CHAPTER 29

# Chapter Goals

- Describe the need for DLSw.
- Know the advantages of DLSw over source-route bridging.
- Specify the transport protocol between DLSw switches.
- Understand the basic structure of DLSw.
- Recognize DLSw processes by name and function.
- Understand the circuit establishment process.

# Data-Link Switching

# Background

*Data-link switching (DLSw)* provides a means of transporting IBM Systems Network Architecture (SNA) and network basic input/output system (NetBIOS) traffic over an IP network. It serves as an alternative to *source-route bridging (SRB)*, a protocol for transporting SNA and NetBIOS traffic in Token Ring environments that was widely deployed before the introduction of DLSw. In general, DLSw addresses some of the shortcomings of SRB for certain communication requirements—particularly in WAN implementations. This chapter contrasts DLSw with SRB, summarizes underlying protocols, and provides a synopsis of normal protocol operations.

DLSw initially emerged as a proprietary IBM solution in 1992. It was first submitted to the IETF as RFC 1434 in 1993. DLSw is now documented in detail by IETF RFC 1795, which was submitted in April 1995. DLSw was jointly developed by the Advanced Peer-to-Peer Networking (APPN) Implementors Workshop (AIW) and the Data-Link Switching Related Interest Group (DLSw RIG).

RFC 1795 describes three primary functions of DLSw:

- The Switch-to-Switch Protocol (SSP) is the protocol maintained between two DLSw nodes or routers.
- The termination of SNA data-link control (DLC) connections helps to reduce the likelihood of link layer timeouts across WANs.
- The local mapping of DLC connections to a DLSw circuit.

Each of these functions is discussed in detail in this chapter.

In 1997, the IETF released DLSw version 2 (RFC 2166) which provides enhancements to RFC 1795 document. The additional features include these:

- IP multicast
- UDP unicast responses to DLSw broadcasts
- Enhanced peer-on-demand routing
- Expedited TCP connections

Each of these features enables DLSw as a scalable technology over WANs. In DLSw Version 1, transactions occur with TCP. As a result, many operations in a DLSw environment consumed circuits between peers. For example, a multicast required multiple TCP connections from the source to each peer. With DLSw Version 2, multicast is distributed using unreliable transport following traditional multicast methods.

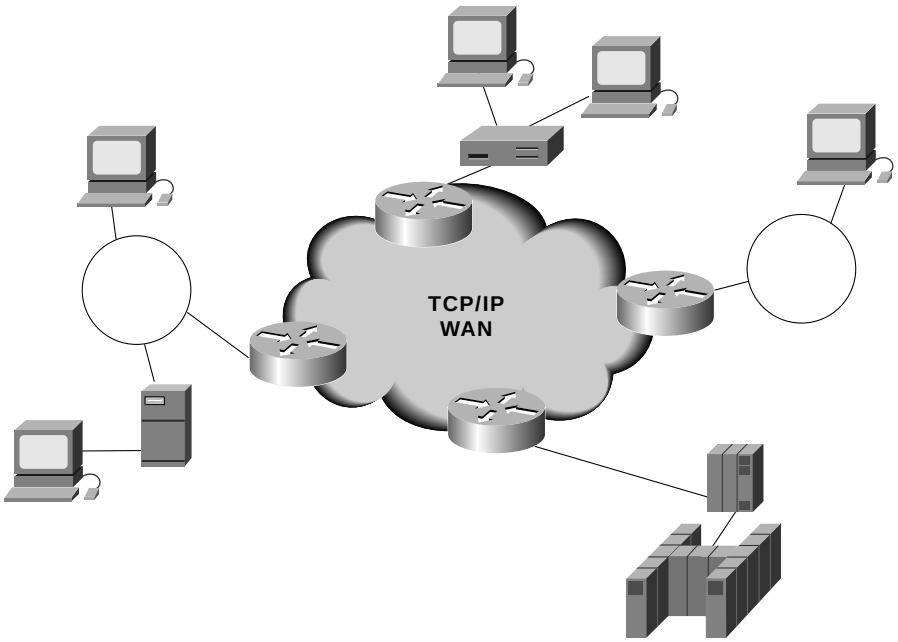
Note that RFC 2166 does not supercede 1795, but it adds functionality and maintains backward compatibility.

Cisco supports a third version of DLSw called DLSw+. DLSw+ predates DLSw Version 2 and provides even further enhancements to basic DLSw. DLSw+ is fully compliant with RFC 1795. The enhancements may be used when both peers are Cisco devices running DLSw+.

This chapter focuses on the basic function of DLSw as defined in RFC 1795.

Figure 29-1 illustrates a generalized DLSw environment.

***Figure 29-1 A DLSw Circuit Facilitates SNA Connectivity over an IP WAN***




# DLSw Contrasted with Source-Route Bridging

The principal difference between SRB and DLSw involves support of local termination. SNA and NetBIOS traffic rely on link layer acknowledgments and keepalive messages to ensure the integrity of connections and the delivery of data. For connection-oriented

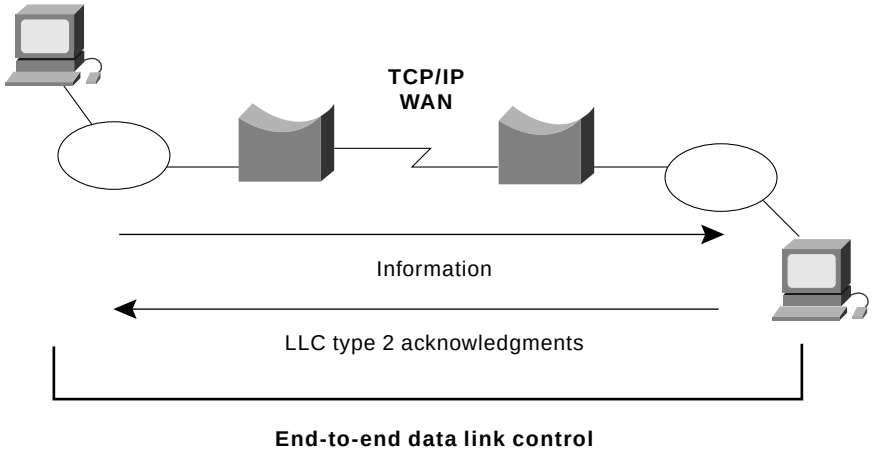
data, the local DLSw node or router terminates data-link control. Therefore, link layer acknowledgments and keepalive messages do not have to traverse a WAN. By contrast, DLC for SRB is handled on an end-to-end basis, which results in increased potential for DLC timeouts over WAN connections.

Although SRB has been a viable solution for many environments, several issues limit its usefulness for transport of SNA and NetBIOS in WAN implementations. Chief among them are the following constraints:

- SRB hop-count limitation of seven hops
- Broadcast traffic handling (from SRB explorer frames or NetBIOS name queries)
- Unnecessary traffic forwarding (acknowledgments and keepalives)
- Lack of flow control and prioritization

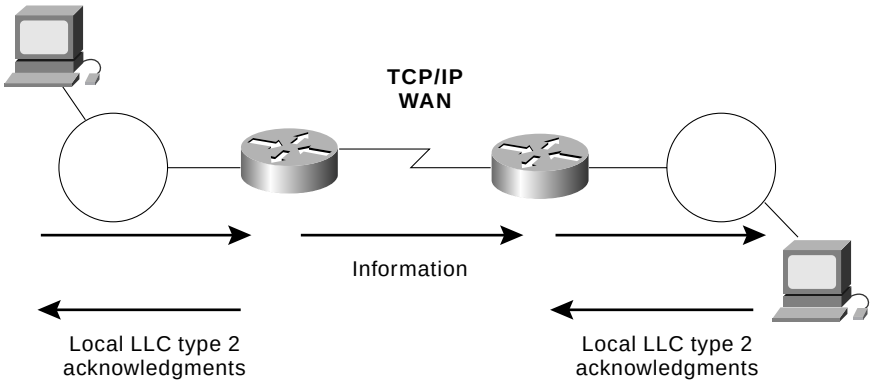
Figure 29-2 illustrates the basic end-to-end nature of an SRB connection over a WAN link.

***Figure 29-2 SRB Provides an End-to-End Connection over an IP WAN***




Local termination of DLC connections by DLSw provides a number of advantages over SRB-based environments. DLSw local termination eliminates the requirement for link layer acknowledgments and keepalive messages to flow across a WAN. In addition, local termination reduces the likelihood of link layer timeouts across WANs. Similarly, DLSw ensures that the broadcast of search frames is controlled by the DLSw when the location of a target system is discovered. Figure 29-3 illustrates the flow of information and the use of local acknowledgment in a DLSw environment.

*Figure 29-3 DLSw Uses Local Acknowledgment to Control Data Flow*

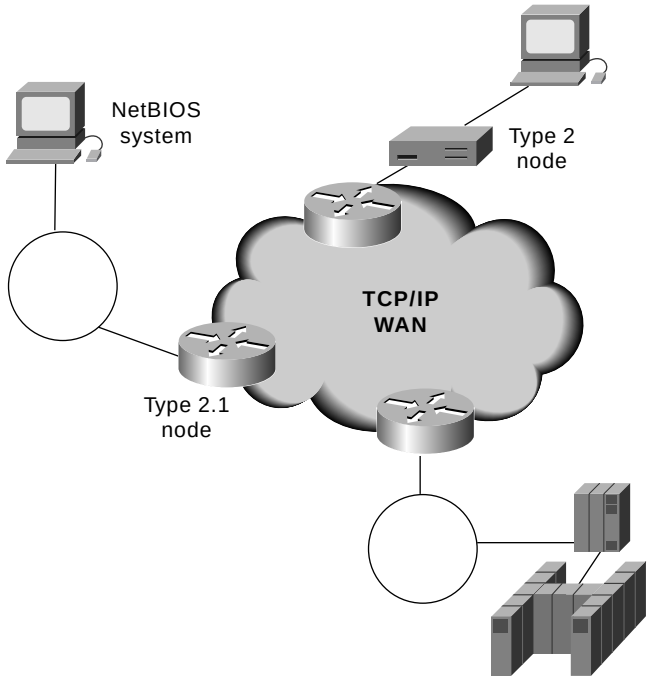



# DLSw SNA Support

One of the advantages inherent in DLSw is that it supplies broader device and media support than previously available with SRB. DLSw accommodates a number of typical SNA environments and provides for IEEE 802.2-compliant LAN support, which includes support for SNA physical unit (PU) 2, PU 2.1, and PU 4 systems and NetBIOS-based systems.

DLSw provides for Synchronous Data Link Control (SDLC) support, covering PU 2 (primary or secondary) and PU 2.1 systems. With SDLC-attached systems, each SDLC PU is presented to the DLSw Switch-to-Switch Protocol (SSP) as a unique Media Access Control (MAC)/service access point (SAP) address pair. With Token Ring–attached systems, a DLSw node appears as a source-route bridge. Remote Token Ring systems accessed via a DLSw node are seen as attached to an adjacent ring. This apparent adjacent ring is known as a virtual ring created within each DLSw node. Figure 29-4 illustrates various IBM nodes connected to a TCP/IP WAN through DLSw devices, which, in this case, are routers.

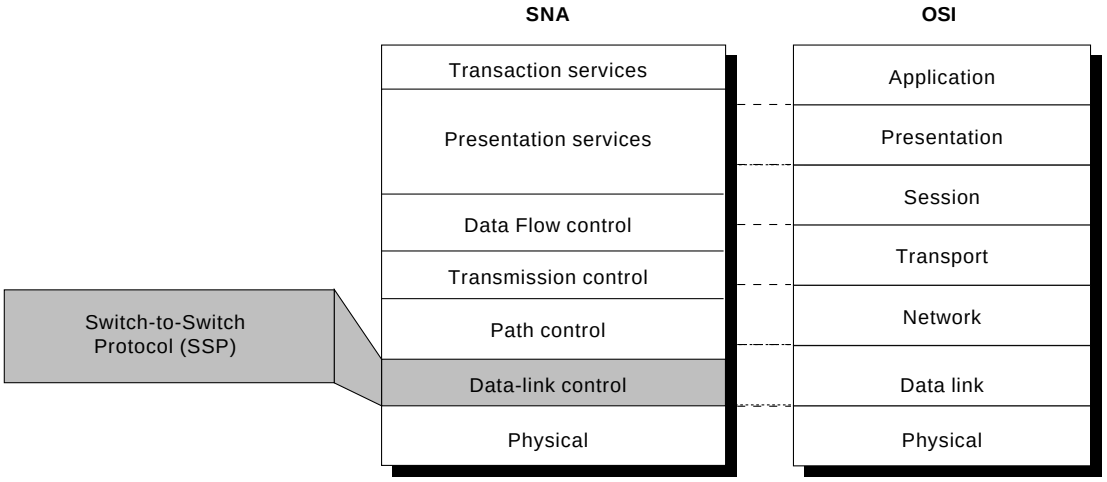
*Figure 29-4 SNA Nodes Connect Through TCP/IP WAN via DLSw*




# DLSw Switch-to-Switch Protocol

*Switch-to-Switch Protocol (SSP)* is a protocol used between DLSw nodes (routers) to establish connections, locate resources, forward data, and handle flow control and error recovery. This is truly the essence of DLSw. In general, SSP does not provide for full routing between nodes because this is generally handled by common routing protocols such as RIP, OSPF, or IGRP/EIGRP. Instead, SSP switches packets at the SNA data link layer. It also encapsulates packets in TCP/IP for transport over IP-based networks and uses TCP as a means of reliable transport between DLSw nodes. Figure 29-5 illustrates where SSP falls in the overall SNA architecture, and shows its relationship to the OSI reference model.

*Figure 29-5 SSP Maps to the Data Link Components of SNA and the OSI Reference Model*




# DLSw Operation

DLSw involves several operational stages. Two DLSw partners establish two TCP connections with each other. TCP connections provide the foundation for the DLSw communication. Because TCP provides for reliable and guaranteed delivery of IP traffic, it ensures the delivery and integrity of the traffic that is being encapsulated in the IP protocol, which, in this case, is SNA and NetBIOS traffic. After a connection is established, the DLSw partners exchange a list of supported capabilities. This is particularly vital when the DLSw partners are manufactured by different vendors. Next, the DLSw partners establish circuits between SNA or NetBIOS end systems, and information frames can flow over the circuit.

## DLSw Processes

The overall DLSw operational process can be broken into three basic components: capabilities exchange, circuit establishment, and flow control. In the context of DLSw, *capabilities exchange* involves the trading of information about capabilities associated with a DLSw session. This exchange of information is negotiated when the session is initiated and during the course of session operations. *Circuit establishment* in DLSw occurs between end systems. It includes locating the target end system and setting up data-link control connections between each end system and its local router. DLSw *flow control* enables the establishment of independent, unidirectional flow control between partners. Each process in discussed in the sections that follow.

### DLSw Capabilities Exchange

DLSw capabilities exchange is based on a switch-to-switch control message that describes the capabilities of the sending data-link switch. A capabilities exchange control message is sent after the switch-to-switch connection is established or during run time, if certain operational parameters that must be communicated to the partner switch have changed. During the capabilities exchange, a number of capabilities are identified and negotiated. Capabilities exchanged between DLSw partners include the following:

- DLSw version number

- Initial pacing window size (receive window size)
- NetBIOS support
- List of supported link SAPs (LSAPs)
- Number of TCP sessions supported
- MAC address lists
- NetBIOS name lists
- Search frames support

## DLSw Circuit Establishment

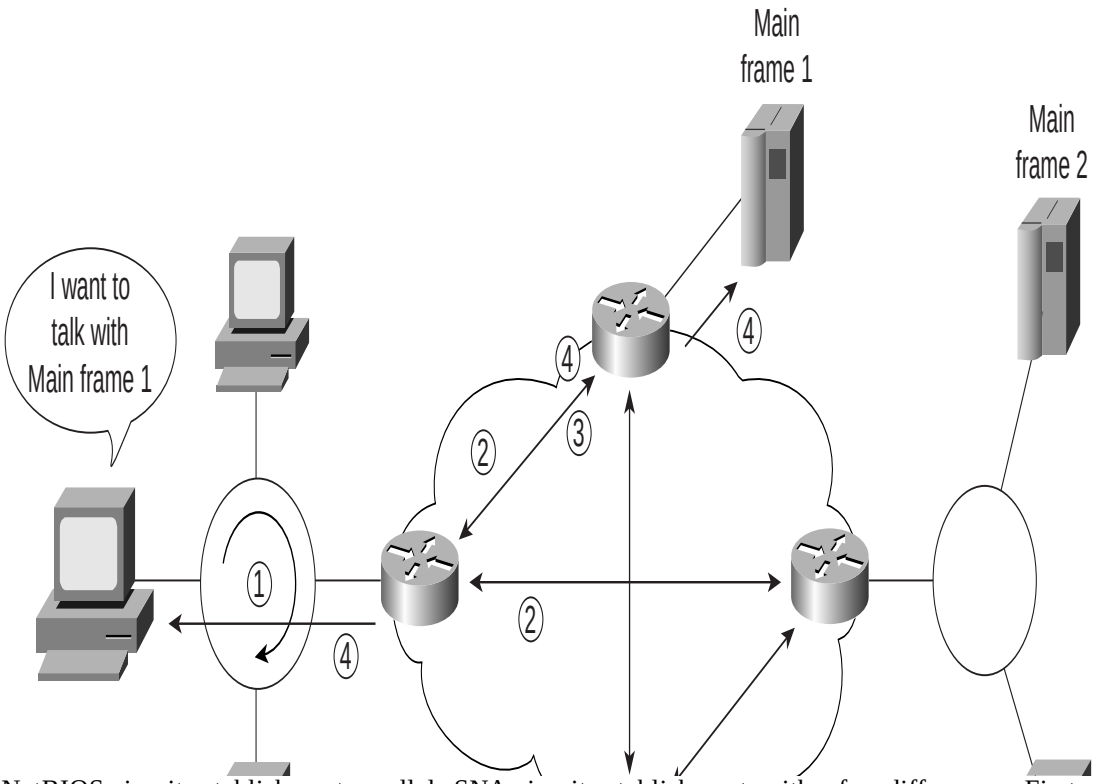
The process of circuit establishment between a pair of end systems in DLSw involves locating the target end system and setting up data-link control (DLC) connections between each end system and its local router. The specifics of circuit establishment differ based on traffic type.

One of the primary functions of DLSw is to provide a transport mechanism for SNA traffic. SNA circuit establishment involves several distinct stages and is illustrated in Figure 29-6.

First, SNA devices on a LAN find other SNA devices by sending an explorer frame with the MAC address of the target SNA device. When a DLSw internetworking node receives an explorer frame, that node sends a *canureach frame* to each of its DLSw partners. The function of this frame is to query each of the DLSw partners to see whether it can locate the device in question. If one of the DLSw partners can reach the specified MAC address, the partner replies with an *icanreach frame*, which indicates that a specific DLSw partner can provide a communications path to the device in question.

After the canureach and icanreach frames have been exchanged, the two DLSw partners establish a circuit that consists of a DLC connection between each router and the locally attached SNA end system (for a total of two connections) and a TCP connection between the DLSw partners. The resulting circuit is uniquely identified by the source and destination circuit IDs. Each SNA DLSw circuit ID includes a MAC address, a link-service access point (LSAP), and the DLC port ID. Circuit priority is negotiated at circuit setup time.

*Figure 29-6 DLSw Circuit Establishment Flow*




NetBIOS circuit establishment parallels SNA circuit establishment, with a few differences. First, with NetBIOS circuit establishment, DLSw nodes send a name query with a NetBIOS name (not a canureach frame specifying a MAC address). Similarly, the DLSw nodes establishing a NetBIOS circuit send a name recognized frame (not an icanreach frame).

## DLSw Flow Control

DLSw flow control involves *adaptive pacing* between DLSw routers. During the flow-control negotiation, two independent, unidirectional flow-control mechanisms are established between DLSw partners. Adaptive pacing employs a windowing mechanism that dynamically adapts to buffer availability. Windows can be incremented, decremented, halved, or reset to zero. This allows the DLSw nodes to control the pace of traffic forwarded through the network to ensure integrity and delivery of all data.

### DLSw Flow-Control Indicators

*Granted units* (the number of units that the sender has permission to send) are incremented with a flow-control indication (one of several possible indicators) from the receiver. DLSw flow control provides for the following indicator functions:

- **Repeat**—Increments granted units by the current window size.
- **Increment**—Increases the window size by 1 and increases granted units by the new window size.
- **Decrement**—Decrements window size by 1 and increments granted units by the new window size.
- **Reset**—Decreases window size to 0 and sets granted units to 0, which stops all transmission in one direction until an increment flow-control indicator is sent.
- **Half**—Cuts the current window size in half and increments granted units by the new window size.

- **Flow**—Control indicators and flow-control acknowledgments can be piggybacked on information frames or sent as independent flow-control messages. Reset indicators are always sent as independent messages.

### Adaptive-Pacing Examples

Examples of adaptive-pacing criteria include buffer availability, transport utilization, outbound queue length, and traffic priority. Examples of how each can be used to influence pacing follow:

- **Buffer availability**—If memory buffers in a DLSw node are critically low, the node can decrement the window size to reduce the flow of traffic. As buffer availability increases, the node then can increase the window size to increase traffic flow between the DLSw partners.
- **Transport utilization**—If the link between two DLSw partners reaches a high level of utilization, the window size can be reduced to lower the level of link utilization and to prevent packet loss between the nodes.
- **Outbound queue length**—Traffic forwarded by a DLSw node typically is placed into an outbound queue, which is a portion of memory dedicated to traffic being forwarded by one device to another. If this queue reaches a specified threshold or perhaps becomes full, the number of granted units can be reduced until the queue utilization is reduced to a satisfactory level.
- **Traffic priority**—One of the unique capabilities of the SSP is its capability to prioritize specific traffic. These priorities are identified by the Circuit Priority field in the DLSw message frame. By providing a varying number of granted units to specific DLSw circuits, the nodes can maintain different levels of priority to each circuit.

# DLSw Message Formats

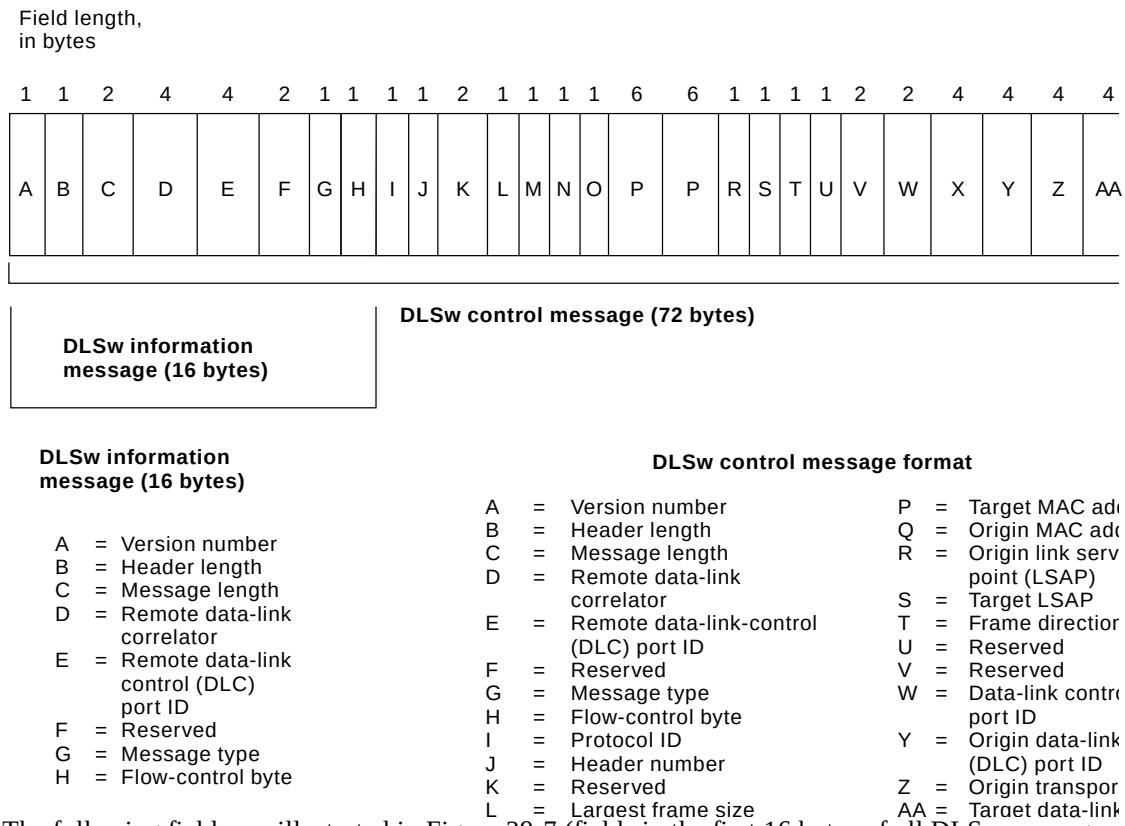
Two message header formats are exchanged between DLSw nodes:

- Control
- Information

The control message header is used for all messages except information frames (Iframes) and independent flow control messages (IFCMs), which are sent in information header format.

Figure 29-7 illustrates the format of the DLSw Control and Information fields. These fields are discussed in detail in the subsequent descriptions.

***Figure 29-7 DLSw Control and Information Frames Have Their First 16 Bytes in Common***




The following fields are illustrated in Figure 29-7 (fields in the first 16 bytes of all DLSw message headers are the same):

- **Version number**—When set to 0x31 (ASCII 1), indicates a decimal value of 49, which identifies this device as utilizing DLSw version 1. This will allow future interoperability between DLSw nodes using different versions of the DLSw standard. Currently, all devices utilize DLSw version 1, so this field will always have the decimal value of 49.
- **Header length**—When set to 0x48 for control messages, indicates a decimal value of 72 bytes. This value is set to 0x10 for information and independent flow control messages, indicating a decimal value of 16 bytes.
- **Message length**—Defines the number of bytes within the data field following the header.
- **Remote data-link correlator**—Works in tandem with the remote DLC port ID to form a 64-bit circuit ID that identifies the DLC circuit within a single DLSw node. The circuit ID is unique in a single DLSw node and is assigned locally. An end-to-end circuit is identified by a pair of circuit IDs that, along with the data-link IDs, uniquely identifies a single end-to-end circuit. Each DLSw node must keep a table of these circuit ID pairs: one for the local end of the circuit and the other for the remote end
of the circuit. The remote data-link correlator is set equal to the target data-link correlator if the Frame Direction field is set to 0x01. It is equal to the origin data-link correlator if the Frame Direction field is set to 0x02.
- **Remote DLC port ID**—Works in tandem with the remote data-link correlator to form a 64-bit circuit ID that identifies the DLC circuit within a single DLSw node. The circuit ID is unique in a single DLSw node and is assigned locally. The end-to-end circuit is identified by a pair of circuit IDs that, along with the data-link IDs, uniquely identifies a single end-to-end circuit. Each DLSw device must keep a table of these circuit ID pairs: one for the local end of the circuit and the other

for the remote end of the circuit. The remote DLC port ID is set equal to the target DLC port ID if the Frame Direction field is set to 0x01. It is equal to the origin DLC port ID if the Frame Direction field is set to 0x02.

- **Message type**—Indicates a specific DLSw message type. The value is specified in two different fields (offset 14 and 23 decimal) of the control message header. Only the first field is used when parsing a received SSP message. The second field is ignored by new implementations on reception, but it is retained for backward compatibility with RFC 1434 implementations and can be used in future versions, if needed.
- **Flow-control byte**—Carries the flow-control indicator, flow-control acknowledgment, and flow-control operator bits.
- **Protocol ID**—When set to 0x42, indicates a decimal value of 66.
- **Header number**—When set to 0x01, indicates a value of 1.
- **Largest frame size**—Carries the largest frame size bits across the DLSw connection. This field is implemented to ensure that the two end stations always negotiate a frame size to be used on a circuit that does not require DLSw partners to resegment frames.
- **SSP flags**—Contains additional information about the SSP message. Flag definitions (bit 7 is the most significant bit, and bit 0 is the least significant bit of the octet) are shown in Table 29-1.

***Table 29-1 SSP Flag Definitions***

| Bit Position | Name | Meaning |
|---|---|---|
| 7 | SSPex | 1 = Explorer message (canureach or icanreach). |
| 6 through 0 | Reserved | None. Reserved fields are set to 0 upon transmission and are ignored upon receipt. |

- **Circuit priority**—Provides for unsupported, low, medium, high, and highest circuit priorities in the 3 low-order bits of this byte. At circuit start time, each circuit endpoint provides priority information to its circuit partner. The initiator of the circuit chooses which circuit priority is effective for the life of the circuit. If the priority is not implemented by the nodes, the unsupported priority is used.
- **Target MAC address**—Combines with the target link SAP, origin MAC address, and origin SAP to define a logical end-to-end association called a data-link ID.
- **Origin MAC address**—Serves as the MAC address of the origin end station.
- **Origin LSAP**—Serves as the SAP of the source device. The SAP is used to logically identify the traffic being transmitted.
- **Target LSAP**—Serves as the SAP of the destination device.
- **Frame direction**—Contains the value 0x01 for frames sent from the origin DLSw to the target DLSw node, or 0x02 for frames sent from the target DLSw to the origin DLSw node.
- **DLC header length**—When set to 0 for SNA and 0x23 for NetBIOS datagrams, indicates a length of 35 bytes. The NetBIOS header includes the following information:
  - Access Control (AC) field
  - Frame Control (FC) field
  - Destination MAC address (DA)
  - Source MAC address (SA)
  - Routing Information (RI) field (padded to 18 bytes)
  - Destination service access point (DSAP)

  - Source SAP (SSAP)
  - LLC control field (UI)

- **Origin DLC port ID**—Works in tandem with the origin data-link correlator to form a 64-bit circuit ID that identifies the DLC circuit within a single DLSw node. The circuit ID is unique in a single DLSw node and is assigned locally. The end-to-end circuit is identified by a pair of circuit IDs that, along with the data-link IDs, uniquely identify a single end-to-end circuit. Each DLSw node must keep a table of these circuit ID pairs: one for the local end of the circuit and one for the remote end of the circuit.
- **Origin data-link correlator**—Works in tandem with the origin DLC port ID to form a 64-bit circuit ID that identifies the DLC circuit within a single DLSw node. The circuit ID is unique in a single DLSw and is assigned locally. The end-to-end circuit is identified by a pair of circuit IDs that, along with the data-link IDs, uniquely identify a single end-to-end circuit. Each DLSw node must keep a table of these circuit ID pairs: one for the local end of the circuit and one for the remote end of the circuit.
- **Origin transport ID**—Identifies the individual TCP/IP port on a DLSw node. Values have only local significance. Each DLSw node must reflect the values, along with the associated values for the DLC port ID and the data-link correlator, when returning a message to a DLSw partner.
- **Target data-link correlator**—Works in tandem with the target DLC port ID to form a 64-bit circuit ID that identifies the DLC circuit within a single DLSw node. The circuit ID is unique in a single DLSw node and is assigned locally. The end-to-end circuit is identified by a pair of circuit IDs that, along with the data-link IDs, uniquely identifies a single end-to-end circuit. Each DLSw node must keep a table of these circuit ID pairs: one for the local end of the circuit and one for the remote end of the circuit.
- **Transport ID**—Identifies the individual TCP/IP port on a DLSw node. Values have only local significance. Each DLSw node must reflect the values, along with the associated values for the DLC port ID and the data-link correlator, when returning a message to a DLSw partner.

# Review Questions

**Q—***DLSw provides link layer acknowledgments. What is meant by link layer acknowledgments? Why is this advantageous?*

**A—**Link-layer acknowledgments (acks) refer to a process within the *broadcast domain* of the end device. The acks are between the end device and the local DLSw switch (router). Without link layer acks, the ack must reach all the way to the other end device. The ack may need to cross several LAN segments and a wide-area network. While crossing the WAN, significant propagation delay may be introduced, causing the protocol to time out and fail.

**Q—***DLSw SSP uses what transport protocol? What are the advantages and disadvantages of this selection?*

**A—**SSP uses TCP. This has the normal advantages of a reliable transport protocol, in which the data flow is monitored and retransmitted if any data is lost (sequence numbers and acks). However, TCP may not scale well when there are many DLSw switches that need to establish a peer-to-peer relationship.

**Q—***List and describe the three operational phases of DLSw.*

**A—**In phase one, DLSw peers establish two TCP connections. In phase two, the peers exchange capabilities with each other. This helps to ensure that the peers use the same options. It is particularly necessary to do this in an environment in which DLSw components come from multiple vendors. In

phase three, the circuit establishment phase, end devices establish a connection to their intended end-device target. This involves a local connection between the end device and the DLSw switch, and for the DLSw switches to discover what DLSw peer to send the data to.

**Q—***What protocols does DLSw support?*

**A—**SNA and NetBIOS. Both of these depend upon link layer acknowledgments.

**Q—***What is the normal Layer 2 process employed without DLSw?*

**A—**Before DLSw, systems used source-route bridging (SRB). However, SRB doesn't scale well in a WAN environment because of the hop-count limitation (7) and the inefficient handling of broadcast traffic.

**Q—***DLSw defines two message types. What are they, which has the larger header, and is there anything in common between them?*

**A—**The two messages are for control and information flow. The control frame has a 72-byte header, and the information message supports a 16-byte header. Therefore, the control frame has a larger header. The first 16 bytes of both headers have the same format.