End Site Control Plane System Project

Phil DeMar (FNAL), Dantong Yu (BNL),
Martin Swany (Univ. Of Delaware)

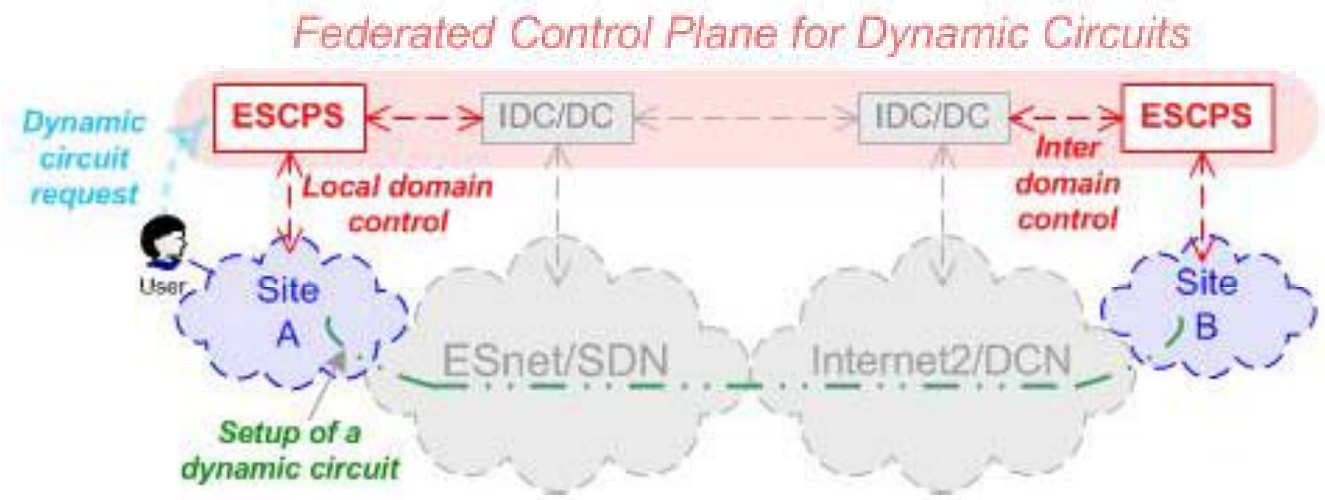
CHEP 2010
Academia Sinica
Taipei, Taiwan

Outline

- ESCPS Basic Function(s)
- Model
- Design

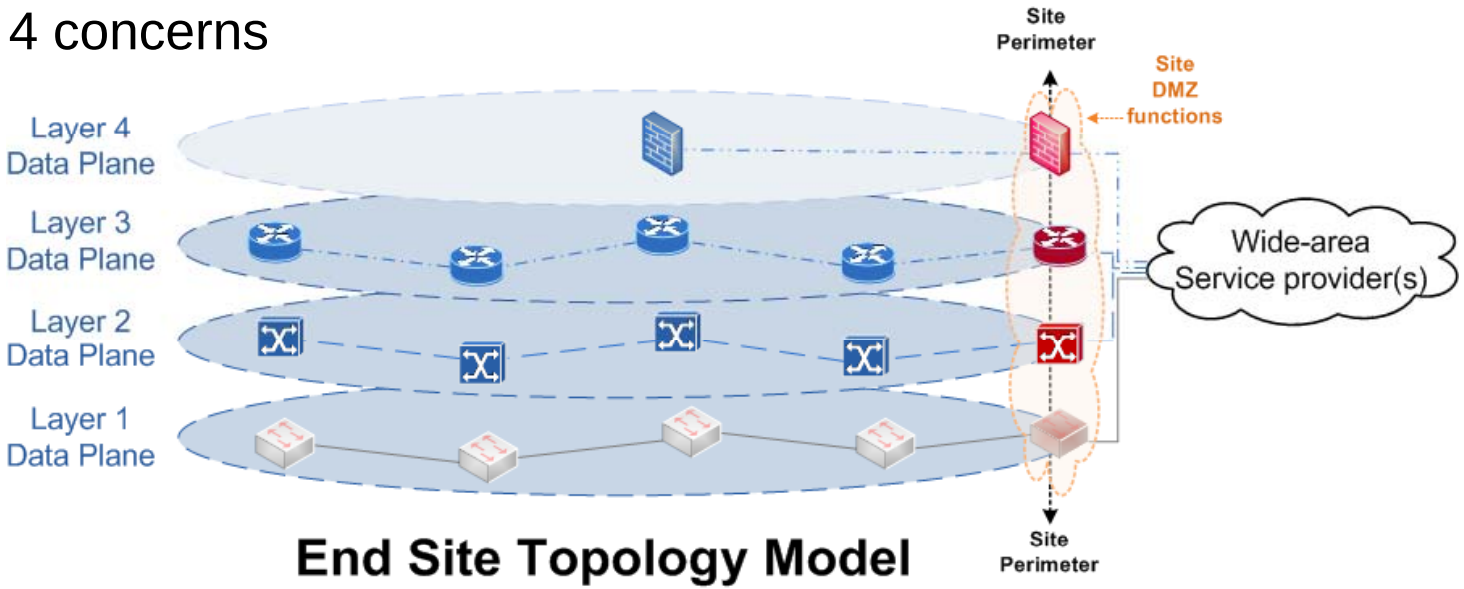
End Site Control Plane System Project

- Network service to facilitate site use of circuit services:
  - Accept and process user/app requests for circuit services
  - Provide local interface to & coordination of WAN circuit services
  - Configure local network infrastructure for use of circuits
  - Long term vision: End site component of federated control plane for circuit services



What's Unique About End Site Circuit Services?

- Same general multi-layer architecture as WANs
- Unique aspects of end sites:
  - Circuit termination pts.
  - Virtual path beyond circuit termination pt
  - Site perimeter
  - Layer 4 concerns



End Site Topology Model

Network Models

- Elements of a typical model:
  - Hosts
  - Devices (router, switch)
  - Physical links (interfaces, ports)
- Elements of ESCPS model
  - Aggregated Flow End Entities
  - Virtual Paths
  - Rules:
    - configuration units that need to be deployed to create a desired virtual path
    - elements of traditional network models used as attributes or parameters of these rules

Basic Definitions

Flows

- ***Flow:*** Unidirectional stream of packets identified by common set of keys:
  - Source/dest. IP addr., source/dest. port #, protocol ID, others (ie. DSCP field)
- ***Application Flow:*** Bidirectional aggregation of flows from a common data movement
- ***Aggregate Flow End-Point Entity (AFEE):*** A physical or logical entity that sources or sinks application flows

Circuits / Paths

- ***Virtual circuit:*** Dedicated network path between layer 3 end points, with no intermediate layer 3 hops, that provides specific network services for designated application flows
- ***Virtual network service path:*** The portion of an end-to-end network path for which network services are provided
- ***End-to-end network path:*** Network path between source & destination AFEs

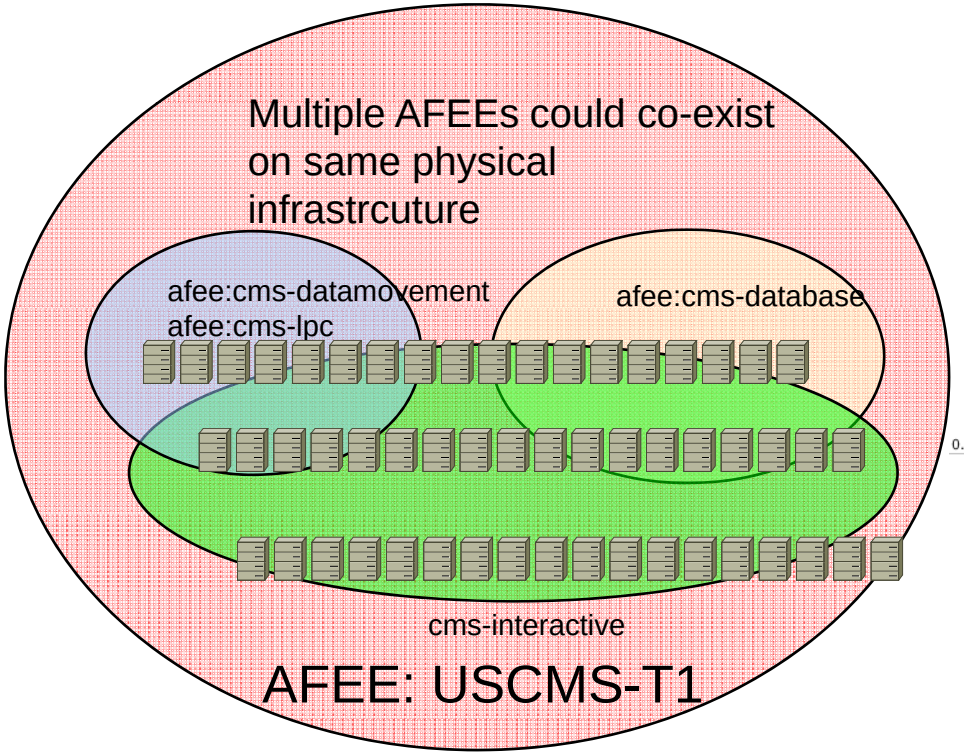
Devices

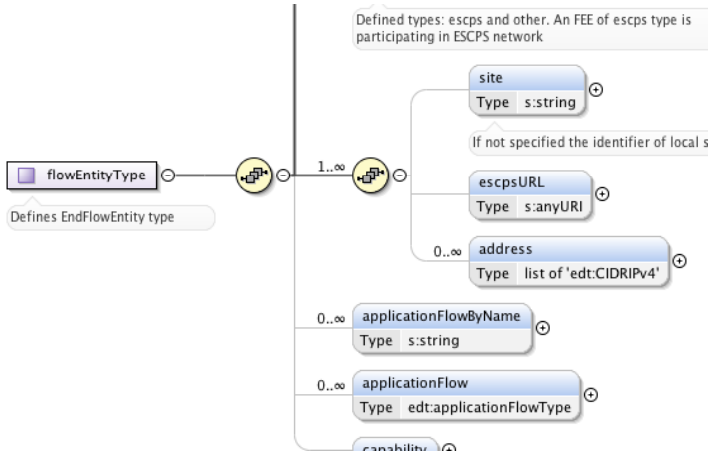
- ***Circuit Termination Device:*** Network device on which a circuit terminates
- ***Circuit Transit Device:*** Network device that supports a transiting circuit
- ***Circuit Admission Device:*** Network perimeter device interface where a circuit enters a network domain

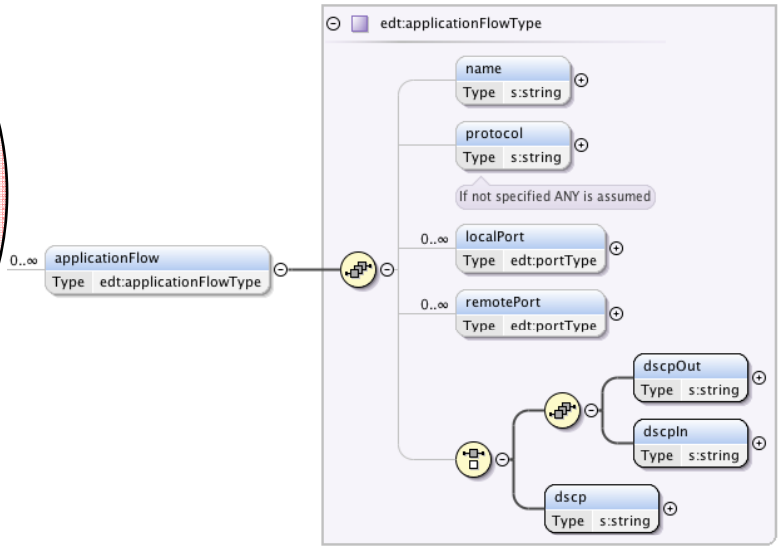
Aggregated Flow End Entity (AFEE)

Aggregated Flows (AGGFLOWS): *multiple application flows aggregated for each direction of traffic (bidirectional flows) that associated with one or multiple applications or aggregated based on other schemes as defined by site's needs.*

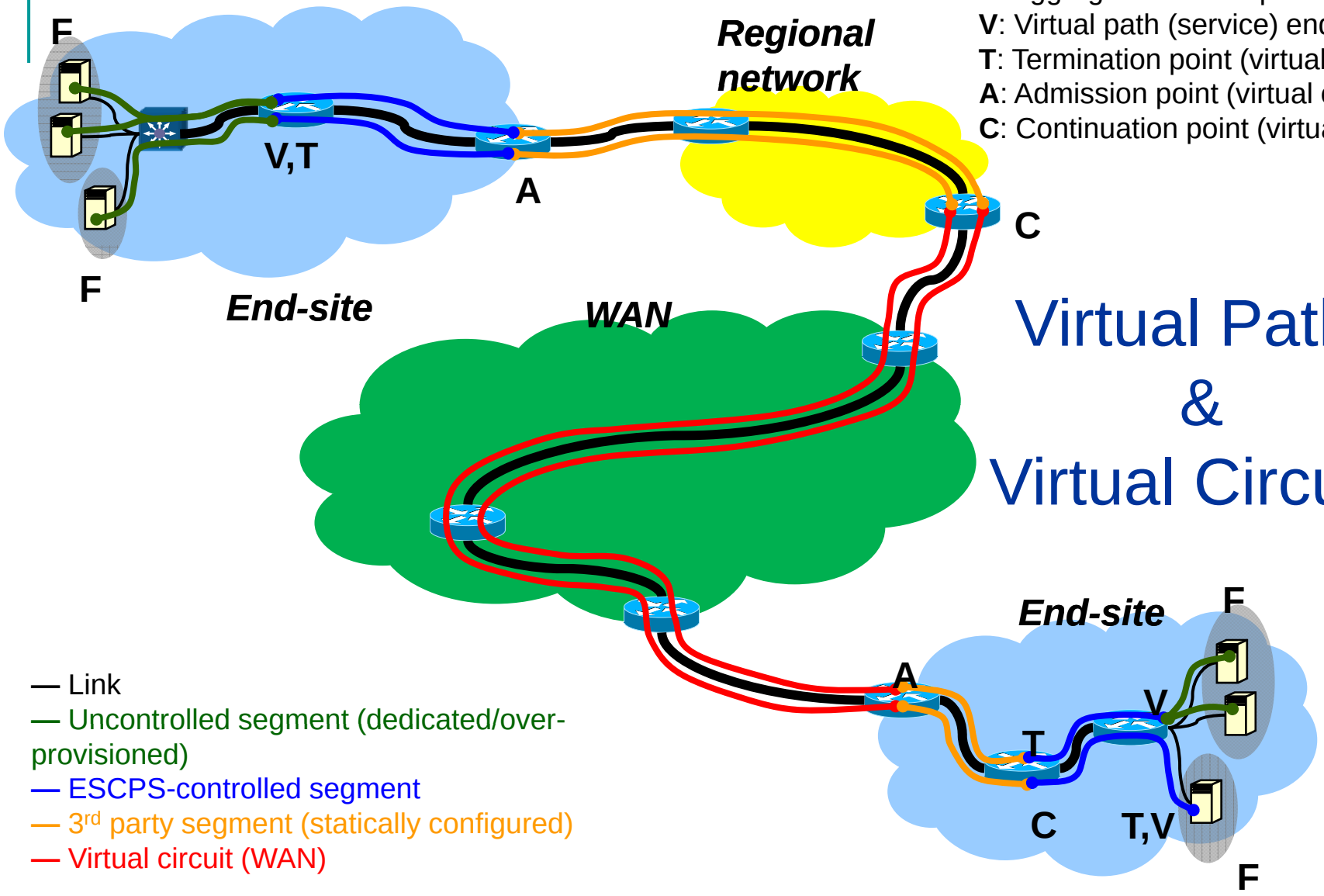
Aggregated Flows End Entity (AFEE): *An entity of IP network that generate or sinks aggregated flows.*







F: Aggregate flow endpoint
V: Virtual path (service) endpoint
T: Termination point (virtual circuit)
A: Admission point (virtual circuit)
C: Continuation point (virtual circuit)

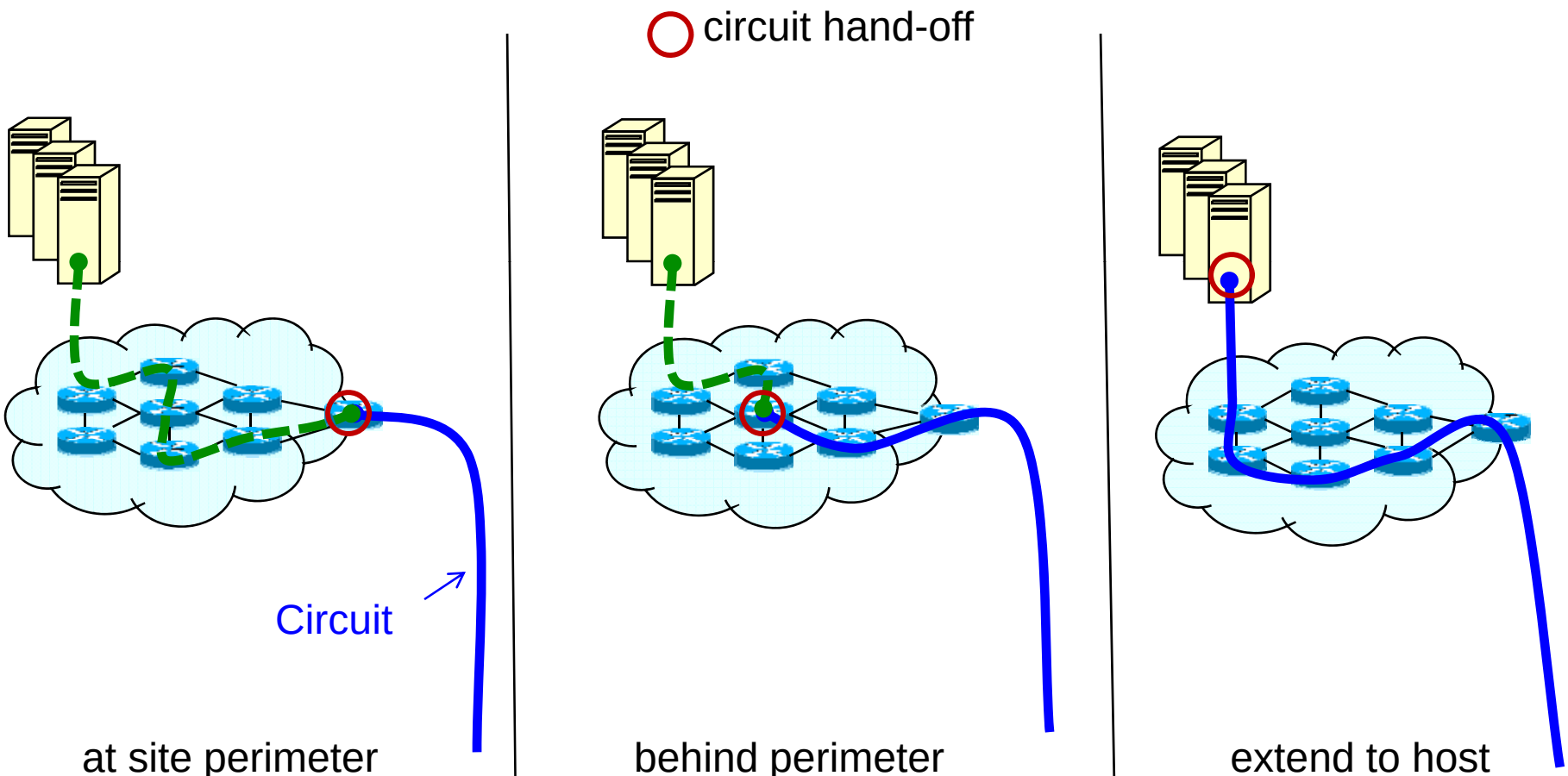


Virtual Paths & Virtual Circuits

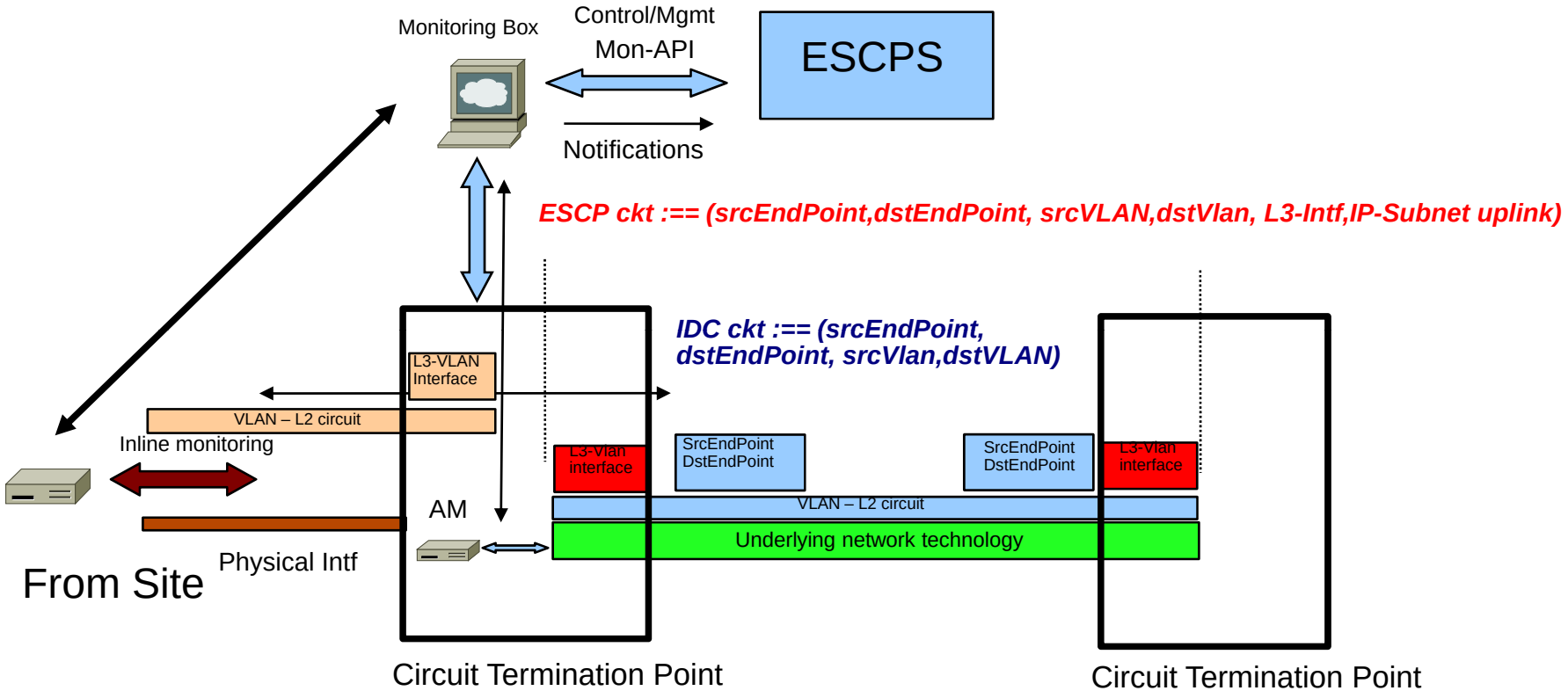
— Link
— Uncontrolled segment (dedicated/over-provisioned)
— ESCPS-controlled segment
— 3rd party segment (statically configured)
— Virtual circuit (WAN)

Office of Science
U.S. DEPARTMENT OF ENERGY
Fermilab
BROOKHAVEN NATIONAL LABORATORY

An Objective:
Selectable Circuit Termination Points



Circuit model from ESCPS perspective

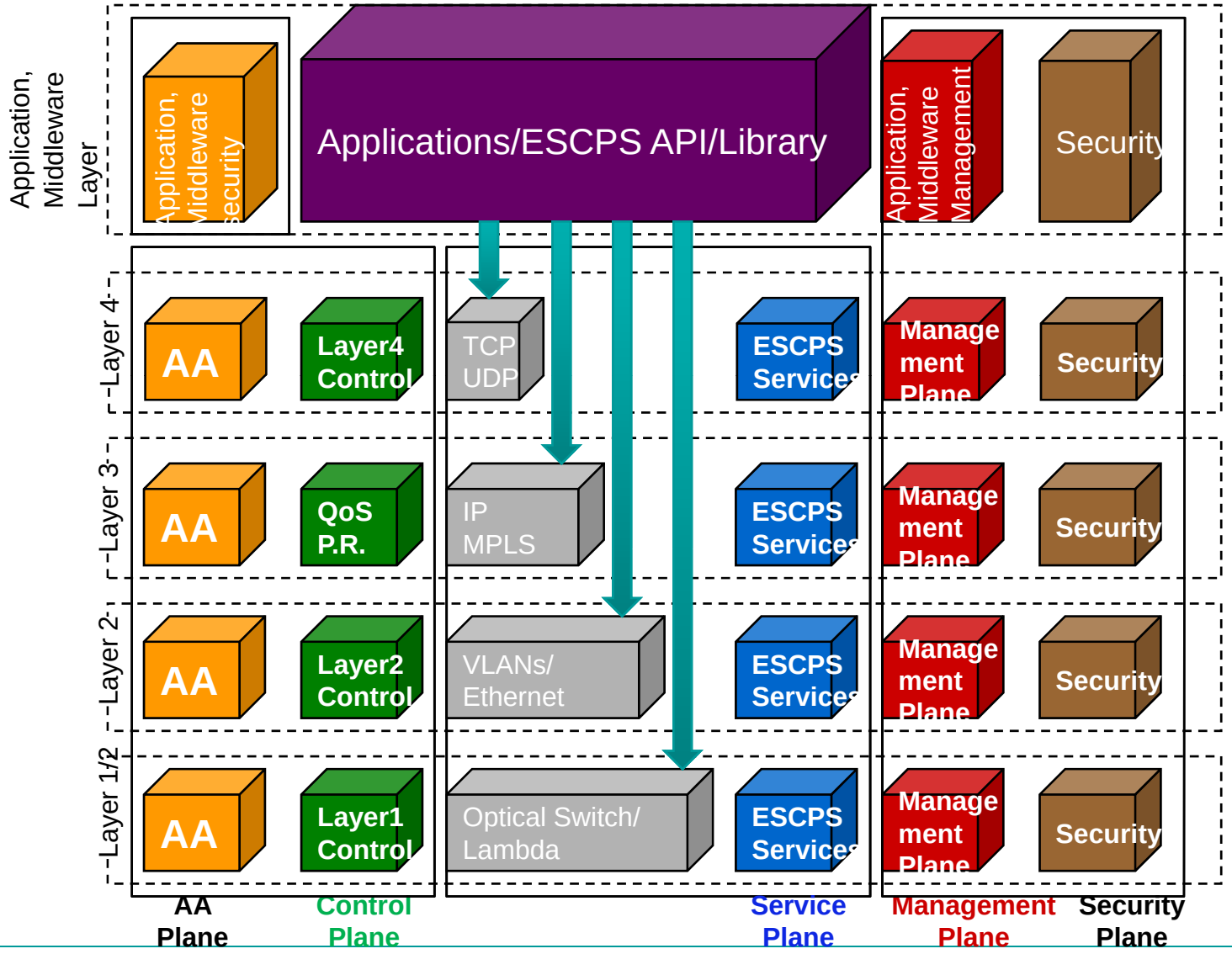


The same circuit but from different perspectives

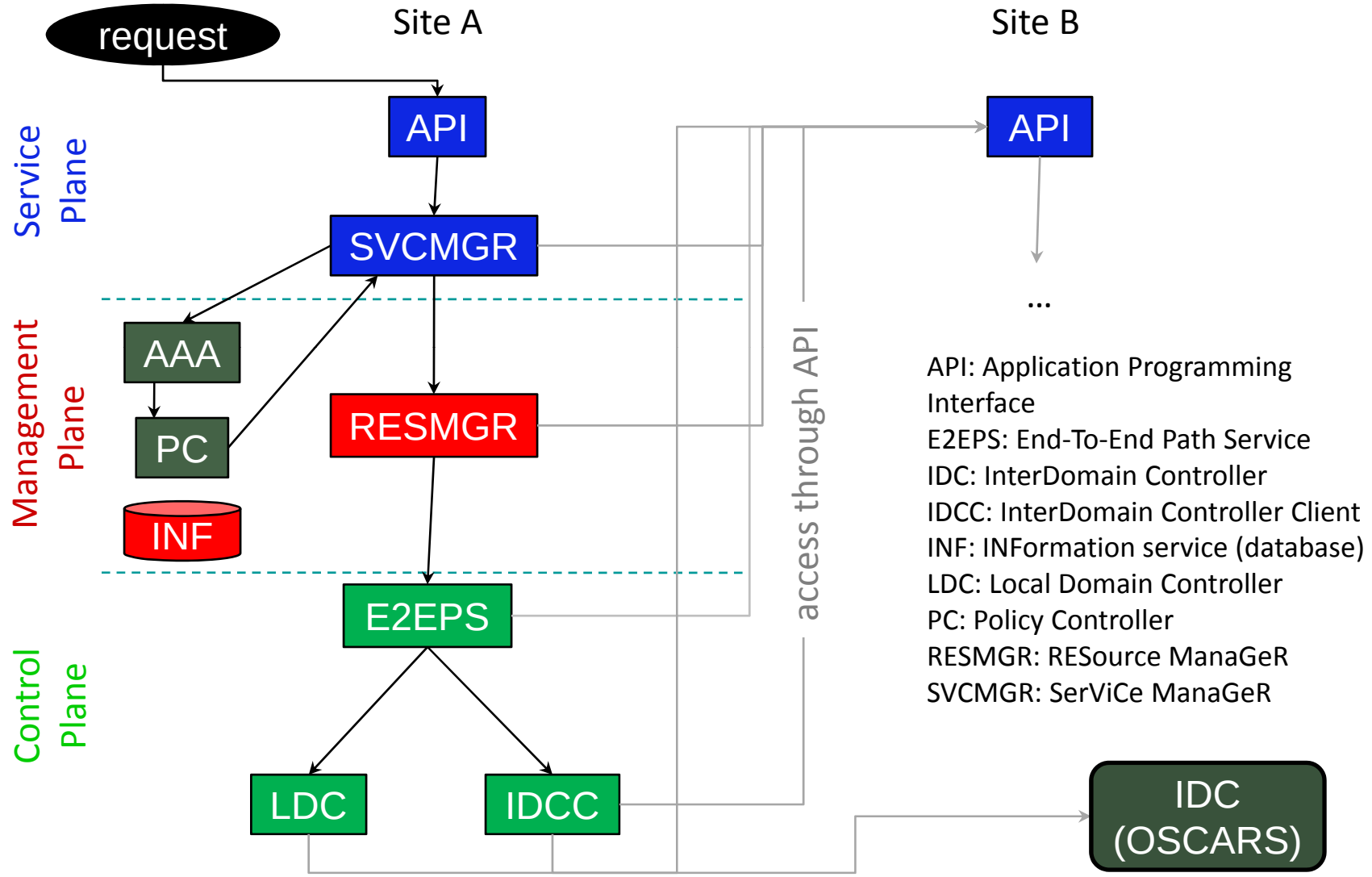
Site ckt:== (vlan,l3vlan-intf,ipsubnet uplink)
ESCPS ckt:== (srcEndPoint,dstEndPoint, Vlan,L3-vlan intf, IP subnet uplink)
IDC Ckt :== (srcEndPoint, dstEndPoint, vlan)

Office of Science U.S. DEPARTMENT OF ENERGY
Fermilab
BROOKHAVEN NATIONAL LABORATORY

Multi-Layer Capability Planes (End System perspective)



ESCPS Architecture

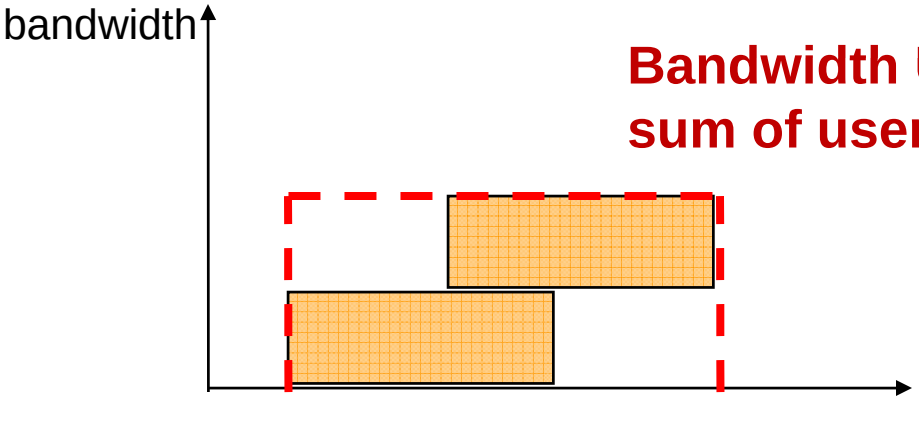


Components (I)

- API: main layer of communication between clients/ESCPS and ESCPS site instances
- SVCMGR: interprets & handles API requests
  - Coordinates with AAA and policy control components
  - Maintains service tickets & assigns unique request IDs
  - Engages RESMGR for service processing
- RESMGR:
  - Manages resources & policies associated with service request
  - Negotiates reservations across all domains
    - BAGs: intersection & fitting of resource requests results in set of solutions
  - Coordinates with E2EPS for realizing a network path reservation

RESMGR: Reservation Consolidation

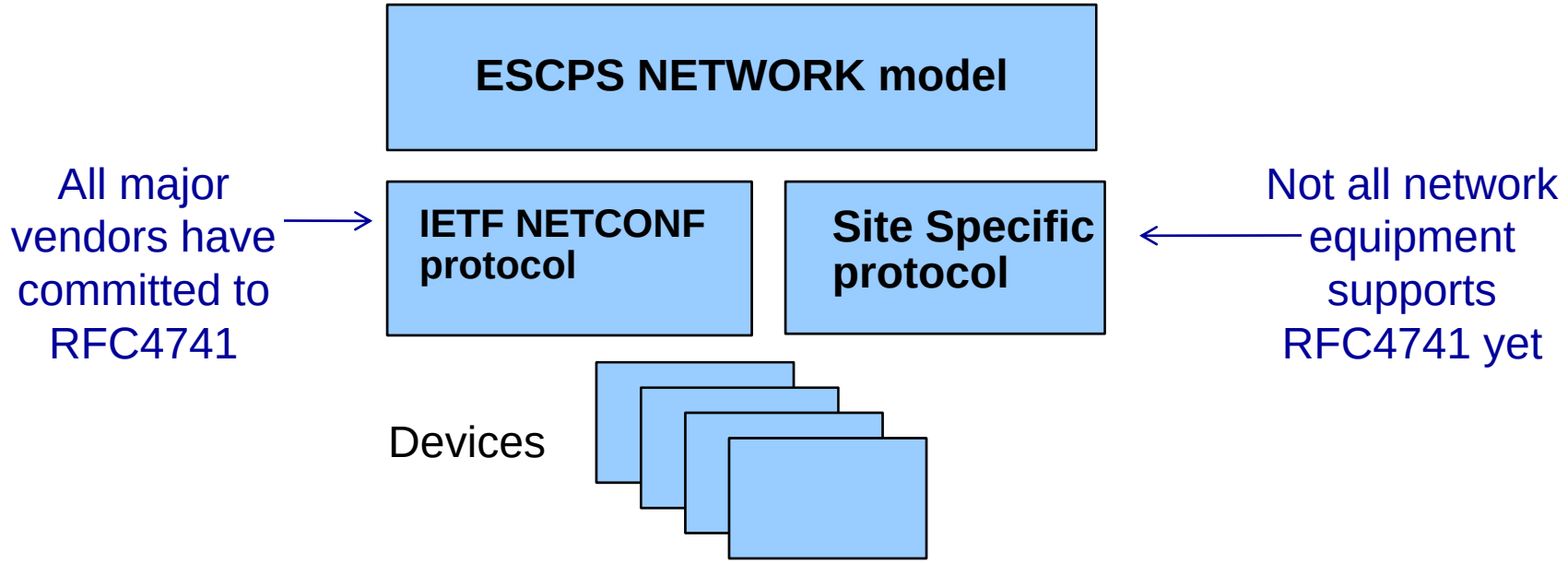
- Motivation
  - To survive from a limited number of VLAN (circuit) IDs
  - To reduce WAN operations (setup and tear-down)
- Idea
  - Use one VLAN (WAN reservation) for multiple user reservations
  - However, bandwidth will not be fully utilized



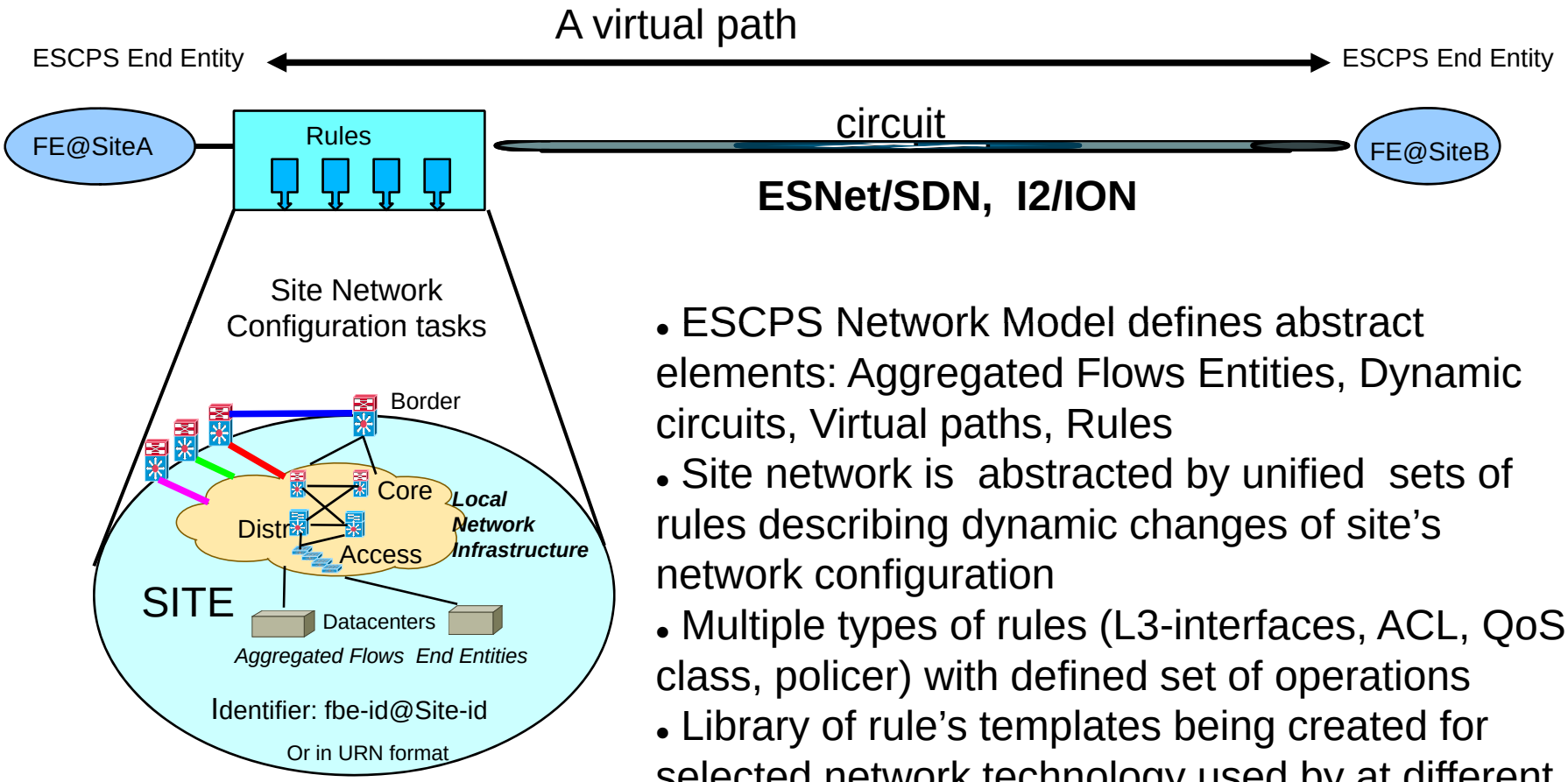
Bandwidth Utilization =
sum of user reservations / WAN reservation

LDC

- Local Domain Controller (LDC):
  - Configures site LAN according to network model
    - Must be capable of dealing with multiple layers
  - Based on RULES (site-specific configuration units)
  - Extensible to support multiple hardware types & layers
    - Goal is to conform to standards

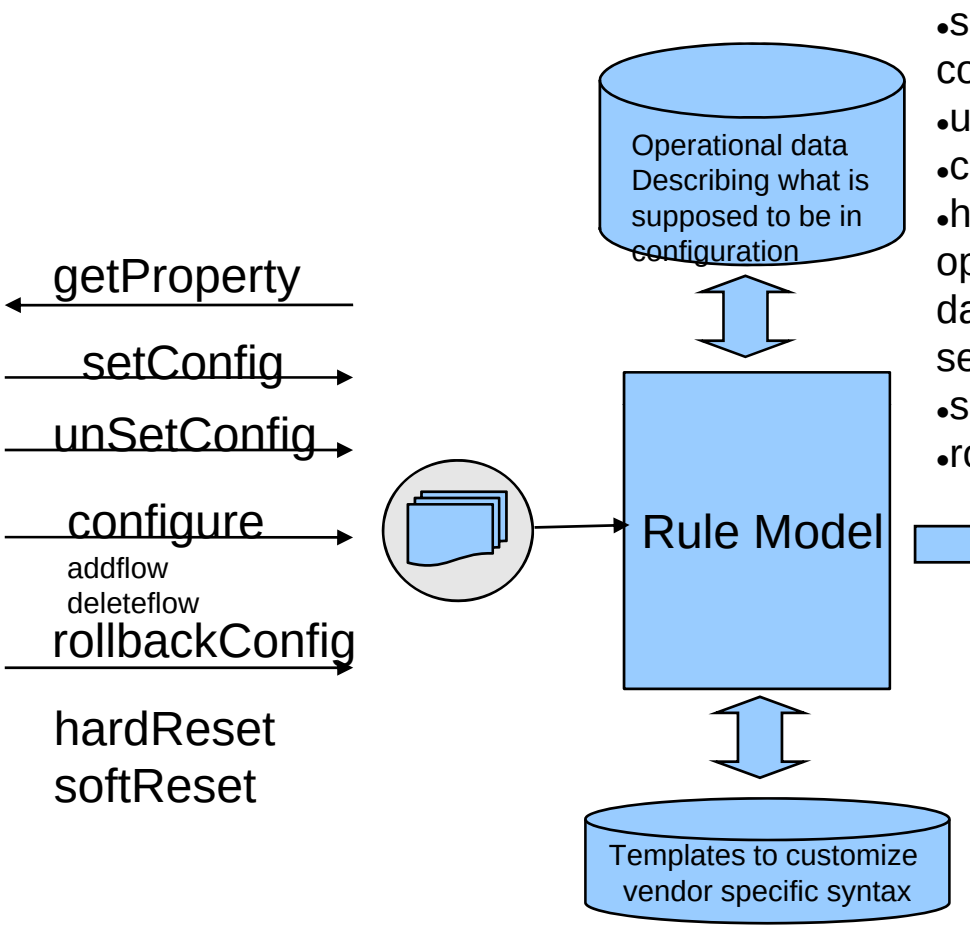


LDC in the ESCPS Model



- ESCPS Network Model defines abstract elements: Aggregated Flows Entities, Dynamic circuits, Virtual paths, Rules
- Site network is abstracted by unified sets of rules describing dynamic changes of site's network configuration
- Multiple types of rules (L3-interfaces, ACL, QoS class, policer) with defined set of operations
- Library of rule's templates being created for selected network technology used by at different sites

LDC: Unified Rule Interface

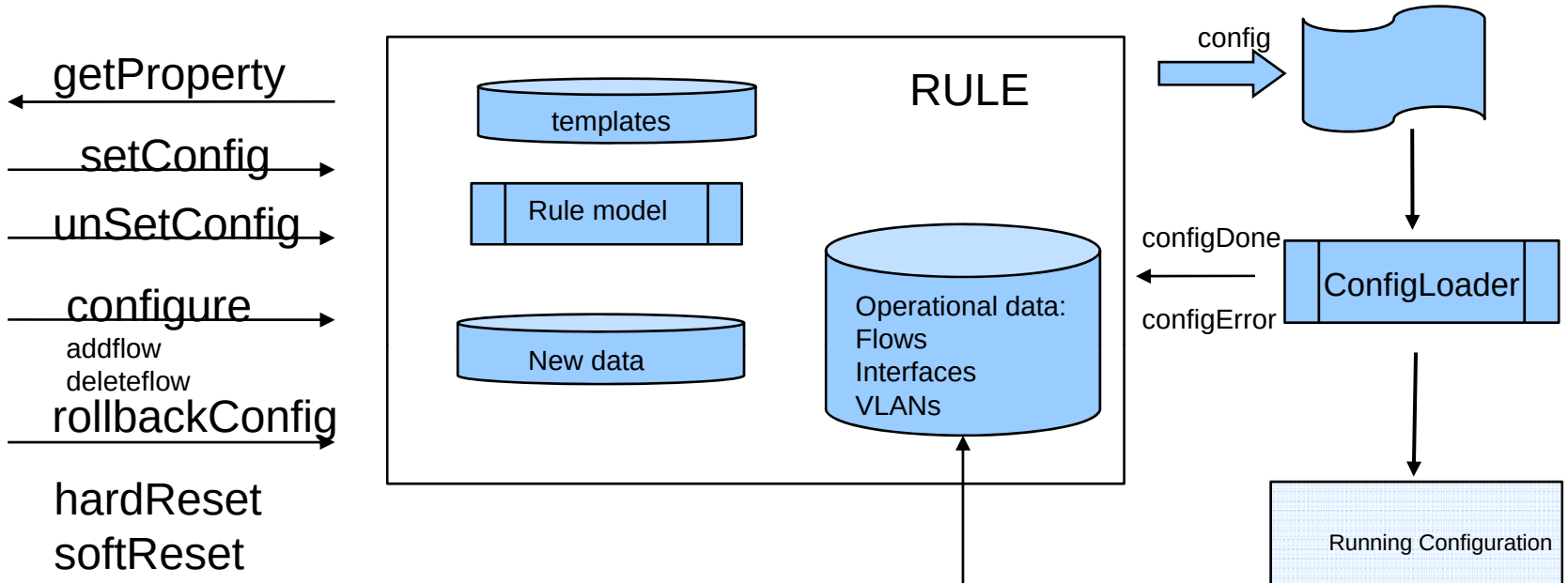


- setConfig – create fragment of new configuration
- unSetConfig – remove configuration fragment
- configure – configure selected rule's property
- hardReset – recreate configuration with all operational data removed and then do unSetConfig/ setConfig
- softReset – remove all operational data
- rollback – undo the last change

A fragment of configuration to upload in devices

Properties: (e.g. IP address, netmask,FlowsToAdd, FlowsToDelete, Status)

LDC: Architecture of a Simple Type Rule



Rule's model describes configuration fragments created by this rule, their type, properties and operations.

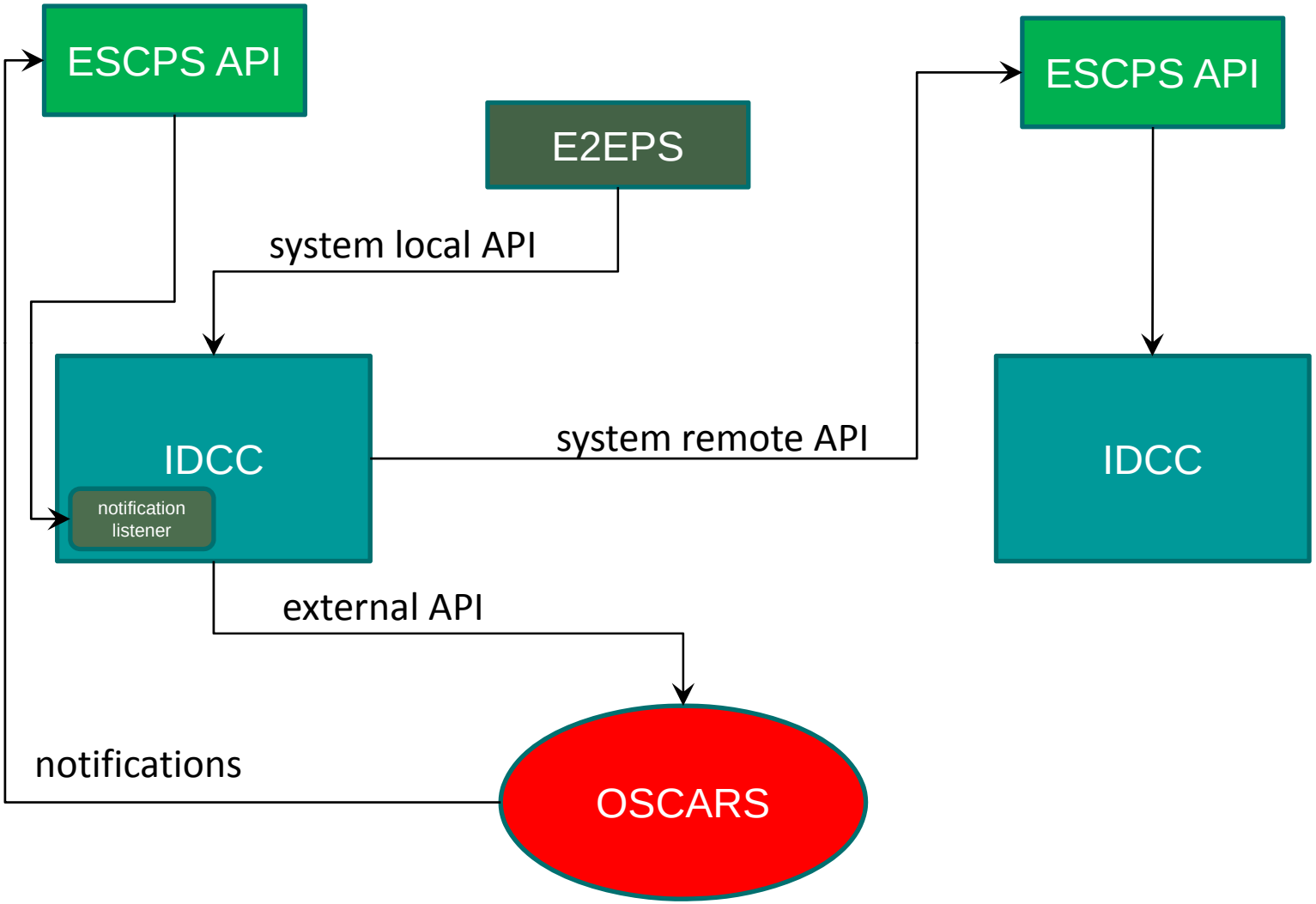
Examples of rules

route-mapType	VLANType
ACLType	QoS-ClassType
L3-InterfaceType	policerType

IDCC

- Basic role: interface/negotiate with IDC (OSCARS) for transit circuit reservations
- 3 APIs:
  - System local: E2EPS, others to IDCC
    - Core methods for managing circuits (e.g., create, teardown) , inquire about circuits, associate monitors with circuits
  - System remote: IDCC to remote IDCC
    - Methods for negotiating primary/secondary role between sites (prevents race conditions when reserving circuits, especially during trial-and-error iterations)
    - Methods for negotiating VLAN and IP address space to be used with a particular circuit
    - Distribute mutual exclusion-based negotiation protocol to avoid race conditions when deciding roles
  - External: IDCC to IDC
    - Methods for invoking the full set of IDC calls, including notifications

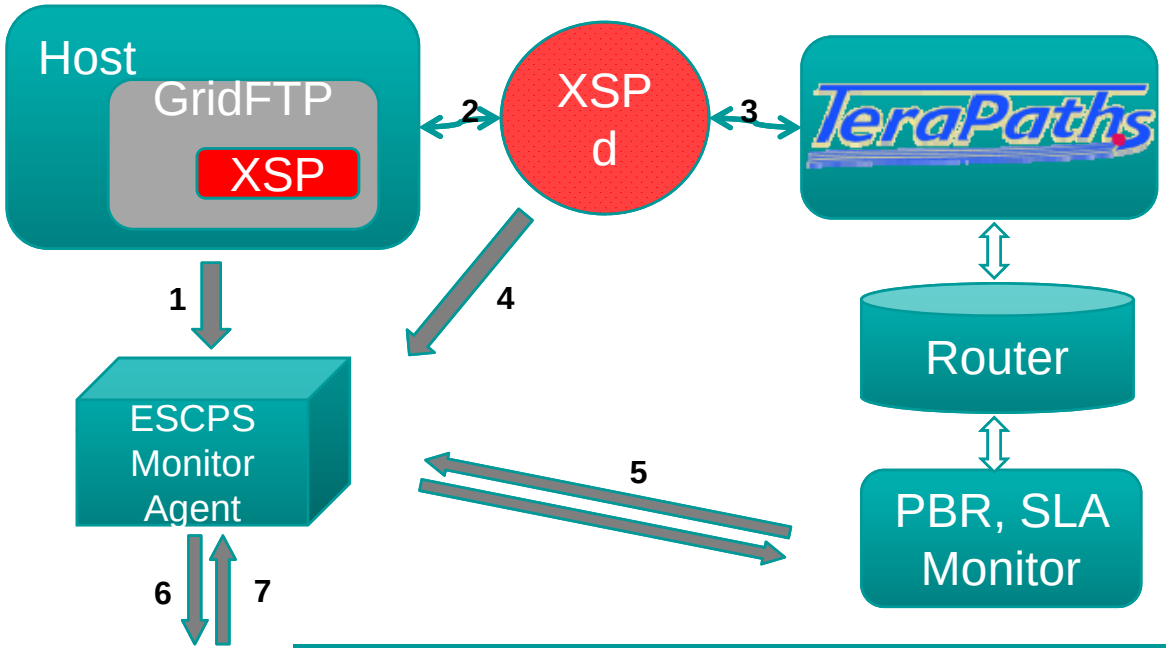
InterDomain Controller Client (IDCC)

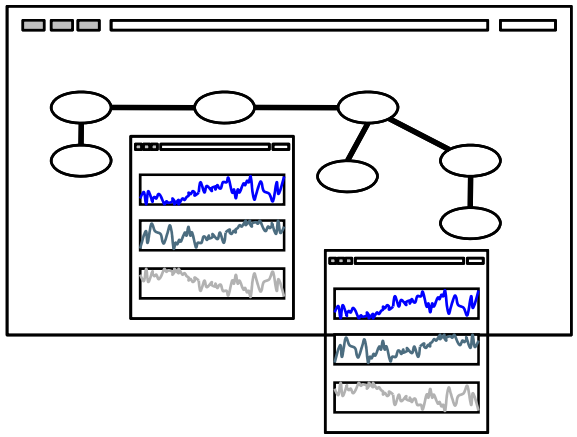


ESCPS Monitoring Agent

- Agent Roles:
  - Define topology, circuit, and measurement data models
  - Collect and cache measurement data
  - Handle requests for measurement retrieval and visualization
  - Maintain end-site topology information, map measurements to topological elements
  - Allow for registering monitored circuits and updating circuit status
- Agent provides WS-API for managing monitored circuits (*notify*, *update*, *remove* calls).
- Requests circuit-specific measurements (router looking glass) based on circuit status
- Provides functionality to poll and passively collect SNMP and host-specific measurements (eventually to invoke and collect active measurements)

ESCPS Monitoring Architecture





1. Host statistics collection and reporting with event daemon.
2. XSP client requests path on application request, sends path and application-specific parameters.
3. XSPd signals Terapaths to reserve path based on local config.
4. XSPd monitors path status and reports state to monitoring agent.
5. Agent requests PRB “policy matched” counters from looking glass monitor when path is active.
6. Monitoring agent caches measurement data (SNMP and active).
7. The agent visualizes the network topology and handles client requests for dynamically updated charts and path status.

Putting It Altogether - ESCPS workflow

1) User sends advance reservation request with source, destination, minimum bandwidth requirements and time period
2) Local service ticket gets created with unique ID
3) Authentication, authorization and policy validation
4) Local validation of resources
   - Find remote ESCPS
5) Initiate service request with remote ESCPS
6) Synchronization of service tickets
   - Identify which ESCPS will have primary role
7) Primary resource manager develops local and remote bandwidth availability graph (BAG)
8) Request transit circuit (IDCC), with parameters resulting from negotiation through BAG intersection or by trial and fail, iterations)
9) Scheduling of E2E path
   - Coordination with remote ESCPS
   - Collect local configuration
10) Wait for circuit setup
    - Verify transit circuit is in place
11) Local domain path activation (LDC)
    - Synchronize with remote ESCPS
12) Graceful circuit shutdown
    a) Local domain path deactivation
    b) Synchronize with remote ESCPS
    c) Transit circuit teardown
13) Ticket closeout and acknowledgement

ESCPS Project Status:

- A joint collaboration between Brookhaven National Lab, Fermilab, and the University of Delaware:
  - Completed year 1 of a three year project
- ESCPS component design specs completed and in prototype
- Next step: integration of components into unified framework
  - The major focus area for the coming year…
- Additional areas of effort:
  - Broadening equipment/vendor base of tested devices
  - Building configuration tools to assist site deployment
  - Always looking for candidate test sites…

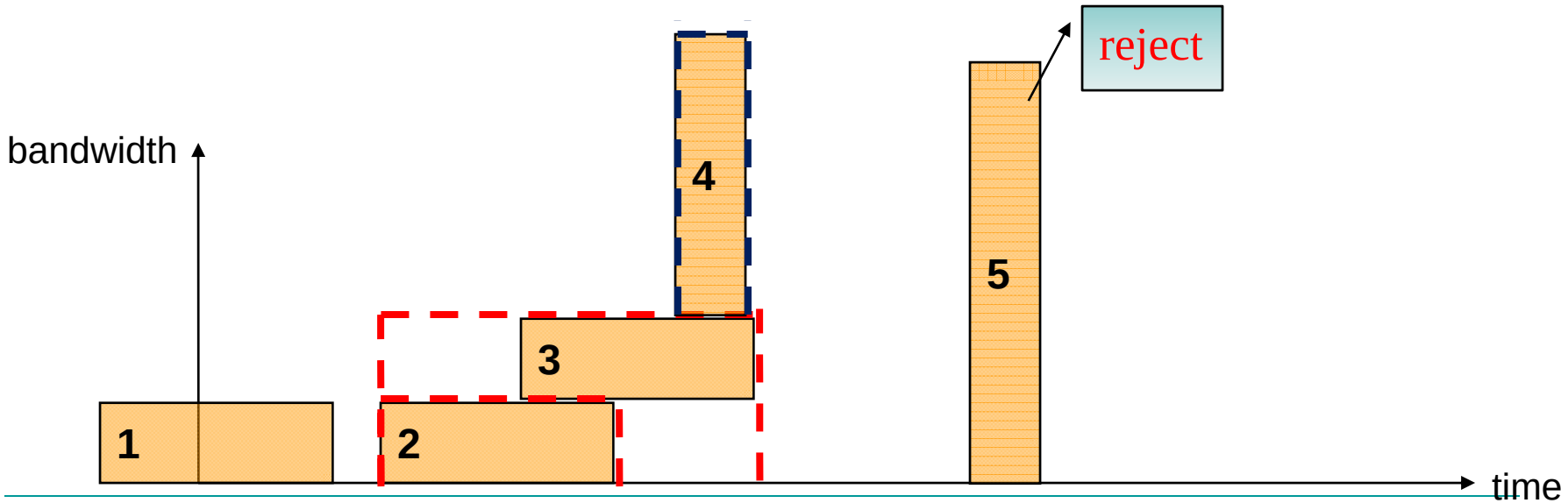
Additional Slides…

Bandwidth Allocation and Circuit Assignment

- Given
  - Offline case: a set of reservation requests
- Decision to make
  - Allocate bandwidths circuits (VLANs)
  - Assign reservation requests to circuits
- Objective
  - Maximize the number of requests that can be satisfied
- Main Constraints
  - Each reservation must be assigned to one circuit
  - The total capacity WAN provides
  - The bandwidth utilization must be higher than a given value
  - The number of available circuit IDs are constrained by a given value

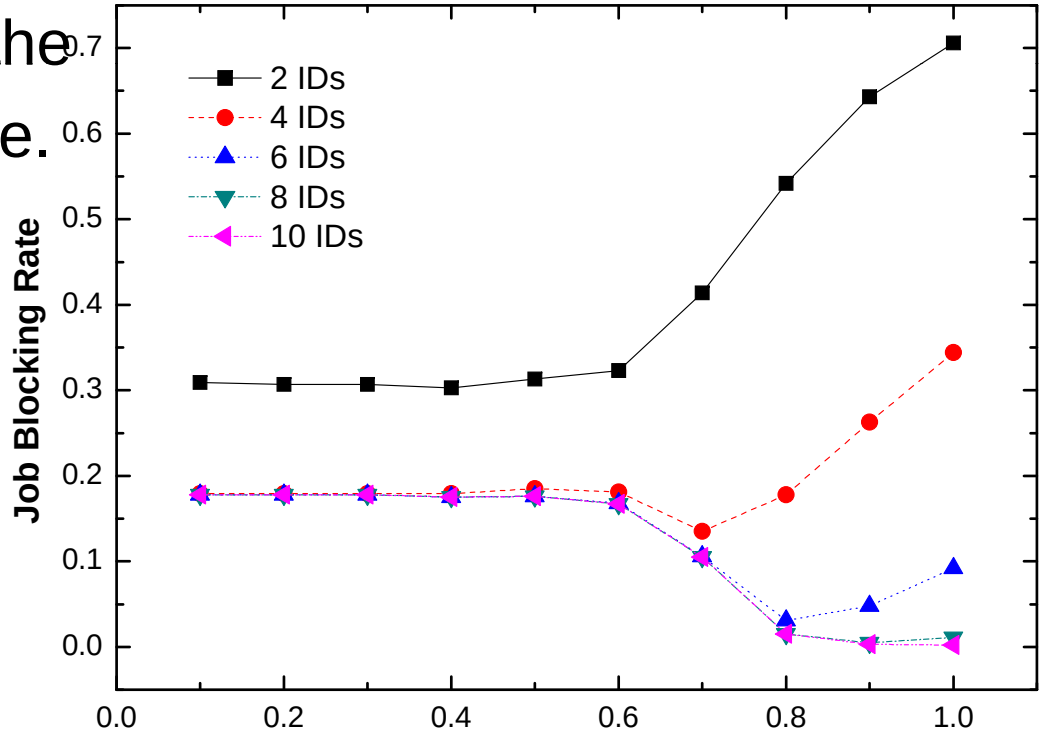
Preliminary Results

- Algorithm Sketch
  - Order requests
  - Use consolidation when possible (bandwidth utilization is high enough)
  - Assign new circuit when necessary (if circuit IDs and bandwidths are available)



Preliminary Results

- Online case
  - Choose an “optimization window” near the new request and perform reservation consolidation
  - Need more IDs for the higher utilization rate.

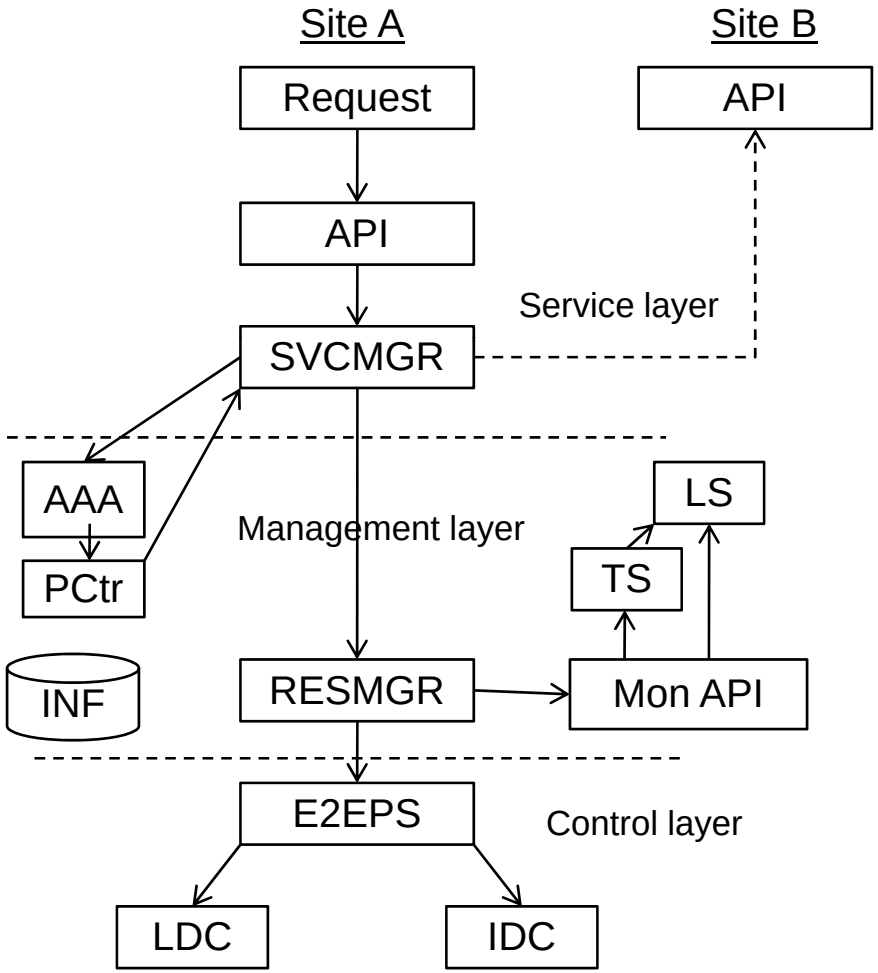


LDC – A local domain controller

Modifying on-demand Site Network to accomodate requested services

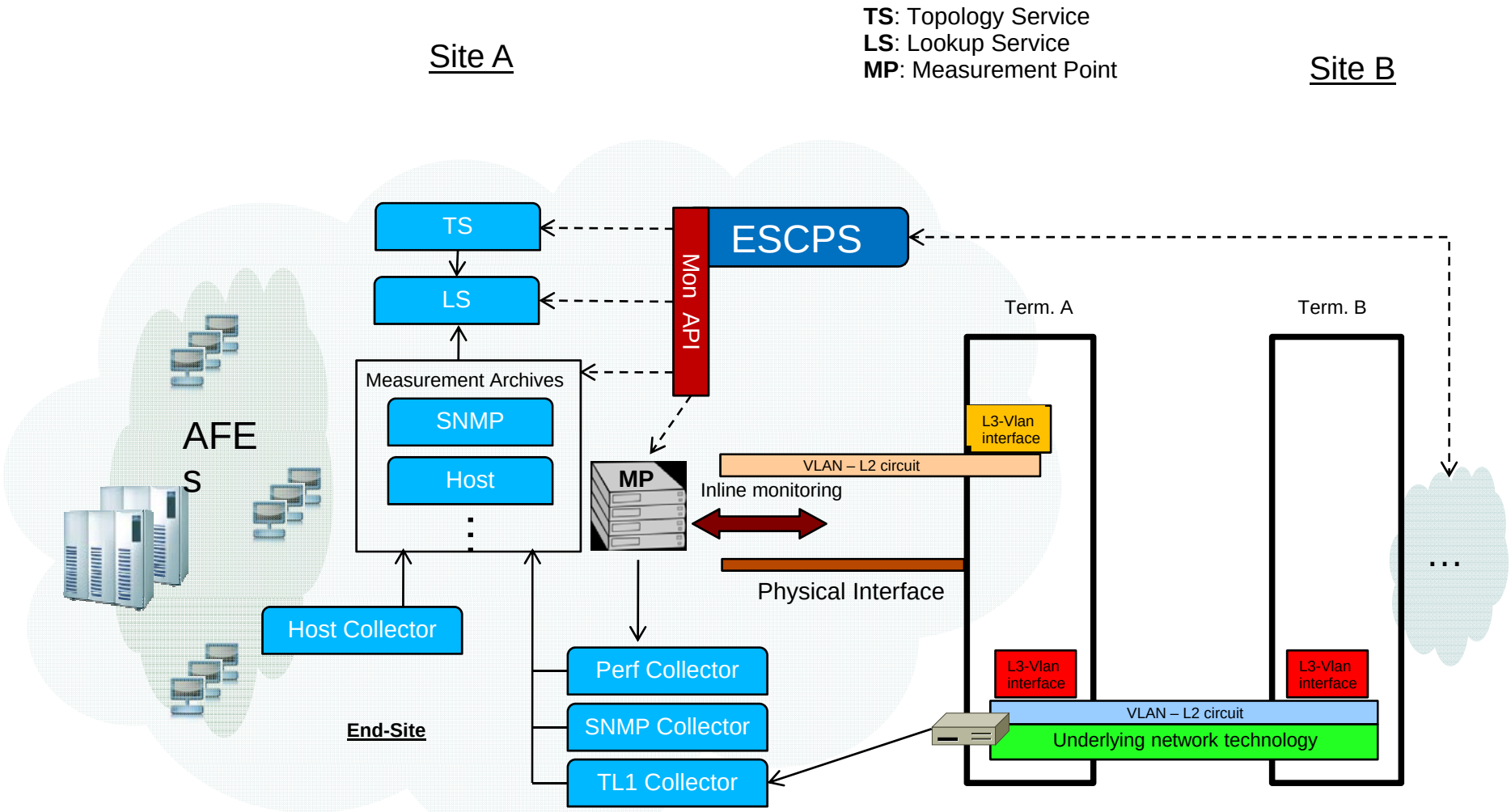
- setSiteVirtualPath (srcAFE, dstAFE, cktID, flows, BW, QoS)
- getVirtualPathStatus(Id)
- GetVirtualPathInfo (ID)
- updateSiteVirtualPath(Id, params)
- addFlows (src, srcOptionalSrcAttr, dst, dstOptionalAttr, circuit, flows, BW, QoS) – multi form of input parameters will be supported
- deleteFlows
- getKnownAFE – return information about locally defined AFEs, lust or XML-Docs
- getAFEInfo - return info about specific AFE
- getKNownCircuit
- getCircuitInfo
- getQoSModel


ESCPS Monitoring Workflow



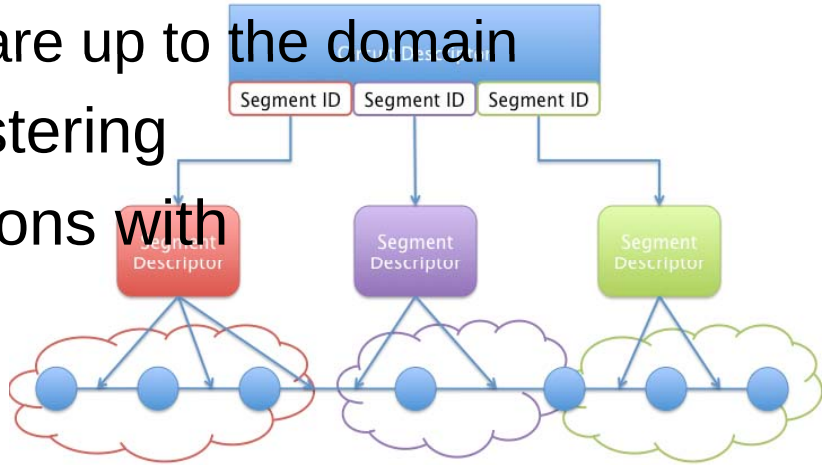
1. Default monitoring profile created or specified by user via request.
2. Monitoring profile verified by AAA and policy components (PCtr)
3. Establish monitoring session with remote ESCPS agent (SVCMGR)
4. Monitoring profile verified by remote ESCPS
   a. The “master” is also responsible for initiating measurement collection
5. Once circuit ID is known (IDC), monitoring API performs lookup on monitored segments
   a. Collection begins on available segments based on advertised topology information (LS-TS)
   b. Active measurements to be performed based on active monitoring profile
6. Similar case for the local domains (A and B) with registered local segments (LS)
   a. Aggregates generated and exposed through monitoring API (web-app or other service)
7. Collection stops once circuit is terminated or when requested
8. Historical monitoring data saved or processed for analysis (INF)

ESCPS Monitoring Components



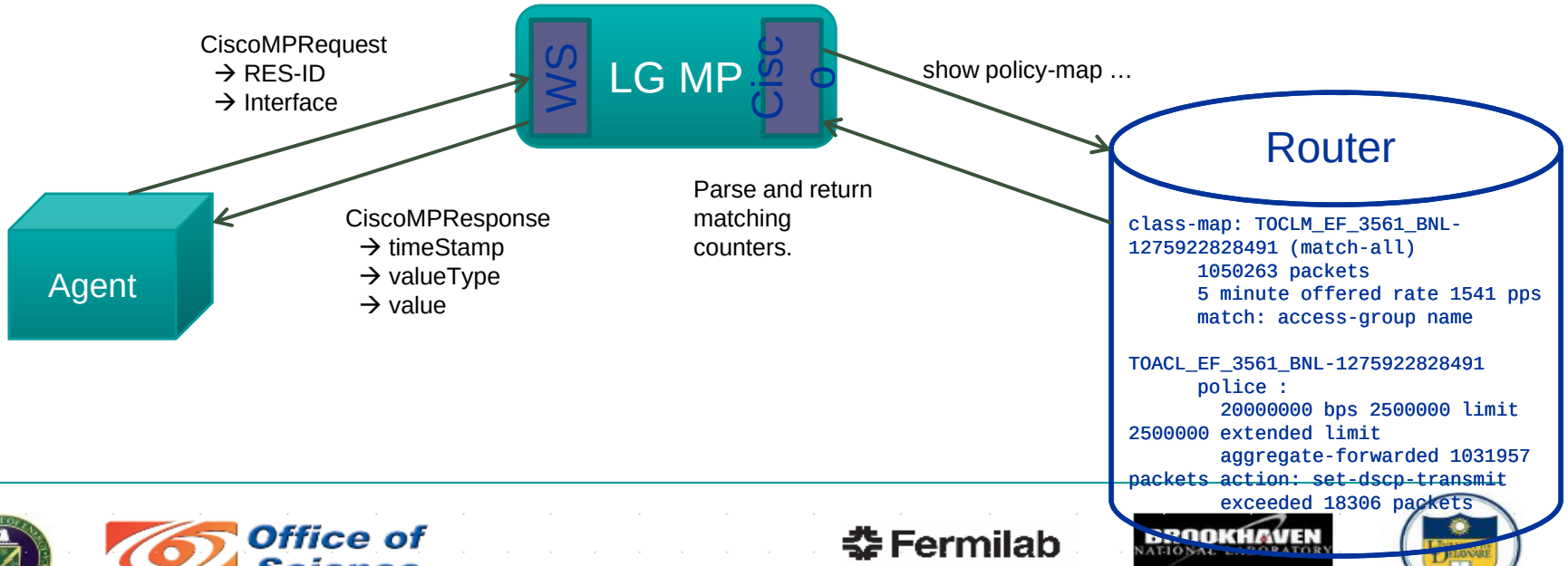
Monitoring Architecture Details

- Monitoring agent must construct a circuit descriptor
  - The requesting domain assumes role of the “circuit description agent”
  - Includes details about the domain-specific segments
  - Identifier, bandwidth, source, destination, etc.
- Identifiers for each segment defined by network model
  - Contents of the description are up to the domain
- Agent is responsible for registering segment and circuit descriptions with the lookup infrastructure



Universal Looking Glass MP (PBR, SLA Mon.)

- Provides a WS-interface to switch/router metrics within the end-site
- Will support a number of devices (currently Cisco 6509)
- May also be enabled to invoke and return results of active measurements (traceroute, ping, etc) or enabling passive measurements (TCPdump, Ethereal) at points along the path



XSP

- Proof-of-concept deployment as SVCMGR for Terapaths reservations

- XSPd (*daemon*) implements protocol frontend
  - Accepts on-demand reservation requests from clients
  - Signals Terapaths to allocate a circuit and monitors circuit status