Internet Engineering Task Force
Internet Draft
Document: draft-walker-ccamp-req-00.txt
Expires: August 2001

J. Jiang
D. Walker
J. Wang
SS8 Networks Inc

February 2001

Common Control and Measurement Plane Framework and Requirements

<draft-walker-ccamp-req-00.txt>

Status of this Memo

Abstract

This document describes architectural and protocol requirements for the Common Control and Measurement Plane.

Table of Contents

1. Introduction

As networking technology continues to evolve, there is an ever-increasing number of transport layer protocols that one is likely to encounter in developing end-to-end solutions. Along with the growing stringency of Service Level Specifications (SLS), there is both a need to be able to provide a finer level of control over network traffic in terms of the level of service that can be delivered by the various technologies, as well as a need to ensure that the network is providing the required level of service.

The various network technologies, such as MPLS label switching, ATM, Diffserv, optical switching, and more, frequently come with a unique set of mechanisms that offer ISPs the tools they need to control and monitor technology-specific or even vendor-specific islands. The unique nature of these islands creates complex problems when larger networks are created by interconnecting such islands.

This draft presents a framework and set of generic requirements that are independent of the underlying technology and which can be used to ensure that the network can be monitored and controlled to provide specific levels of policy, security, and quality of service characteristics.

Networks can be functionally divided into three planes of activity: a data or transport plane, a control plane, and a management plane.

The control plane consists of logical entities (Control Elements) which perform network level coordination functions such as: state information management (acquisition, representation, dissemination), decision making (e.g. path selection), and action invocation (e.g. signalling).

The transport plane provides consists of entities (such as layer 2 and layer 3 switches, routers, and others, collectively referred to in this document as Transport Elements) which primarily switch or forward data (bearer or signalling) traffic. These entities may be statically or dynamically configured in order to determine how particular traffic is to be treated.

The measurement plane provides transport level resource status information to interested parties in order that appropriate policies may be applied (e.g. allowing routers to determine the appropriate next hop destination for outgoing packets).

The framework proposed in this draft suggests that Control Elements are able to control and monitor one or more Transport Elements. While the document presents a discussion on the relative merits of centralized and distributed control networks, it should be emphasized that CEs are logical entities which may or may not be co-located with TEs in actual implementations.

It must be noted that the requirements set out in this draft may be partially satisfied by extending existing protocols, such as COPS [7], MEGACO [2], OSPF [3], and others.

2. Definitions

Service domain: Service domain defines a portion of the network under one service providerÆs administration. All the network elements within a service domain have consistent view of the network and policy.

Clearing House (CH): Given the large number of access networks belonging to different service domains, it is not possible to have SLS between all domains on the Internet. A clearinghouse facilitates the authorization and logging or accounting between service domains for premium services. This does not preclude however some domains to have direct bilateral agreements, so as not to use any clearinghouse service when exchanging traffic.

Control and Measurement Plane: The control and measurement plane is a functional layer which is built on top of transport network to control the transport elements to perform service management, traffic engineering, policy control, and QoS control functions. The control and measurement plane is one of the three dimensions of a service providerÆs network, which includes transport plane (data plane), control and measurement plane and management plane.

Control Element (CE): The network components providing control capability for traffic engineering, service management, protection/restoration, policy control and end-to-end QoS control. These components communicate with TEs to collect network status and resource information, compute source route or perform path provisioning for tunnel management, execute policy logic, update its policy information base, and exchange this information with other CEs.

Transport Element (TE): The network components providing transport function to switch or forward bearer traffic. Examples of TEs include MPLS LSRs, ATM switches, Lambda switches, DiffServ capable routers, PSTN-IP gateways, etc. A TE communicates with CE to report network resource and status information, receive and execute policy decisions from CE for traffic engineering, service management, protection/restoration, policy control and end-to-end QoS control.

Peer: CEs are connected to each other via a logical link, or an association. The two CEs that have associations form a peer relationship. This peer relationship is abbreviated to as peer in this draft.

Internal peer: An internal peer is a peer relationship between two CEs in the same service domain.

External peer: An external peer is a peer relationship between two CEs in two different service domains.

Internal CE: An internal CE is a CE that has no external peer.

Border CE: A border CE is a CE that has at least one external peer.

3. Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC-2119 [4].

4. Common Control and Measurement Plane

A network can be functionally divided into three planes: a data or transport plane, a control plane, and a management plane. The control plane consists of network control elements. The network control plane elements perform network level coordination functions including: state information management (acquisition, representation, dissemination), decision making (e.g., path selection), and action invocation (e.g., signalling). In order to manage state information of the network, measuring and monitoring the network resource and status is the key function. To emphasize this function, the control plane is referred to as control and measurement plane in this draft.

4.1. Functions of the Control and Measurement Plane

This plane is designed to perform the following functions:

- Traffic engineering

  It is able to control the traffic flows in the network so that the network resource is utilized in a most efficient fashion. With this feature, this plane must be able to handle various types of traffic in terms of their QoS requirements including delay, packet loss, bandwidth requirements, etc over a mixed underlying transport networks. This feature requires the CEs to collect traffic and resource information in the network, to compute the best path for each flow, and to issue control commands to TEs. This plane must support both time-dependent and state-dependent traffic engineering.

- Support end-to-end QoS

  This plane needs to support end-to-end QoS for its customers. For this purpose, the CE must have not only the network resource knowledge of its own service domain, but also access to the performance measurements of the other service domains a particular

  flow is to transverse. These performance measurements are collected by TEs and CEs in a service domain and may be exchanged with CEs in other domains.

- Support policy

  Various levels of policy need to be supported by this plane. These include service policy, customer policy, resource policy, network functional policy, and network element policy. This plane must support policy creation, modification, and deletion. It must also support service policy advertisement among service domains. It must support three types of policy: QoS policy, service policy and traffic engineering policy. The framework must be such that is easily extended to support other policies.

- Support service provisioning and management

  For a service provider, the traffic engineering and QoS control is based on the customerÆs service profile and its service policy. The control plane must support customer service provisioning and management. This include communication with Service Management System (SMS) for Local Service Level Agreements (SLA) specification, SLA negotiation, service creation, modification and deletion. It should also be able to perform SLA advertisement among the CEs within the same service domain and LSA exchange among CEs in different service domains. It should also be able to allocate service to specific customer flows as required by SMS.

In addition to the generic management function as described above, this plane must also support management of particular services, such as VPN service.

4.2. Centralized Architecture

The control and measurement plane can be deployed in two different architectures: the control function is separate from the TEs or control integrated with the TEs. The former is referred to as centralized control and the latter is referred to as distributed control. Neither model should necessarily rule out the other, so that it is possible to have a centralized architecture with some CEs just happening to be co-resident with TEs, and on the other hand it is also possible to have a distributed architecture where the TE function on some physical entities is null.

With centralized architecture, the control function is allocated in one or a few centralized CEs that are physically separated from the TEs. The interaction between the CE and the TEs is via a set of protocols as defined and discussed in this draft. These protocols must be independent of the underlying transport network. It is each TEÆs responsibility to translate its technology sub-network specific resource representation into the abstracted common representation.

All the CEs form a control network. The control network may consist of one or more CEs depending on the size of the network and capacity of the CE. In case of multiple CEs, a mechanism must be defined for these CEs to communicate and synchronize policy, resource and traffic information, and provisioned service.

The centralized architecture has the following advantages:

- Easier to benefit from management information continuously collected by NMS (Network Management System)

  This information, such as performance alarms, failure alarms, and traps can be used with other information for the control elements to make control decisions.

  With distributed architecture, a distributed routing protocol relies mainly on timers and missing PDUs to detect a failure between two adjacent switching nodes.

- Easier for policy control

  Policy control consists of policy creation, installation, modification, deletion advertisement, and policy decision making. In reality, policy is usually service provider based (service policy, customer policy, accounting policy) or network based (network function specific and network element policy). To be able to provide end-to-end QoS, one service domain needs to exchange service level policy with its neighboring domains. With a centralized architecture, it is easier to maintain policy consistency because the policy control is performed at one (or a few) central place(s).

  With distributed approach, the policy creation needs to be done repeatedly on every transport elements. The policy advertisement between different domains are even more difficult with distributed architecture.

- Easier for traffic engineering of mixed underlying transport network

  A service providerÆs network may consist of mixed types of sub-networks. For example, a GMPLS network may consist of two Packet Switched Capable (PSC) MPLS sub-networks connected by one Lambda Switched Capable (LSC) MPLS sub-network [5]. With centralized architecture, a centralized decision can be easily made based on its consistent and complete view of the underlying network.

  On the other hand, with distributed architecture, the routing protocol are used to build and maintain a logical model of the network. Because not all routing entities have the same view of the overall network (e.g., two ATM label switching networks connected with one lambda switching network, the ATM switch in an MPLS-ATM network has different view from Lambda switch in an MPLS-

  optical network in terms of network topology, network resource and congestion status), a best decision based on entire network is difficult to make.

- Easier to operate

  With centralized architecture, new features or policies can be introduced with a simple upgrade.

  With distributed architecture, upgrading every switch with new routing software is difficult.

- Better flexibility

  With centralized architecture, a set of protocols between CE and TE must be well defined. This provides a flexibility where the control and measurement function can be allocated. It also allows separate TE and CE development to optimize their functionality.

- Better information consistency

  With centralized architecture, information is stored in a few central places. The possibility of session setup failure due to inconsistent information is lower than that in distributed architecture.

- Offload LSRs

  With centralized architecture, the separate control and measurement plane takes care of all the control and measurement tasks. LSRs can concentrate on real time traffic switching.

  With distributed architecture, some non-real-time tasks (topology synchronization, policy advertisement, etc.) must also be executed at TEs and compete with real time tasks for CPU time.

- Easier for end-to-end QoS control

  For end-to-end QoS control, a decision maker needs to have knowledge not only the traffic and resource in its own network, but also those in other domains. It introduces a lot of overhead to make the information available to every switch rather than to only a few central control elements.

- Easier to extend for control of other networks

  In the future, when new types of networks are included into service providerÆs network, it is easier to accommodate them into a centralized control and measurement plane because the this plane is an abstract and common plane and all the transport technology specific function is kept by each TE.

4.3. Distributed Architecture

In a distributed architecture, each TE communicates with other TEs to collect network topology, resource, and traffic information and performs route computation by itself. Each switch maintains the policy and service profile for all its customers. The advantage of distributed architecture over centralized one are as follows:

- Better survivability

  With distributed architecture, if one TE fails, only the traffic handled by this LSR is affected. The rest of the network will continue to work.

  On the other hand, with centralized architecture, if a CE fails, all the services on the switches under the control of that CE will be impacted.

- Easier to make use of existing routing protocols

  Distributed IP routing (OSPF, IS-IS) has been deployed on TEs; suggestions have been made to extend these protocols to support traffic engineering and QoS [6]. These are fully distributed protocols.

- Complex overall architecture

  With centralized architecture, because the number of network elements that can be managed by one control element is limited by its capacity, multiple control elements may need to be deployed in parallel. Then another centralized component on top of the control elements must be deployed to take care of end-to-end on-demand services. That makes the overall architecture complicated. With distributed architecture, each transport element takes care of its own for all the control capability. No complicated hierarchy is involved.

- Better session setup latency

  With a distributed approach, a tunnel setup message does not have to go CE so the session setup latency is reduced. The same reasoning applies for protection/restoration.

5. Architectural Requirements

The objective is to build a common plane for various underlying transport networks. This plane has a common interface to the underlying transport elements. The high level architectural requirements are described below.

5.1. Independence from Underlying Transport Networks

The underlying transport network can be based on any type of transport technology. The interface between control elements and transport elements must be generic. It must be suitable for any type of networks. The parameters passed at the interface must be abstract and suitable for carrying topology, resource, traffic, and policy decision information for any type of networks. It is up to the transport element to map its technology specific presentation of above to the standard interface.

The architecture must be extensible to support more functions and other network transport elements.

5.2. Scalable to Very Large Networks

Contemporary public networks are growing very fast with respect to network size and traffic volume. The architecture must be designed to work with small network consisting a few tens of TEs to a big network consisting of a few thousands TEs.

5.3. Flexibility

With different transport networks, and at different stages of deployment of the architecture, there may be different solutions to the same issue. The architecture must be flexible in adopting different mechanisms.

For example, the measurement results can be obtained in different ways: using the measurement protocol as discussed in this draft, using OSPF-TE when it is widely deployed in the network, or using MIBs uploading. The architecture must be flexible to allow different mechanisms to be easily plugged in.

In another example, a particular MPLS subnet may have its own built-in traffic engineering mechanism. The architecture must allow the transport elements to choose which mechanism (at control element or of its own) to use.

In another scenario, a third party policy engine is already deployed in the service providerÆs network, this architecture must allow the policy engine to be plugged into the control plane.

Another example is that at the early deployment stage, some network parameters (e.g., CE to TE association) may be statically provisioned and at advanced stage they may be obtained by protocol (e.g., auto-discovery). As for provisioning, the plane should also provide sufficient configuration options so that a network administrator can tailor the system to a particular environment.

5.4. Steady State Operation

The architecture must be such that the entire control system can reach steady state fast. For example, this requires the routing computation be relatively independent of dynamically changeable parameters.

5.5. Minimized Overhead

This architecture should not introduce significant transport and processing overhead. For this purpose, the control protocols should be as simple as possible. The amount of information should be minimized and the format to represent the information should be efficient.

5.6. Minimized Impact on Real-Time Performance

With more functionality introduced into the control plane, session setup latency will be degraded. The architecture must be designed so that this impact is minimized.

5.7. Simplicity

The system should be as simple as possible, consistent with the intended applications. The system should be relatively easy to use (i.e., clean, convenient, and intuitive user interfaces).

Simplicity in user interface does not necessarily imply that the TE system will use naive algorithms. Even when complex algorithms and internal structures are used, such complexities should be hidden as much as possible from the network administrator through the user interface.

5.8. Survivability

It is critical for an operational network to recover promptly from network failures and to maintain the required QoS for existing services. Survivability generally mandates introducing redundancy into the architecture, design, and operation of networks.

Survivability can be addressed at the device level by developing network elements that are more reliable; and at the network level by incorporating redundancy into the architecture, design, and operation of networks. This draft requires that a philosophy of robustness and survivability should be adopted in the architecture, design, and operation of control and measurement plane.

5.9. Interoperability

Whenever feasible, control and measurement systems and their components should be developed with open standards based interfaces to allow interoperation with other systems and components.

6. Proposed High Level Architecture

Based on the functions described in Sections 4 and 5, the proposed architecture for the control plane is described in this section.

6.1. Architecture Overview

As illustrated in Figure 1, the control and measurement plane is separated from and built on top of the transport network. The entire controlled network is divided into service domains. One domain is under management of a single service provider and the network elements within one domain share consistent network view and policy view. The entire controlled service domain consists of one or more CEs and multiple TEs. Each TE is under the control of one CE. One CE can control multiple TEs.

```
                                   +------------------+
                                   |      Clearing    |
                                   |       House      |
```

```
                                      +-------------------+
                                        A              A
                                        |              |
                                        |(R5)          |(R5)
    +-----------------------------------|----+    +----|-------------+
    |                                   |    |    |    |             |
    |                                   V    |    |    V             |
    |  +-------------+      +-------------+  |    | +-------------+  |
    |  |   Control   |<---->|   Control   |<------->|   Control   |  |
    |  |   Element   | (R3) |   Element   |  |(R4)| |   Element   |  |
    |  +-------------+      +-------------+  |    | +-------------+  |
    |       A   A                A   A       |    |      A   A       |
    |       |   |                |   |       |    |      |   |       |
    |   (R1)|   |(R2)        (R1)|   |(R2)   |    |  (R1)|   |(R2)   |
    |       |   |                |   |       |    |      |   |       |
    |       V   V                V   V       |    |      V   V       |
    |  +-------------+      +-------------+  |    | +-------------+  |
    |  |  Transport  |      |  Transport  |  |    | |  Transport  |  |
   ....|   Element   |......|   Element   |.........|   Element   |....
    |  +-------------+      +-------------+  |    | +-------------+  |
    |                                        |    |                  |
    |               (domain 1)               |    |    (domain 2)    |
    +----------------------------------------+    +------------------+
```

Figure 1: CCAMP Architecture & Reference Points

If there is more than one CE in the network, the CEs are connected via associations in either a partial or full mesh. The CEs and their associations together form a CE network. The links between CEs are logical links, or associations. These CEs and their associations are provisioned so that reachability (directly or indirectly) exists between any pair of CEs.

In case of multiple CEs, each CE is responsible for managing a number of TEs. Any TE is controlled by only one CE at any time. One TE may have associations with more than one controller for protection purpose.

Each TE communicates with its CE using two protocols: a control protocol and a measurement protocol. The control protocol is used for the CE to send policy decisions, tunnel setup information (e.g., source routing path), and traffic filters for mapping incoming traffic to the tunnel to be setup. The measurement protocol is used for TEs to report network status and resource information to the CE.

Because multiple CEs may be deployed in one service domain, these CEs need to communicate to each other so that they have the consistent view about the service profiles of customers, policies, network resource and status. For this purpose, these CEs need to speak another protocol with its peers: inter-CE protocol.

The inter-CE protocol consists of two parts: intra-domain part and inter-domain part. For the intra-domain part, the protocol allows the CEs to share all policy and network information with each other. No security check or policy filtering logic is required. While for the inter-domain part, only the customer level policy and service capability is exchanged. Security mechanisms must be applied to inter-domain communication. Only the border CE needs to support both intra-domain part and inter-domain part. The internal CEs only need to support inter-domain part.

In this architecture, we propose direct communication between CEs for inter-domain communication. Another alternative is to exchange policy information between service domains via a Clearing House. This alternative, however, is not addressed in this draft.

6.2. Single Protocol or Separate Protocols

From a functional point of view, the system requires two protocols, control protocol and measurement protocol. These two protocols can be kept separate or combined together.

With a single protocol, only one association needs to be established and maintained between each TE and its CE. The messages exchanged can be reduced because some information of two protocols can be carried in a single message. This reduces the protocol overhead.

With separate protocols, it is easy to develop each protocol independently and to incorporate other protocols into the architecture. For example, when TE-OSPF is widely deployed, it can be used for measurement and reporting purpose, and therefore no new measurement protocol is needed.

Although more investigation is required before reaching an agreement on single protocol or separate protocol, in this draft, we describe the requirements separately.

6.3. Reference Points

In this architecture illustrated in Figure 1, the following reference points are defined.

- Reference Point R1

  Policy control information flow between CE and TE is captured in R1. The information across this point communicates policy-based session setup request and decision, traffic filter decision, and policy installation request between CE and TE. This protocol is a client-server protocol with the TE as client and the CE as the server.

- Reference Point R2

  Transport Elements uploading information and/or measurement information flow between CE and TE is captured in this reference point. The information across this reference point communicates TE topology, resource, network status and measurement information. The protocol used at this interface is client-server protocol with TE as client and CE as server.

- Reference Point R3

  Information flowing between two internal CEs is captured in this point. The information across this reference point communicates network topology, resource, and status information of the portion of the network and TEs under its control. It also communicates policy information and service capability information learned from other domains. The protocol used at this point is a peer protocol.

- Reference Point R4

  Information flowing between two border CEs in different domains is captured in this point. The information across this reference point communicates pricing, authorization, usage, policy, and service capability information. The policy information flowing at this point includes customer specific policy, service specific policy, and resource policy. It is used for advertising, negotiating and notifying policy information. The policy information across this point can be either both globally

  available policy information and peer domain specific policy information (if clearing house is not available) or only peer domain specific policy (if clearing house is used for global available policy information). This protocol is a peer protocol.

- Reference Point R5

  The information flowing between CE and Clearinghouse is captured in this reference point. The information flowing across this reference point is inter-domain pricing, authorization, and usage information as well as customer, service, and resource specific policy. The protocol used at this point is client-server protocol with CE as client and CH as server.

7. TE Functional Requirements

The underlying network could be MPLS network, ATM network, optical switching network, etc. or any combination of the above. However, the following assumptions are made about the network and the TE:

- The TEs are connected to each other in a arbitrary topology (meshed, star, tree, etc)

- One TE can have different types of interfaces: different MPLS capable interfaces, or non-MPLS interfaces.

- Every MPLS capable interface has IP address, implemented IP stack, running IP routing, running MPLS signaling (e.g., LDP and CR-LDP)

- TEs that have both MPLS and non-MPLS interfaces are able to do traffic mapping between non-MPLS traffic (packets, time slots, lambdas, physical interfaces) and MPLS traffic according to a traffic classifier

- TEs that have different types of MPLS interface are able to map between those interfaces

- Every TE is able to perform resource reservation and release

- Every TE is able to collect network topology and status information and report it to CE

- Every TE is able to perform performance measurement and report the results to CE.

- Every TE is able to collect and report network resource usage information and report it to CE

- Every TE supports the control protocol and measurement protocol as described in this draft, including establishing and maintaining association with CE, generating, receiving, and processing protocol messages, switchover to a backup CE in case that the primary CE failure is detected, etc.

- Every TE must support either provisioned CE assignment or CE auto-discovery.

- Every TE is able to enforce policy decision it received from CE

8. CE Functional Requirements

The CE is the core component of this architecture. It must provide the following capabilities.

8.1. Association Establishment and Management

These requirements for a CE to establish and maintain associations with TEs and its CE peers are addressed by each protocol in seubsequent sections. For the purpose of completion, they are also listed here.

- It is able to establish and maintain association with its intra-domain peers and inter-domain peers

- It is able to monitor whether its peers are alive

- It is able to delete the association with a peer when the peer fails or the peer relationship is removed by operator

- It must support auto-discovery of CE by TE

- When a new TE added into the network, the CE is able to coordinate with other CEs to decide which CE is to control the new TE.

- It is able to establish and maintain associations with the TEs under its control

- It is able to reassign a TE under its control to another CE and communicate this reassignment with TE and CE.

- It is able to detect its peerÆs failure or its TEÆs failure and close the association

8.2. Tunnel Management

- Tunnel routing involves the selection of a path from the originating node to the destination node in a network. CE should support time-dependent routing and state-dependent routing.

- The architecture also allows other routing engine or routing mechanisms to be plugged in. In this case, the CE must also be

able to decide which routing mechanism to be used for a particular tunnel setup request according to its local policy.

- It is able to compute and setup a path according to the traffic and QoS requirements.

- It is able to manage routing table from different route mechanisms and perform route lookup based on its local policy.

- It is able to instruct TEs to establish tunnels according to the path it specified

- It is able to maintain all the information related to each tunnel originating from the controlled TE. The tunnel could be any type of point-to-point, point-to-multipoint or multipoint-to-point.

- It is able to instruct TEs to modify an established tunnel without affecting existing traffic

- It is able to delete a tunnel upon request or due to network failure

8.3. Resource Management

- It is able to store the network topology and resource formation in a way that it is easy to be advertised and easy to be used for route computation

- It must maintain the network resources information for any type of interfaces

- It is able to perform admission control upon a request for tunnel establishment based on resource availability, setup requirements and its local policy

- It must be able to update the resource utilization of the underlying network upon tunnel setup or release

- It must be able to update its resource utilization information upon report from TE or other CEs

- It must be able to advertise any topology change reported by TEs under its control to other CEs within the same domain

- It must be able to advertise any resource utilization change calculated by itself or reported by TEs to other CEs within the same domain

8.4. QoS policy capability

- It must be able to make Policy Decision upon the request from TE, other network components such as SIP proxy server, or provision

- It must support QoS policy management. It is able to create and maintain a policy database in a format that is easy to update and easy to apply.

- It must be able to communicate with a separate policy repository using a standard protocol

- It must support both policy provisioning and policy outsourcing modes as defined in COPS [7]. For provisioning mode, it is able to install polices to the TEs that are under its control.

- It must support policy management so that the service provider is able to create, modify or delete policy via a standard user interface (CLI, GUI).

- It must be able to distribute new policy items to its intra-domain peers. The new policy could be created by an operator, or learned from its neighbor domain peers.

- It must be able to advertise its policy to other service domains according to its filtering policy.

- It must be able to negotiate the service, pricing, and customer policy with other service domains.

- It must support various types of policies.

- The policy framework must be extensible to include other policy in addition to QoS policy

8.5. Service provisioning and control

- It must be able to interact with Service Management System (SMS) to create, modify, and delete services

- It must be able to interact with SMS to provision services

- It must able to provision services based on Service Level Specification (SLS) with its access customers

- It must be able to provision services based on Service Level Agreement (SLA) with its peer service providers

- It should be able to exchange SLA with other service domains

8.6. OAM&P

A CE must be able to perform the following standard OAMP functions:

- Operation management: load/boot, software/hardware upgrade, capability to enable or disable resource and/or features.

- Configuration management: provisioning and configuring components and applications

- Performance management: performance monitoring, data collecting and analysis

- Accounting management: gathering statistics and usage information for accounting or billing purposes

- Fault management: problems/symptoms report and handling

8.7. Robustness

The control architecture must provide three level protections:

- Network level protection: When one CE fails, other CEs will automatically take care of all the TEs under failed CEÆs control.

- Link level protection: Physical or logical link failure should not cause the association termination.

9. General Protocol Requirements

In the control architecture described in Section 6, three protocols have been defined. They are control protocol, measurement protocol, and inter-CE protocol. The inter-CE protocol is divided into two portions: intra-domain part and inter-domain part. This section discusses general protocol requirements that apply to all three protocols.

9.1. Transport Network Assumptions

The protocols must assume that the underlying network:

- May be over large shared networks.

- Assures reliable delivery of messages.

- Does not guarantee message delivery delay.

- Does not guarantee ordering of messages: sequenced delivery of messages associated with the same source of events is not assumed.

9.2. Association requirements

For any of the three protocols to function, an association must be established between two parties. The following are association related requirements.

Each protocol must

- be able to establish, maintain and terminate association between two communication parties (between CE and TE or between two CEs)

- allow the association to be specified by provisioning

- allow the association between CE and TE to be established by auto-discovery

  Each TE is able to discover and registered with CE automatically. CEs should be able to decide which CE should control the discovered TE.

- provide a method for the TE to inform a CE that the it received a command that is under the control of a different CE

- support a method for the TE to inform a CE that it cannot handle any more requests

- allow a CE to terminate an association and redirect a TE to another CE

9.3. Protocol performance requirements

Each of the three protocols:

- should minimize message exchanges between TE and CE and between CEs

- should make efficient use of the underlying transport mechanism

  For example, protocol PDU sizes vs. transport MTU sizes needs to be considered in designing the protocols.

- must not contain inherent architectural or signaling constraints that would limit peak throughput rates or the number of TEs a CE can control

- should allow for default/provisioned settings so that commands need only contain non-default parameters

9.4. Transport Requirements

Each of the three protocols:

- must provide the ability to abort delivery of obsolete messages at the sending end if their transmission has not been successfully completed

  For example, aborting a command that has been overtaken by events.

- should support priority messages

  The protocol should allow a command precedence to allow priority messages to supercede non-priority messages.

- should support large fan-out at the CE

- must provide a way for one entity to correlate commands and responses with the other entity

- must provide a reason for any command failure

- must assure that loss of a packet not stall messages not related to the message(s) contained in the packet lost

9.5. Security requirements

Security mechanisms may be specified as provided in underlying transport mechanisms, such as IPSEC. The protocol, or such mechanisms, must:

- allow for mutual authentication at the start of a CE-TE association, especially for inter-domain associations

- allow for preservation of the control messages once the association has been established

- allow for optional confidentiality protection of control messages

- allow a choice in the algorithm to be used

- across untrusted domains in a secure fashion

- define mechanisms to mitigate denial of service attacks

In addition, it may be desirable for the protocol to be able to pass through commonly used firewalls.

9.6. Other Requirements

Each of the three protocols must:

- support multiple operations to be invoked in one message and treated as a single transaction

- be both modular and extensible

  Not all implementations may wish to support all of the possible extensions for the protocol. This will permit lightweight implementations for specialized tasks where processing resources are constrained. This could be accomplished by defining particular profiles for particular uses of the protocol.

- be flexible in allocation of intelligence between CE and TE

  For example, an CE may want to allow the TE to assign particular

TE resources in some implementations, while in others, the CE may want to be the one to assign TE resources for use. In another example, CE may allow TE to do path computation in some implementations, while in others, the CE does the path computation by itself and the TE must take that path.

- support scalability from very small to very large TEs

  The protocol must support TEs with capacity ranging from one to millions of connections.

- support scalability from very small to very large CE span of control (i.e. The protocol should allow CEs to control from one to a few thousands of TEs)

- support the needs of an edge TE that supports small number of tunnels, and the needs of large TEs supporting tens of thousands of tunnels

  Protocol mechanisms favoring one extreme or the other should be minimized in favor of more general-purpose mechanism applicable to a wide range of TEs. Where special purpose mechanisms are proposed to optimize a subset of implementations, such mechanisms should be defined as optional, and should have minimal impact on the rest of the protocol.

- facilitate TE and CE version upgrades independently of one another (the protocol must include a version identifier in the initial message exchange)

- facilitate the discovery of the protocol capabilities of the one entity to the other

- specify commands as optional (can be ignored) or mandatory (must be accepted or rejected)

- within a command, specify parameters as optional (can be ignored) or mandatory (must be accepted or rejected).

10. Control Protocol Requirements

The control protocol is running between CE and controlled TEs. In addition to the general protocol requirements listed in Section 9, this protocol must meet the following requirements.

10.1. Resource requirements

The control protocol must

- support resource allocation for use by a particular tunnel and support its subsequent release at various granularities

- allow modification of resource reservation without affecting existing services

- allow release in a single exchange of messages, of all resources associated with a particular set of connectivity and/or association between a given number of terminations

- not require the TE to maintain a sense of future time: a resource allocation/reservation remains in effect until explicitly released by the CE

- provide a method for the CE to request that the TE to release all resources currently in use, or reserved, for any or all tunnels

- provide a way for the TE to indicate that it was unable to perform a requested action because of resource exhaustion, or because of temporary resource unavailability

10.2. Tunnel Requirements

The control protocol must:

- support establishment, modification and deletion of connections involving any types of layer 1 and layer 2 networks and any combinations

- support establishment, modification and deletion of tunnels involving any amount of resource reservation

- support unidirectional, symmetric bi-directional, and asymmetric bi-directional tunnels

- support point-to-point, point-to-multiple, and multiple-to-point tunnels

- allow TE to request CE for a tunnel setup (including admission control, policy control, path computation, etc.)

- allow CE to specify the entire path or partial path for a tunnel

- allow the specification of traffic filter (classifier) for the tunnel with the granularity of the traffic filter as following:

  PQ (Port Quadruples): same IP source address prefix, destination address prefix, TTL, IP, protocol and TCP/UDP source/destination ports

  PQT (Port Quadruples with TOS): same IP source address prefix, destination address prefix, TTL, IP, protocol and TCP/UDP source/destination ports, and same IP header TOS field (including precedence and TOS bits)

  HP (Host Pair): same specific IP source and destination addresses

  HPT (Host Pair with TOS): same specific IP source and destination addresses with same IP header ToS field

  NP (Network Pair): same IP source and destination address prefix (variable length)

  DN (Destination Network): same IP destination network address prefix (variable length)

  ER (Egress Router): same egress router ID

  NAS (Next-hop AS): same next-hop AS number

  DAS (Destination AS): same destination AS number

  SST (Source Specific Tree): same source address and multicast group

  SMT (shared multicast Tree): same multicast group address

  Same source and destination IP address with same DiffServ PHB

  Same source and destination IP address with same RSVP flow ID

- allow dynamic modification of traffic filter to add or remove any flows to/from the tunnel without affecting existing service

- support rerouting of an existing tunnel to a different path

- allow CE to specify the priority of the tunnel

- allow the TE to report events such as resource reservation and tunnel setup completion

10.3. Event Processing and Scripting

The control protocol must allow CE to enable/disable monitoring for specific supervision events

10.4. Policy Requirements

The control protocol must:

- allow TE to communicate policy request (usually together with tunnel setup request) to CE

- allow CE to communicate policy decision information to TE (usually together with explicit path information for the tunnel)

- allow CE to install policy to TE

- allow CE to modify the installed policy at TE

10.5. Media transformation Requirements

The control protocol must allow CE to instruct TE about mediation/adaptation (or traffic mapping) of flows between different types of transport interfaces.

10.6. Operation/management Requirements

The control protocol must:

- support detection and recovery from loss of contact due to failure/congestion of communication links or due to CE or TE failure

- support detection and recovery from loss of synchronized view of resource and tunnel states between CE and TEs (e.g. through the use of audits)

- provide a means for CE and TE to provide each other with booting and reboot indications, and what the TE's configuration is

- permit more than one backup CE and provide an orderly way for the TE to contact one of its backup CEs

- provide for an orderly switch back to the primary CE after it recovers

- provide a mechanism so that when a CE fails, tunnels already established can be maintained

  The protocol does not have to provide a capability to maintain tunnels in the process of being connected, but not actually connected when the failure occurs.

10.7. Error Control

The control protocol must:

- allow for the TE to report reasons for abnormal failure of lower layer tunnels

- allow the TE to notify the CE that an interface was terminated and communicate a reason when an interface is taken out-of-service unilaterally by the TE due to abnormal events.

- allow the CE to acknowledge that some resource has been taken out-of-service.

- allow the TE to request the CE to release some resource and communicate a reason.

- allow the CE to specify its decision to take resource down, leave it as is or modify it.

10.8. Management Requirements

The control protocol must:

- provide information on:

  . mapping between resources and supporting physical entities

  . statistics on quality of service on the control and signaling interfaces

  . statistics required for traffic engineering within the TE

- allow the TE to provide to the CE all information the CE needs to provide in its MIB

- allow the TE to provide the number of policy query, execution, and advertisements

11. Measurement Protocol Requirements

The measurement protocol also runs between CE and TEs. In addition to the general protocol requirements listed in Section 9, this protocol must meet the following requirements.

11.1. Topology and resource information

The following information must be reported to the CE immediately after a CE-TE association is established, whenever a network

topology changed (node or link added into or removed from the network), and upon the request from CE:

- TE must report underlying network topology information. Each TE is only responsible for reporting its own interfaces.

- For each interface TE reports interface type (e.g., pure IP, RSVP, DiffServ, PSC, TDM, LSC, or FSC), local and remote IP addresses, and total network resource allocated to be used by this Control System in both directions.

- For each interface, TE reports bandwidth reservation granularity (e.g., number of bytes, slot rate, lambda capacity).

- For each interface, TE reports performance parameters including propagation delay and packet loss rate.

The following information must be reported upon request from CE or whenever a pre-specified network resource threshold is crossed due to establishment of new tunnels or release or modification of an existing tunnels:

- For a successfully established tunnel, the originating TE reports the committed resource reservation.

- For tunnel release not triggered by CE, TE reports resource release by indicating to CE the tunnel ID of the tunnel that has been released.

11.2. TE Capability Information

The protocol must allow TE to indicate to CE its capabilities as listed below.

- Whether it is an internal TE or border TE

- Whether it is able to perform tunnel merge

- What kinds of traffic mapping it supports

- Whether it is able to setup uni-directional, synchronous bi-directional, or asynchronous bi-directional tunnels

11.3. Status Information

The measurement protocol must allow the CE to request and the TE to report the following:

- status and all information about the interface when a new interface is added or activated.

- link failure or deactivation

- congestion status in the network

11.4. Tunnel Information

In most cases, CE will keep all the tunnel related information. There may be cases CE needs to request that information from the TE. The protocol must allow:

- CE to request and TE to report tunnel related information (source and destination IP address, traffic filter, merge point, etc.)

- CE to request and TE to report all tunnels associated with a particular interface.

11.5. Performance Information

The protocol must allow the CE to request and TE to report performance information such as round-trip delay, packet loss rate, etc. for a particular tunnel or a particular interface.

11.6. Statistics Information

In most cases, the CE keeps all the statistics information for all the TEs under its control. However, there may be cases that CE needs to request the information from each a particular TE. So the protocol must allow the CE to request and TE to report the following statistics information:

- the number of tunnels that meet certain requirements (on the node, on a particular interface, to a particular IP address, duration exceeding 10 min, etc.)

- the duration of a particular tunnel

- the whole profile of a particular tunnel

11.7. Accounting Requirements

The measurement protocol must:

- support a common identifier to mark resources related to one tunnel

- support collection of specified accounting information from TEs

- provide the mechanism for the CE to specify that the TE report accounting information automatically at end of a session, in mid-

session upon request, at specific time intervals as specified by the TEs and at unit usage thresholds as specified by the CE

- specifically support collection of:

  . Start and stop time, by media flow

  . Volume of content carried (e.g. number of packets/cells transmitted, number received with and without error, inter-arrival jitter), by media flow

- allow the CE to have some control over which statistics are reported, to enable it to manage the amount of information transferred

11.8. Event Processing and Scripting

The measurement protocol must allow CE to enable/disable monitoring for specific supervision events.

11.9. Operation/Management Requirements

The measurement protocol must:

- Support detection and recovery from loss of contact due to failure/congestion of communication links or due to CE or TE failure.

- Support detection and recovery from loss of synchronized view of resource and connection states between CE and TEs (e.g. through the use of audits).

- Provide a means for CE and TE to provide each other with booting and reboot indications, and what the TE's configuration is.

- Permit more than one backup CE and provide an orderly way for the TE to contact one of its backups.

- Provide for an orderly switch back to the primary CE after it recovers.

- Provide a mechanism so that when a CE fails, tunnels already established can be maintained. The protocol does not have to provide a capability to maintain tunnels in the process of being connected, but not actually connected when the failure occurs.

11.10. Error Control

The measurement protocol must

- allow for the TE to report reasons for abnormal failure of lower layer tunnels

- allow the TE to notify the CE that an interface was terminated and communicate a reason when an interface is taken out-of-service unilaterally by the TE due to abnormal events

- allow the CE to acknowledge that some resource has been taken out-of-service

12. Inter-CE Protocol Requirements

This protocol consists of two portions: internal part and external portion. There are some common requirements that apply to both internal and external portion. Some other requirements are specific for internal portion or external portion.

12.1. Common requirements

The following requirements apply for both internal portion and external portion. Both inter-CE protocol must CEs, both in the same domain and in different domains, to:

- support arbitrary network topology of Controllers (meshed, star, tree, etc.)

- allow the Controller peer relationship be provisioned

- support automatic peer discovery

- support detection and recovery from loss of contact due to failure/congestion of communication links or due to Controller failure

- support detection and recovery from loss of synchronized view of resource and connection states between Controllers

- provide a mechanism so that when a Controller fails, connections already established can be maintained

  The protocol does not have to provide a capability to maintain connections in the process of being connected, but not actually connected when the failure occurs.

12.2. Internal capability

The following information is exchange between CEs so that all the CEs within a domain have a consistent view of the network. The inter-CE protocol must allow CEs in the same domain to:

- exchange topology information

- exchange network resource information

- exchange network status information

- exchange tunnel and its allocated resource information

- advertise policy information within the service domain

- negotiate the new assignment of TEs from one CE to another

12.3. External capability

The inter-CE protocol must allow CEs in different domains to:

- exchange service level policy

- exchange pricing and usage information

- exchange performance measurements of their service domain

- exchange Service Level Agreement (SLA)

13. Security Considerations

Security requirements for the protocols are listed in Section 10.5.

14. References

1 Bradner, S., "The Internet Standards Process -- Revision 3", BCP 9, RFC 2026, October 1996.

2 Cuervo, F., et al, "Megaco Protocol Version 1.0", RFC 3015, November 2000.

3 Moy, J., "OSPF Version 2", STD 54, RFC 2328, April 1998.

4 Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

5 Ashwood-Smith, P., et al, "Generalized MPLS - Signaling Functional Description," Internet Draft, <draft-ietf-mpls-generalized-signaling-01.txt>, work in progress.

6 Katz, D., and Yeung, D., "Traffic Engineering Extensions to OSPF," Internet Draft, <draft-katz-yeung-ospf-traffic-03.txt>, work in progress.

7 Durham, D., et al, "The COPS (Common Open Policy Service) Protocol", RFC 2748, January 2000.

15. Author's Addresses

Jianping Jiang
SS8 Networks Inc.
55 Commerce Valley Drive West
Toronto, ON L3T 7B9
Canada
Phone: +1 905 889 5900
Email: jainping@ss8.com

Dave Walker
SS8 Networks Inc.
495 March Road
Ottawa, ON K2K 3G1
Canada
Phone: +1 613 592 2100
Email: drwalker@ss8.com

Jianli Wang
SS8 Networks Inc.
495 March Road
Ottawa, ON K2K 3G1
Canada
Phone: +1 613 592 2100
Email: jianli@ss8.com

Full Copyright Statement