Transporting Voice by Using IP

Voice over UDP, not TCP

- Speech
  - Small packets, 10 – 40 ms in duration
  - Occasional packet loss is not a catastrophe
  - Delay-sensitive
    - TCP: connection set-up, ack, retransmit → delays
  - 5 % packet loss is acceptable if evenly spaced
    - Resource management and reservation techniques
    - A managed IP network
  - In-sequence delivery
    - Mostly yes
- UDP was not designed for voice traffic (not exactly a marriage made in heaven)

Real-Time Transport Protocol

- RTP: A Transport Protocol for Real-Time Applications
  - RFC 1889
  - RTP – Real-Time Transport Protocol
  - RTCP – RTP Control Protocol
- UDP
  - Packets may be lost or out-of-sequence
- RTP over UDP
  - A sequence number
  - A time stamp for synchronized play-out and for delay and jitter calculation
  - Does not solve the problems; simply provides additional information

RTCP

- A companion protocol
- Exchange messages between session users
- # of lost packets, delay and inter-arrival jitter
- The actual voice packets are carried within RTP packets.
- RTCP packets are used for the transfer of the quality feedback.
- RTCP is implicitly open when an RTP session is open
- E.g., RTP/RTCP uses UDP port 5004/5005

RTP Payload Formats [1/2]

- RTP carries the actual digitally encoded voice
  - RTP header + a payload of voice/video samples
  - UDP and IP headers are attached
- Many voice- and video-coding standards
  - RTP must include a mechanism for the receiving end to know which coding standard is being used
  - A payload type identifier in the RTP header
    - Specified in RFC 1890
    - New coding schemes have become available
      - GSM Enhanced Full-rate (EFR) coder
    - See Table 2-1 and Table 2-2
  - A sender has no idea what coding schemes a receiver could handle.

RTP Payload Formats [2/2]

- Separate signaling systems
  - Capability negotiation during the call setup
  - SIP (Session Initiation Protocol) and SDP (Session Description Protocol)
  - A dynamic payload type (payload type numbers 96 to 127) may be used.
    - Support new coding scheme in the future
    - The encoding name is also significant.
      - Unambiguously refer to a particular payload specification
      - Should be registered with the IANA
- RED, "Redundant" payload type
  - Voice samples + previous samples
  - May use different encoding schemes (more bandwidth-efficient)
  - Cope with packet loss

RTP Header Format

<table>
<tr><td>00</td><td>01</td><td>02</td><td>03</td><td>04</td><td>05</td><td>06</td><td>07</td><td>08</td><td>09</td><td>10</td><td>11</td><td>12</td><td>13</td><td>14</td><td>15</td><td>16</td><td>17</td><td>18</td><td>19</td><td>20</td><td>21</td><td>22</td><td>23</td><td>24</td><td>25</td><td>26</td><td>27</td><td>28</td><td>29</td><td>30</td><td>31</td></tr>
<tr><td colspan="2">V=2</td><td>P</td><td>X</td><td colspan="4">CC</td><td>M</td><td colspan="7">PT</td><td colspan="16">Sequence Number</td></tr>
<tr><td colspan="32">Timestamp</td></tr>
<tr><td colspan="32">Synchronization Source (SSRC) Identifier</td></tr>
<tr><td colspan="32">Countributing Source (CSRC) Identifier (0 to 15 entries)</td></tr>
</table>

<table>
<tr><td>00</td><td>01</td><td>02</td><td>03</td><td>04</td><td>05</td><td>06</td><td>07</td><td>08</td><td>09</td><td>10</td><td>11</td><td>12</td><td>13</td><td>14</td><td>15</td><td>16</td><td>17</td><td>18</td><td>19</td><td>20</td><td>21</td><td>22</td><td>23</td><td>24</td><td>25</td><td>26</td><td>27</td><td>28</td><td>29</td><td>30</td><td>31</td></tr>
<tr><td colspan="16">Profile-specific informaiton</td><td colspan="16">Length</td></tr>
<tr><td colspan="32">Header extension</td></tr>
</table>

The RTP Header [1/4]

- Version (V)
  - 2
- Padding (P)
  - The padding octets at the end of the payload
  - The payload needs to align with 32-bit boundary
  - The last octet of the payload contains a count of the padding octets.
- Extension (X)
  - 1, contains a header extension

The RTP Header [2/4]

- CSRC Count (CC)
  - The number of contributing source identifiers
- Marker (M)
  - Support silence suppression
  - The first packet of a talkspurt, after a silence period
- Payload Type (PT)
  - In general, a single RTP packet will contain media coded according to only one payload format.
  - RED is an exception.
- Sequence number
  - A random number generated by the sender at the beginning of a session
  - Incremented by one for each RTP packet

The RTP Header [3/4]

- Timestamp
  - 32-bit
  - The instant at which the first sample in the payload was generated
  - The receiver
    - Synchronized play-out
    - Calculate the jitter
    - The initial timestamp is a random number chosen by the sending application.

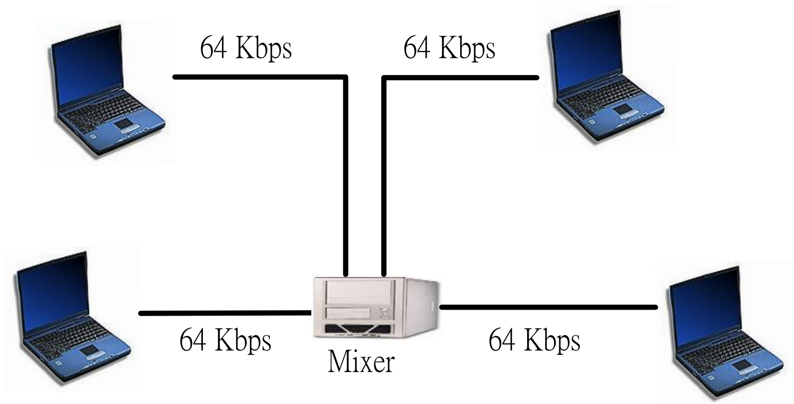
The RTP Header [4/4]

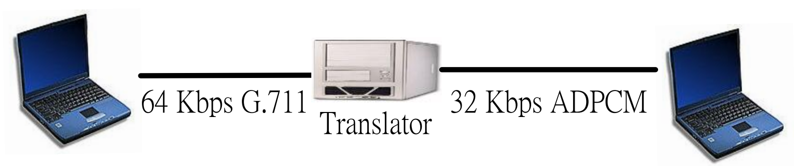
- Synchronization Source (SSRC)
  - 32-bit identifier
  - The entity setting the sequence number and timestamp
    - Normally the sender of the RTP packet
  - Chosen randomly, independent of the network address
    - Meant to be globally unique within a session
  - May be a sender or a mixer
- Contributing Source (CSRC)
  - An SSRC value for a contributor
  - Used to identify the original sources of media behind the mixer
  - 0-15 CSRC entries
- RTP Header Extensions (e.g., additional information for payload format)

<table>
<tr><td>00</td><td>01</td><td>02</td><td>03</td><td>04</td><td>05</td><td>06</td><td>07</td><td>08</td><td>09</td><td>10</td><td>11</td><td>12</td><td>13</td><td>14</td><td>15</td><td>16</td><td>17</td><td>18</td><td>19</td><td>20</td><td>21</td><td>22</td><td>23</td><td>24</td><td>25</td><td>26</td><td>27</td><td>28</td><td>29</td><td>30</td><td>31</td></tr>
<tr><td colspan="16">Profile-specific informaiton</td><td colspan="16">Length</td></tr>
<tr><td colspan="32">Header extension</td></tr>
</table>

Mixers and Translators

- Mixers
  - Enable multiple media streams from different sources to be combined into a single stream
    - If the capacity or bandwidth of a participant is limited
  - An audio conference
  - The SSRC is the mixer
    - More than one CSRC values
- Translators
  - Manage communications between entities that does not support the same coding scheme
  - The SSRC is the participant, not the translator.





The RTP Control Protocol [1/3]

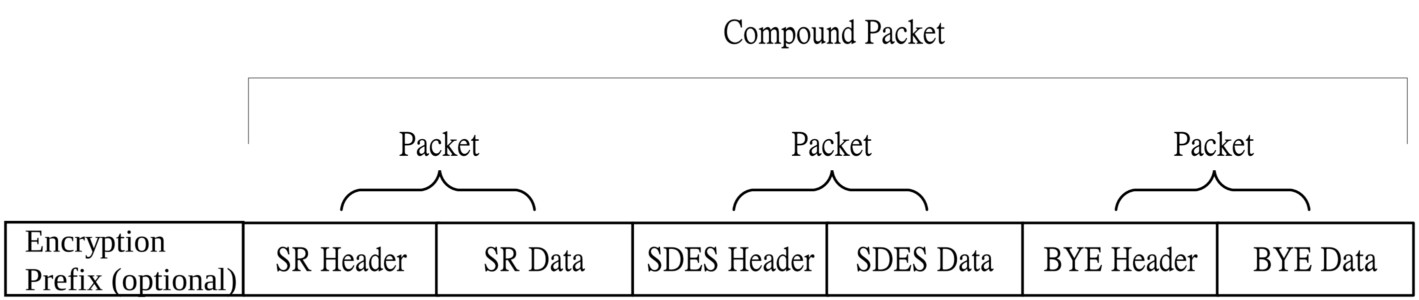
- RTCP
  - A companion control protocol of RTP
  - Periodic exchange of control information
    - For quality-related feedback
  - A third party can also monitor session quality and detect network problems.
    - Using RTCP and IP multicast
- Five types of RTCP packets
  - Sender Report: used by active session participants to relay transmission and reception statistics
  - Receiver Report: used to send reception statistics from those participants that receive but do not send

The RTP Control Protocol [2/3]

- Source Description (SDES)
  - One or more descriptions related to a particular session participant
    - To identify session participants
  - Must contain a canonical name (CNAME)
    - Separate from SSRC which might change
    - When both audio and video streams were being transmitted, the two streams would have
      - different SSRCs
      - the same CNAME for synchronized play-out
- BYE
  - The end of a participation in a session
- APP
  - For application-specific functions

The RTP Control Protocol [3/3]

- Two or more RTCP packets will be combined
  - SRs and RRs should be sent as often as possible to allow better statistical resolution.
  - New receivers in a session must receive CNAME very quickly to allow a correlation between media sources and the received media.
  - Every RTCP packet must contain a report packet (SR/RR) and an SDES packet
    - Even if no data to report
- An example of RTP compound packet



RTCP Sender Report

- SR
  - Header Info
  - Sender Info
  - Receiver Report Blocks
  - Option
    - Profile-specific extension

<table>
<tr><td>00</td><td>01</td><td>02</td><td>03</td><td>04</td><td>05</td><td>06</td><td>07</td><td>08</td><td>09</td><td>10</td><td>11</td><td>12</td><td>13</td><td>14</td><td>15</td><td>16</td><td>17</td><td>18</td><td>19</td><td>20</td><td>21</td><td>22</td><td>23</td><td>24</td><td>25</td><td>26</td><td>27</td><td>28</td><td>29</td><td>30</td><td>31</td></tr>
<tr><td colspan="2">V=2</td><td>P</td><td>X</td><td colspan="4">RC</td><td colspan="8">PT=SR=200</td><td colspan="16">Length</td></tr>
<tr><td colspan="32">SSRC of sender</td></tr>
<tr><td colspan="32">NTP Timestamp (most significant word)</td></tr>
<tr><td colspan="32">NTP Timestamp (least significant word)</td></tr>
<tr><td colspan="32">RTP Timestamp</td></tr>
<tr><td colspan="32">sender's packet count</td></tr>
<tr><td colspan="32">sender's octet count</td></tr>
<tr><td colspan="32">SSRC_1(SSRC of first source)</td></tr>
<tr><td colspan="8">fraction lost</td><td colspan="24">fraction lost</td></tr>
<tr><td colspan="32">extended highest sequence number received</td></tr>
<tr><td colspan="32">interarrival jitter</td></tr>
<tr><td colspan="32">last SR (LSR)</td></tr>
<tr><td colspan="32">Delay since last SR (DLSR)</td></tr>
<tr><td colspan="32">SSRC_2(SSRC of second source)</td></tr>
<tr><td colspan="32">:
:</td></tr>
<tr><td colspan="32">profile-specific extensions</td></tr>
</table>

Header Info

- Resemble to an RTP packet
  - Version
    - 2
  - Padding bit
    - Padding octets?
  - RC, report count
    - The number of reception report blocks
    - 5-bit
      - If more than 31 reports, an RR is added
  - PT, payload type (200)

Sender Info

- SSRC of sender
- NTP Timestamp
  - Network Time Protocol Timestamp
    - The time elapsed in seconds since 00:00, 1/1/1900 (GMT)
    - 64-bit
      - 32 MSB: the number of seconds
      - 32 LSB: the fraction of a seconds (enabling a precision of about 200 picoseconds)
- RTP Timestamp
  - The same as used for RTP timestamps in RTP packets
  - For better synchronization with the sender of the report
- Sender's packet count
  - Cumulative within a session
- Sender's octet count
  - Cumulative within a session

RR blocks [1/2]

- SSRC_n
  - The source identifier of the session participant to which the data in this RR block pertains.
- Fraction lost
  - Fraction of packets lost since the last report issued by this participant
  - By examining the sequence numbers in the RTP header
- Cumulative number of packets lost
  - Since the beginning of the RTP session
- Extended highest sequence number received
  - The sequence number of the last RTP packet received
  - 16 lsb, the last sequence number
  - 16 msb, the number of sequence number cycles

RR blocks [2/2]

- Inter-arrival jitter
  - An estimate of the variance in RTP packet arrival
- Last SR Timestamp (LSR)
  - The last SR received from the source
  - Used to check if the last SR has been received
- Delay Since Last SR (DLSR)
  - The duration in units of 1/65,536 seconds
  - Between the reception of the last sender report from the source and the issuance of this receiver report block

RTCP Receiver Report

- RR
  - Issued by a participant who receives RTP packets but does not send, or has not yet sent
  - Is almost identical to an SR
    - PT = 201
    - No sender information

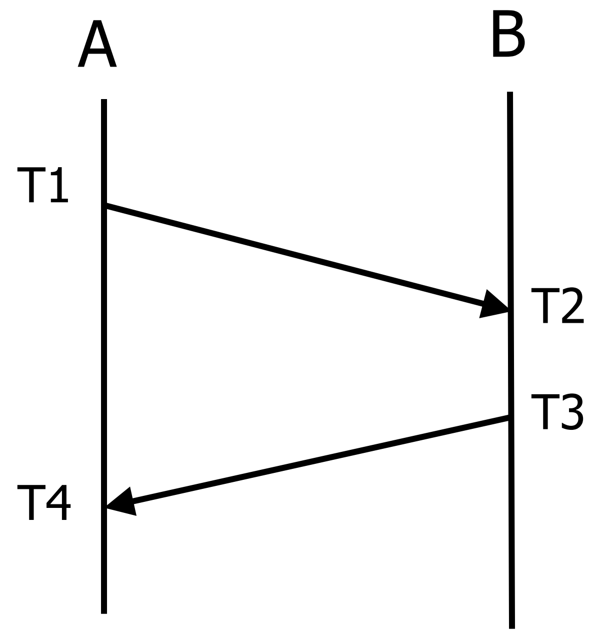
RTCP Source Description Packet

- Provides identification and information regarding session participants
  - Must exist in every RTCP compound packet
- Header
  - V, P, SC, PT=202, Length
- Zero or more chunks of information
  - An SSRC or CSRC value
  - One or more identifiers and pieces of information
    - A unique CNAME (user@host) does not change within a given session.
    - Email address, phone number, name

- RTCP BYE Packet (PT=203)
  - Indicate one or more media sources (SSRC or CSRC) are no longer active
- Application-Defined RTCP Packet (PT=204)
  - For application-specific data
  - For non-standardized application

Calculating Round-Trip Time

- Use SRs and RRs
- E.g.
  - Report A: A, T1 $\rightarrow$ B, T2
  - Report B: B, T3 $\rightarrow$ A, T4
  - RTT = T4-T3+T2-T1
  - RTT = T4-(T3-T2)-T1
  - Report B
    - LSR = T1
    - DLSR = T3-T2



Calculating Jitter

- The variation in delay
- The mean deviation of the difference in packet spacing at the receiver compared to the packet spacing at the sender for a pair of packets
  - This value is equivalent to the derivation in transit time for a pair of packets.
  - S_i = the RTP timestamp for packet i
  - R_i = the time of arrival
  - $D(i,j) = (R_j-R_i)-(S_j-S_i) = (R_j-S_j) - (R_i-S_i)$
- The Jitter is calculated continuously
  - $J(i) = J(i-1) + (| D(i-1,i) | - J(i-1))/16$