Stream Control Transmission Protocol

[Reference: RFC 2960; ietf.org]

THE UNIVERSITY OF TEXAS AT DALLAS

Features of SCTP

- Acked error-free non duplicated user data transfer
- Data fragmentation conforming to discovered path MTU (max transmission unit)
- Sequenced delivery of user messages within each stream (but have multiple streams)
- Option for out-of-order delivery of a user message
- Bundling of multiple user messages into a single SCTP packet
- Fault-tolerance (at network level) by supporting multi-homing at either/both ends of an association and heart-beat/keep alive messages
- Message boundaries preserved
- SACK

Protocol Stack

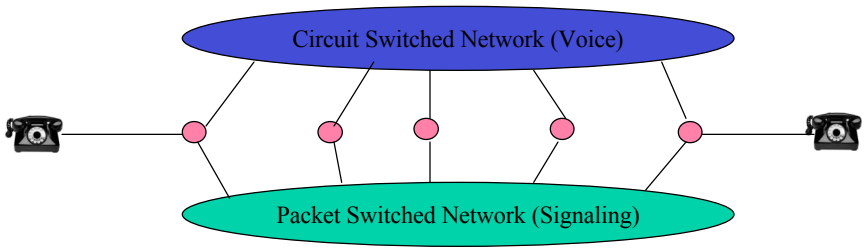
Application
UDP TCP SCTP
IP
DL/MAC/Physical

THE UNIVERSITY OF TEXAS AT DALLAS

Need for SCTP (why not use TCP?)

- Head of line blocking in TCP; not good in many situations
- Applications must add their own record-making; TCP is stream (or byte) oriented
- Fault-tolerance: What if NIC with destination IP address fails?
- TCP vulnerable to Denial of Service attacks
- Main motivator:
  - Carrying SS7 signaling on IP Networks

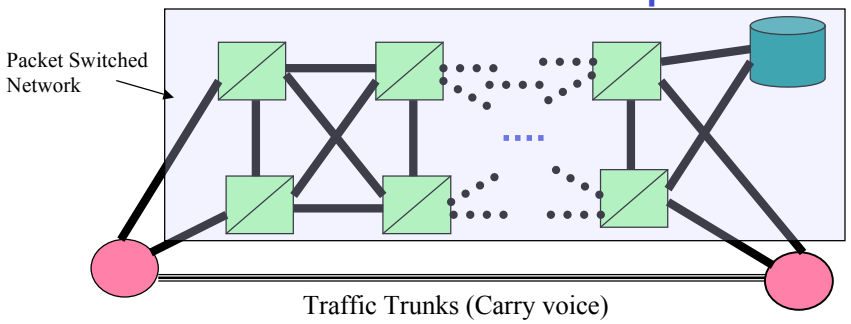
Sample PSTN Network



SSP

THE UNIVERSITY OF TEXAS AT DALLAS

A Sample SS7 Network



Legend — or •••• SS7 Link Set

SCP (Signaling Control Point)

STP (Signaling Transfer Point) (~ Router)

SSP (Signaling Switch Point)

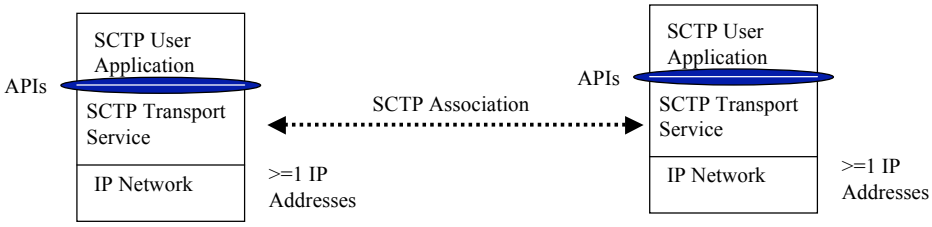
(Packet switched) SS7 Networks's Needs

- Reliability:
  - Multiple parallel links; nodes in mated pairs
- Packet sequencing (between two switches) needed only within each call; not across all calls between two switches
- In VoIP, need to do signaling in IP networks. (Both voice and signaling messages on IP networks)
- Sample Fig

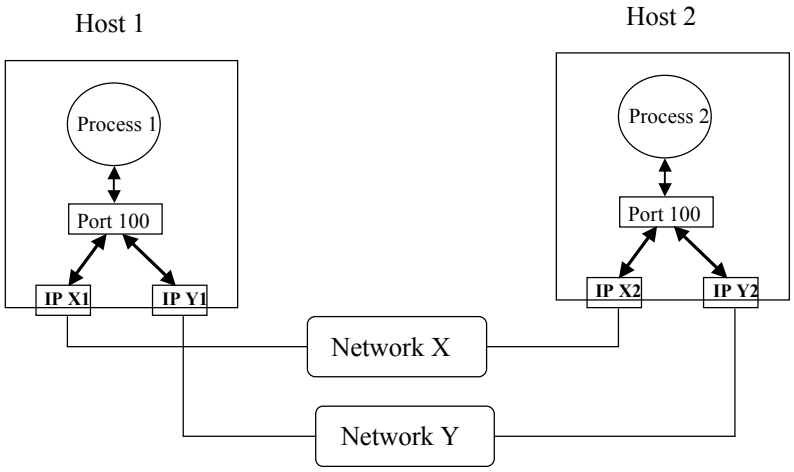
THE UNIVERSITY OF TEXAS AT DALLAS

What if a NIC fails?

- If using TCP, the connection is torn down and new connection (to alternate IP address) made. Lose all current calls being set up
- Using SCTP?



An Example

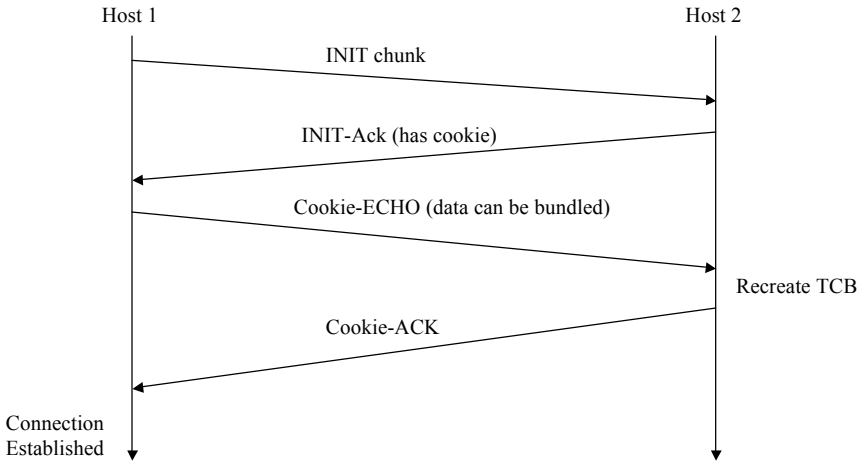


THE UNIVERSITY OF TEXAS AT DALLAS

SCTP Components

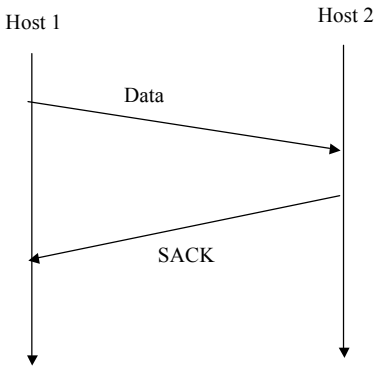
1. Association startup and teardown
2. Sequenced delivery within streams
3. User data fragmentation
4. Sack, congestion avoidance
5. Chunk bundling
6. Packet validation
7. Path management

Association Initiation

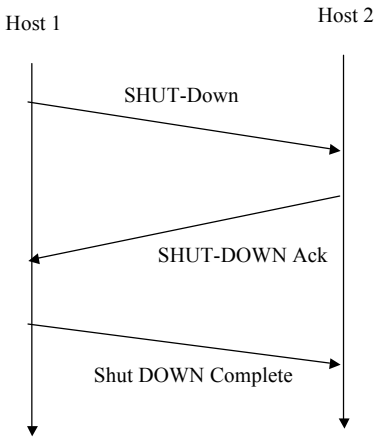


THE UNIVERSITY OF TEXAS AT DALLAS

Data Transmission

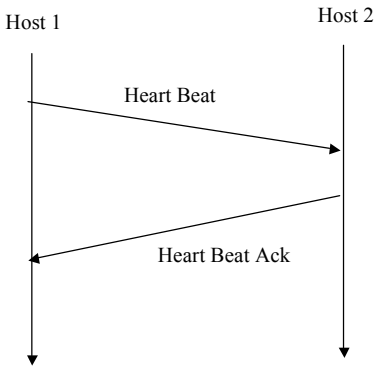


Shut Down

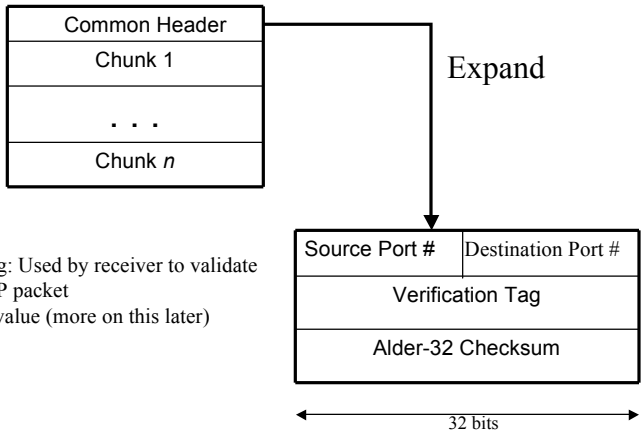


THE UNIVERSITY OF TEXAS AT DALLAS

Heart Beat



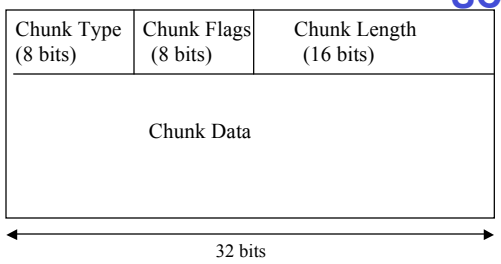
SCTP Packet Format



Verification tag: Used by receiver to validate sender of SCTP packet
= Initiate Tag value (more on this later)

THE UNIVERSITY OF TEXAS AT DALLAS

SCTP Chunk



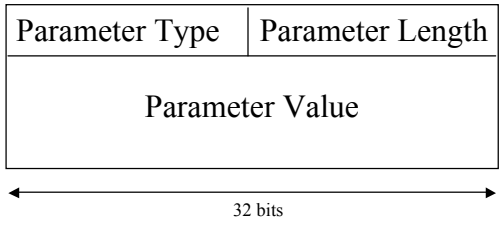
Type:
0 = Payload data
1=INIT
2=INIT-ACK
3=SACK
4=Heart Beat Request
5=Heart Beat ACK
6=Abort
7=Shut Down
8= Shut Down Ack …

Chunk Type is coded carefully.
High order 2 bits say what to do if type in chunk is undefined

00 = stop processing packet and discard packet
(don't process other chunks)
01 = same as above, report unrecognized type
10 = skip this chunk and continue processing
11 = same as 10, report unrecognized type

SCTP Chunk (Continued)

- Flags: 8 bits. Depends on type of chunk.
  - [When not needed, set all to 0]
- Length: >=4 (in bytes). Padded bytes not counted in length
- Depending on chunk type, there may be chunk-specific parameters [will see examples later]



Like chunk type, unrecognized parameter types are handled

THE UNIVERSITY OF TEXAS AT DALLAS

INIT Chunk

Type = 1	Flags = 0	Length (variable)
* Initiate tag *		
Advertised Receiver Window Credit		
# of outbound streams	*# of inbound streams*	
Initial Transmit Sequence Number (TSN)		
Optional/variable length parameters		

* Required Fields

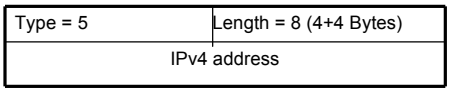
Initiate Tag: Unsigned 32 bit # [randomly chosen]
Receiver of INIT (responding end) records this.
Must be sent by receiver back to sender in every packet

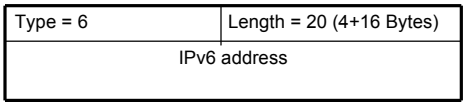
Value = 0 => receiver **aborts** association

INIT Chunk: Continued

- Advertised Window Credit: Buffer size in Bytes. (Cannot be decreased)
- # of outbound and inbound streams should both be greater than 0
- Initial TSN: Sequence numbers (like in TCP)

Optional Variable Length Parameters





THE UNIVERSITY OF TEXAS AT DALLAS

INIT Chunk: Continued

- Incoming packets may be addressed to any one of the multiple IP addresses (specified in optional parameters)
- This is multi homing
- No IP addresses? Use sender's IP address (from IP Packet)

INIT Chunk: Continued

Type = 11	Length
Host name (instead of IP address)	

THE UNIVERSITY OF TEXAS AT DALLAS

INIT-Ack

Type = 2	Flags	Chunk Length (variable)
* Initiate tag *		
Advertised Receiver Window Credit		
# of outbound streams	*# of inbound streams*	
Initial Transmit Sequence Number (TSN)		
COOKIE parameter; others		

* Required Fields

INIT-Ack must contain COOKIE parameter. COOKIE encrypts all state information needed to construct association, COOKIE creation time, and COOKIE life span

INIT-Ack Continued

- COOKIE expires to protect against replay attacks
- Sender of COOKIE destroys all state
  - It will reconstruct later on receiving COOKIE on COOKIE-Echo

COOKIE Echo (S->R)

Type = 10	Chunk Flags	Length >=4 (4+COOKIE L)
COOKIE		
Data Chunks (Optional); may be bundled with COOKIE Echo		

R unpacks COOKIE, reconstructs state, sends COOKIE-ACK to S and can begin accepting data

Data Chunk

Type=0	Flags=UBE	Length0	Length1
TSN	TSN	TSN	TSN
Stream	Number	Str	Seq#
Payloadid0	Payloadid1	Payloadid2	Payloadid3
User	Data	variable	length

Flag Bits UBE are used to indicate:
U –Unordered Data
B –Beginning of Fragmented Message
E –End of Fragmented Message
A user message that fits in one chunk would have both the B and E bits set.

THE UNIVERSITY OF TEXAS AT DALLAS

Packet Loss

- When one packet is lost, retransmission will occur in one of two ways:
  - Repeated SACKs occur reporting the missing packet (via holes) 4 times.
  - A time-out on the packet.
- Receiver SACKs every packet when a hole exists.

Gaps

- Cumulative TSN is the highest consecutive TSN received (no gaps).
- All gaps/fragments reports describe what has been received.
- All gap/fragments numbers are offsets from the cumulative TSN.
- Retransmissions are made to alternate destinations if possible.

THE UNIVERSITY OF TEXAS AT DALLAS

SACK

Type = 3	Flags = 0	Length
Cumulative TSN		
Advertised Receiver Window Credit		
n=# of Gap acks	# of Dup TSNs =m	
Gap Ack Block 1 start	Gap Ack Block 1 end	
. . .	. . .	
Gap ack block n start	Gap ack block n end	
Dup TSN 1		
. . .		
Dup TSN m		

Example

- Received 200, 201, 202, gap, 210, 211, 212, gap, 218, 219, 220
- Cum TSN = 202
- Number of gaps = 2
- Start = 8; End = 10; {received 202, 202+8 to 202+10}
- Start = 16; End = 18

THE UNIVERSITY OF TEXAS AT DALLAS

Sending Large Message

- Message size > path MTU

1. Break message into path-MTU sized chunks
   - Overhead: IP/Data Chunk are common headers
2. Assign these chunks TSNs in sequential order
3. BE bits:
   - 10 = first chunk
   - 00 = middle pieces (pieces other than first and last)
   - 01 = last chunk

- Unfragmented messages have the BE flags set to 11.
- The TSN is useful in guaranteeing correct delivery.

Receiving Large Message

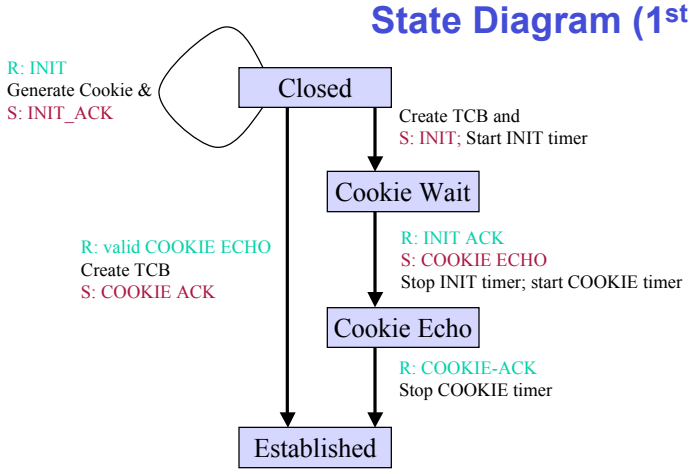
- Save each piece in a re-assembly queue.
- Use TSN to order the pieces.
- When all pieces are received (from 10 to 01) with no missing TSNs, merge all the data together.

THE UNIVERSITY OF TEXAS AT DALLAS

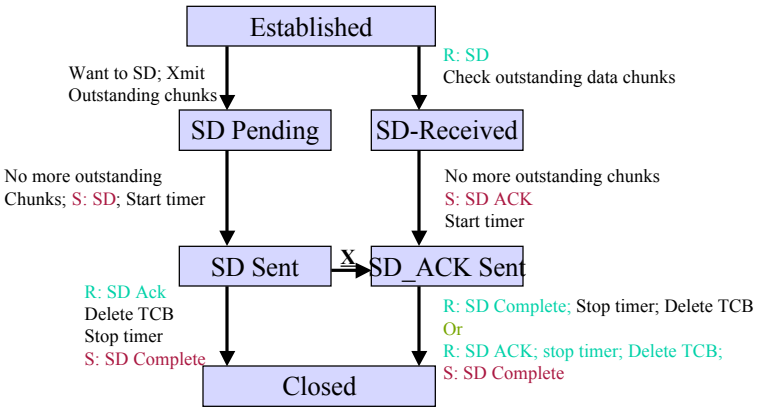
Heart Beat and Fault Management

- HB sent on idle destinations at pre-determined rate
- Receiver responds with Ack.
- Misses? Timeouts on HB?
  - > Threshold? Report destination is down
  - Report to upper layer
  - If down destination is primary, switch all traffic to an alternate IP address

State Diagram (1st Half)



State Diagram (2nd Half)



X :: R: SD;
S: SD Ack; start timer

Retransmission Timer

- RTO
- Compute and manage RTO of each end point
- Like TCP

THE UNIVERSITY OF TEXAS AT DALLAS

Security

- No protection to data contents
  - An attacker can sniff LAN, hijack connection, transmit bad packets, etc.
  - IP Sec can be used when protection against a sniffer is needed
- SCTP uses validation/verification tag to verify packets that belong to an association
  - Blind attacks are not possible
- COOKIE-Protect against SYN attacks

Congestion Control

- Slow start and congestion avoidance per association
- Uses:
  1. Receiver advertised window size
  2. Congestion control window- adjusted by sender based on observed network conditions
  3. Slow start threshold

Need these for each IP address (multi-homed)

Four NACKS (by SACK) => immediate transmission of missing packet (fast retransmit)