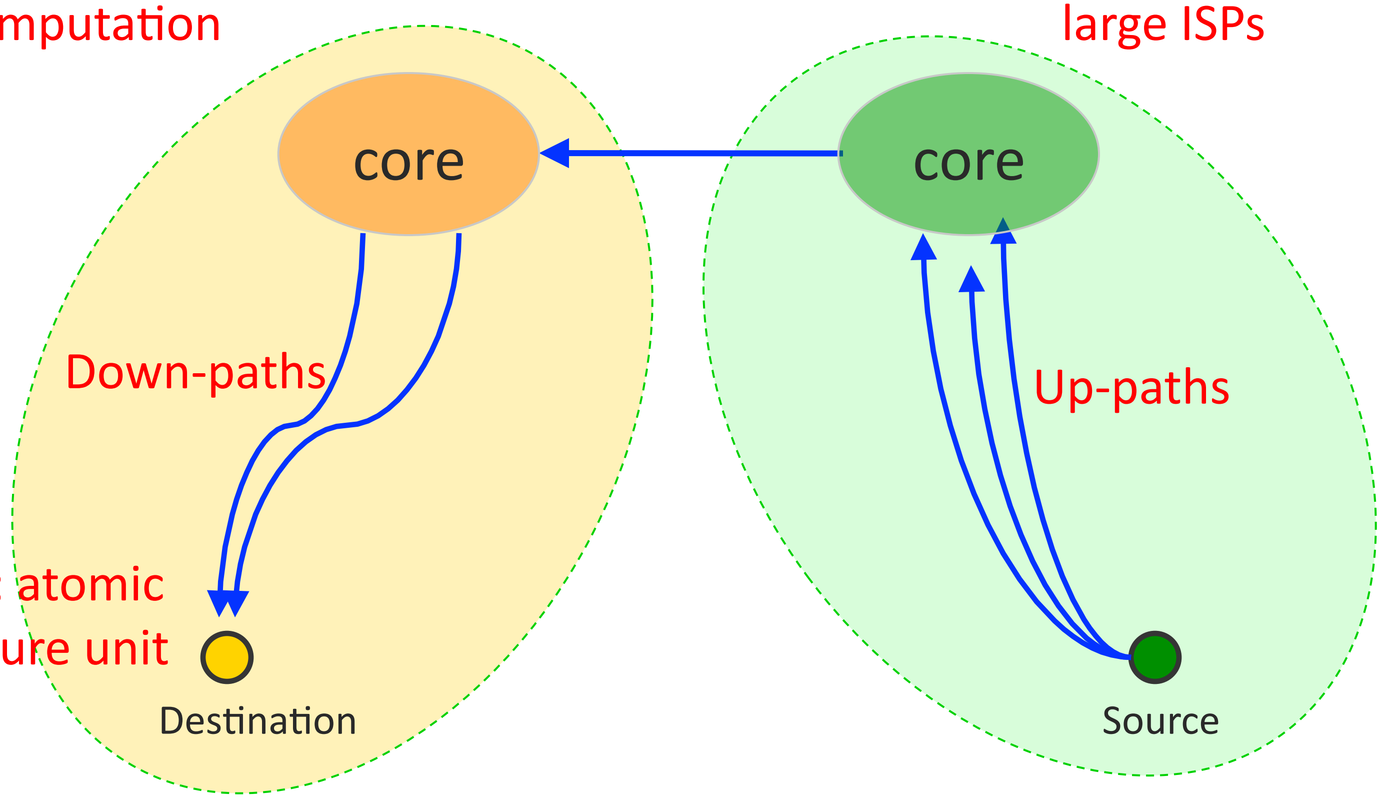


Hierarchical Decomposition

- Global set of TD (Trust Domains)
  - ✓ Map to geographic, political, legal boundaries
- 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 is atomic failure unit, contiguous/autonomous domain
  - ✓ Transit AD or endpoint AD

Hierarchical Decomposition

- Split the network into a set of trust domains (TD)



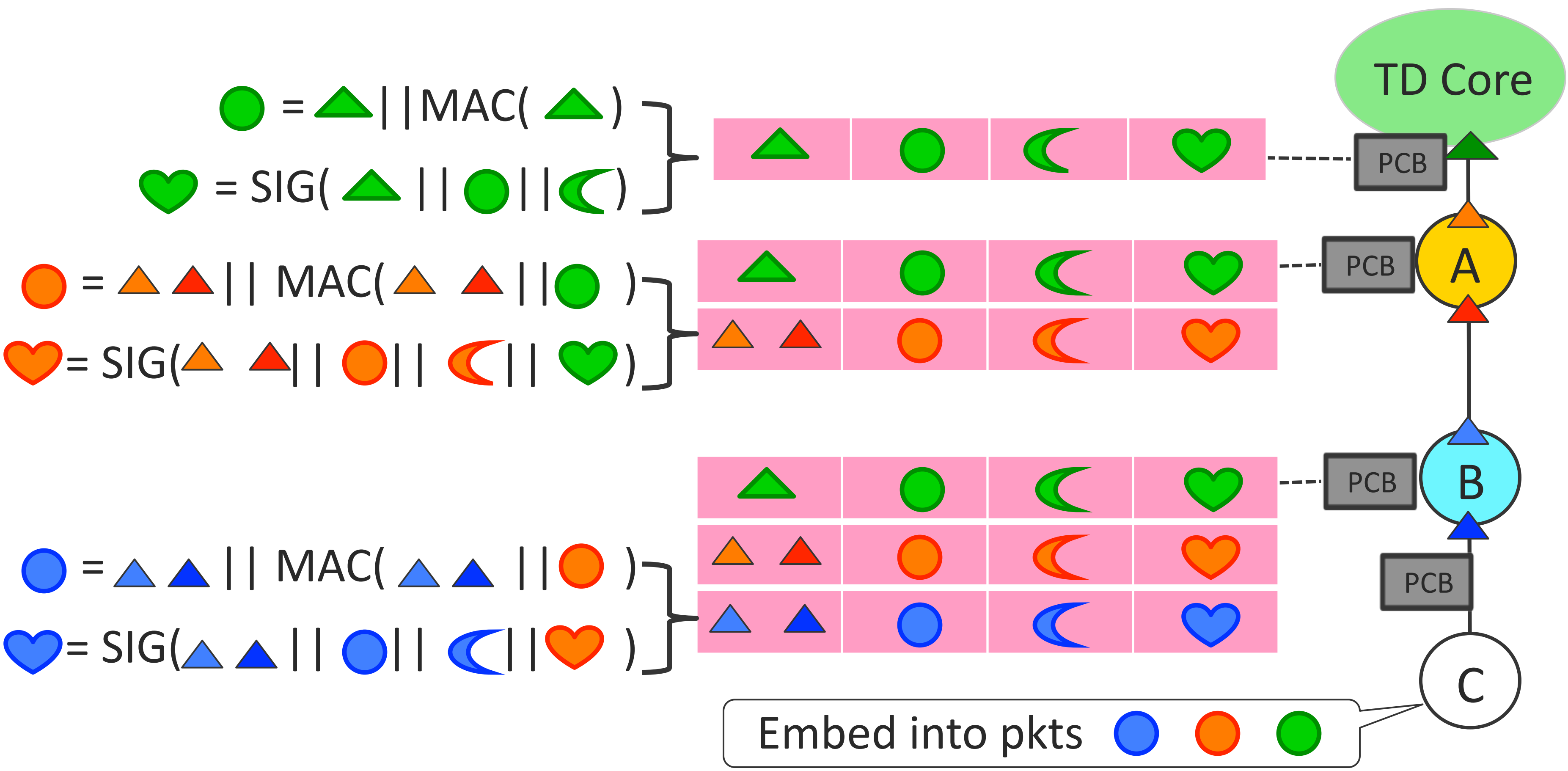
Path Construction

Goal: each endpoint learns multiple verifiable paths to its core

- Discovering paths via Path Construction Beacons (PCBs)
  - ✓ TD Core periodically initiates PCBs
  - ✓ Providers advertise upstream topology to peering and customer ADs
- ADs perform the following operations
  - ✓ Collect PCBs
  - ✓ For each neighbor AD, select which *k* PCBs to forward
  - ✓ Update cryptographic information in PCBs
- Endpoint AD will receive up to *k* PCBs from each upstream AD, and select *k* down-paths and up-paths

Path Construction

: interface : Opaque field : expiration time : signature



Path Construction

Interfaces: I(i) = previous-hop interfaces || local interfaces

Opaque field: O(i) = local interfaces || MAC over local interfaces and O(i-1)

Signature: Σ(i) = sign over I(i), T(i), O(i), and Σ(i-1), with cert of pub key

TC→A: $I(TC)$: ϕ ||{ϕ, $TC1$}

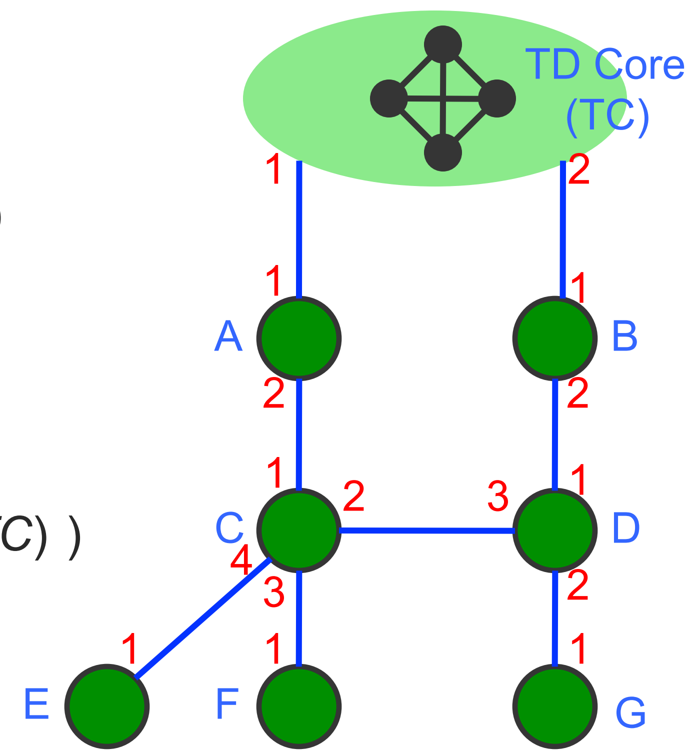
$O(TC)$: {ϕ, $TC1$} ||MAC_{Ktc}({ϕ, $TC1$} || ϕ)

Σ(TC): Sign($I(TC)$ || $T(TC)$ || $O(TC)$ || ϕ)

A→C: $I(A)$: $I(TC)$|| {$A1$, $A2$}

$O(A)$: {$A1$, $A2$} || MAC_{Ka}({$A1$, $A2$} ||$O(TC)$)

Σ(A): Sign($I(A)$ || $T(A)$ || $O(A)$ ||Σ(TC))



Path Construction

Interfaces: I(i) = previous-hop interfaces || local interfaces

Opaque field: O(i) = local interfaces || MAC over local interfaces and O(i-1)

Signature: Σ(i) = sign over I(i), T(i), O(i), and Σ(i-1), with cert of pub key

C? – One PCB per neighbor

C→E: $I(C)$: $I(A)$|| {$C1$, $C4$}

$O(C)$: {$C1$, $C4$} || MAC_{Ka}({$C1$, $C4$} || $O(A)$)

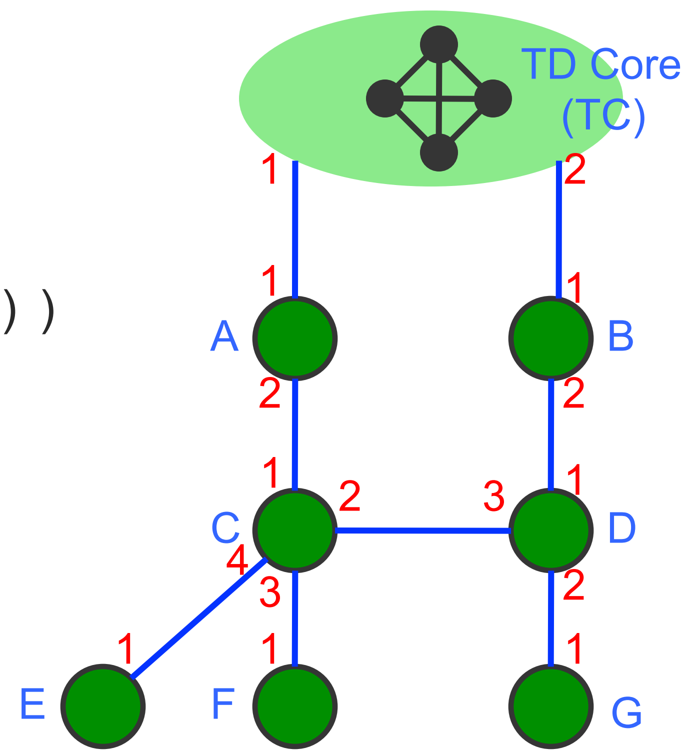
Σ(C): Sign($I(C)$ || $T(C)$ || $O(C)$ ||Σ(A))

Also include peering link!

$I_{C,D}$(C): {$C4$,$C2$} || TD || AID_D

$O_{C,D}$(C): {$C4$, $C2$} ||MAC_{Kc}({$C4$, $C2$})

$\Sigma_{C,D}$(C): Sign($I_{C,D}$(C) || $T_{C,D}$(C) || $O_{C,D}$(C) || $O(C)$)

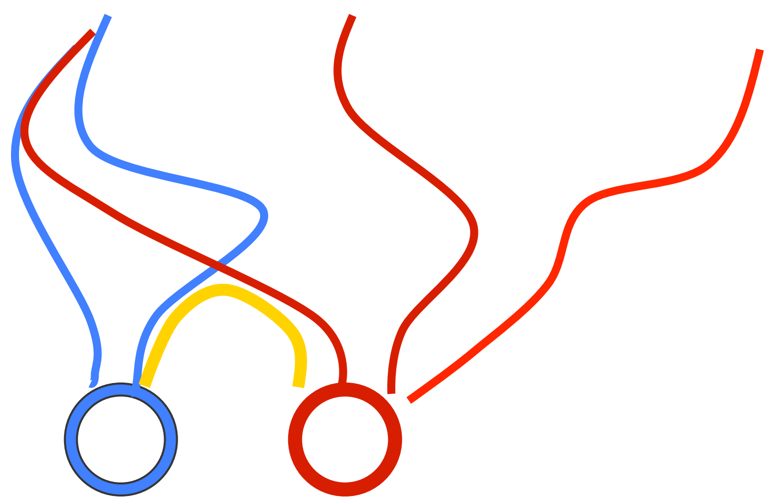


Address/Path Resolution

- TD core provides address/path resolution servers
- Each endpoint is identified as an AID:EID pair. AID is signed by the containing TD, and EID is signed by the containing AD (with AID).
  - ✓ Address is a public key [AIP 2008]
- Each AD registers name / address at address resolution server, uses an up-path to reach TD core
  - ✓ Private key used to sign name→address mapping
- ADs select which down-paths to announce
- ADs sign down-paths with private key and register down-paths with path resolution servers

Route Joining

- Local traffic should not need to traverse TD core
- Sender obtains receiver's *k* down-paths
- Sender intersects its up-paths with receiver's down-paths
- Sender selects preferred routes based on k^2 options



Forwarding

- Down-path contains all forwarding decisions (AD traversed) from endpoint AD to TD core
  - ✓ Ingress/egress points for each AD, authenticated in opaque fields
  - ✓ ADs use internal routing to send traffic from ingress to egress point
- Joined end-to-end route contains full forwarding information from source to destination
  - ✓ No routing / forwarding tables needed!

Discussion

- Incremental Deployment
  - ✓ Current ISP topologies are consistent with the TDs in SCION
  - ✓ ISPs use MPLS to forward traffic within their networks
  - ✓ Only edge routers need to deploy SCION
  - ✓ Can use IP tunnels to connect SCION edge routers in different ADs

- Limitations
  - ✗ ADs need to keep updating down-paths on path server
  - ✗ Increased packet size
  - ✗ Static path binding, which may hamper dynamic re-routing

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

Performance Benefits

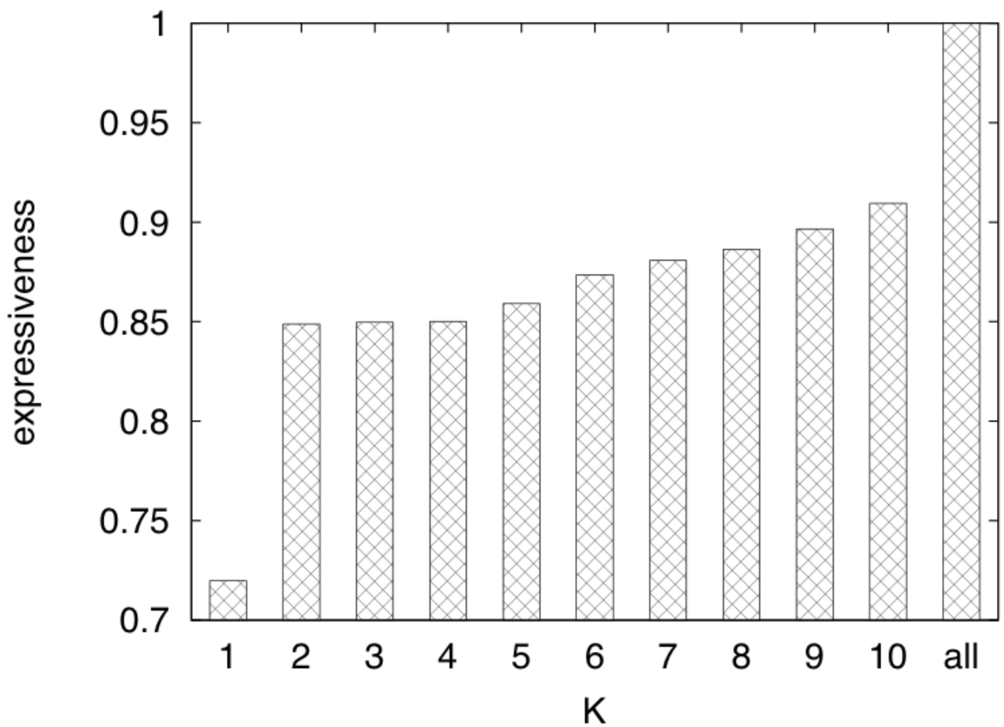
- Scalability
  - Routing updates are scoped within the local TD
- Flexibility
  - Transit ISPs can embed local routing policies in opaque fields
- Simplicity and efficiency
  - No interdomain forwarding table
    - Current network layer: routing table explosion
  - Symmetric verification during forwarding
  - Simple routers, energy efficient, and cost efficient

Evaluation

- Methodology
  - Use of CAIDA topology information
  - Assume 5 TDs (AfriNIC, ARIN, APNIC, LACNIC, RIPE)
  - We compare to S-BGP/BGP
- Metric 1: additional path length (AD hops) compared to BGP
  - *Without* shortcuts: 21% longer
  - *With* shortcuts:
    - 1 down/up- path: 6.7% longer
    - 2 down/up- path: 3.5% longer
    - 5 down/up- path: 2.5% longer

Evaluation (cont'd)

- Metric 2: Expressiveness
  - Fraction of BGP paths available under SCION



Related Work

- Routing security
  - S-BGP, soBGP, psBGP, SPV, PGBGP
  - Only topological correctness; addressed a subset of attacks addressed in SCION
- Routing control
  - Multipath (MIRO, Deflection, Path splicing, Pathlet), NIRA
  - Only given control to the source, and/or little security assurance
- Next-generation architectures
  - HLP, HAIR, RBF, AIP, ICING/IGLOO
  - Focusing on other aspects (reducing routing churns and routing table sizes, enforcing routing policies, and providing source accountability)

Conclusions

- 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