(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
9 August 2012 (09.08.2012)

(10) International Publication Number
WO 2012/106032 A1

(51) International Patent Classification: *H04L 29/06* (2006.01) *H04L 12/56* (2006.01)

(21) International Application Number: PCT/US2011/064592

(22) International Filing Date: 13 December 2011 (13.12.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data: 61/439,245 3 February 2011 (03.02.2011) US

(71) Applicant *(for all designated States except US)*: **INTERDIGITAL PATENT HOLDINGS, INC.** [US/US]; 3411 Silverside Road, Concord Plaza, Suite 105, Hagley Building, Wilmington, Delaware 19810 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **REZNIK, Alexander** [US/US]; 1212 River Road, Titusville, New Jersey 08560 (US). **LIVET, Catherine M.** [FR/CA]; 5023 rue de Brebeuf, Montreal, QC H2J 3L6 (CA). **DIFAZIO, Robert A.** [US/US]; 6 Gwendale Lane, Greenlawn, New York 11740 (US). **PERRAS, Michelle** [CA/CA]; 4434 Christophe-Colomb, Montreal, Québec H2J 3G5 (CA).

(74) Agent: **WOLINSKY, Scott**; Volpe and Koenig, P.C., 30 S. 17th Street, United Plaza, Philadelphia, Pennsylvania 19103 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

(54) **Title**: METHOD AND APPARATUS FOR ESTABLISHING AN END-TO-END MULTIPATH CONNECTION USING A MULTI-CONNECTION TRANSPORT PROTOCOL

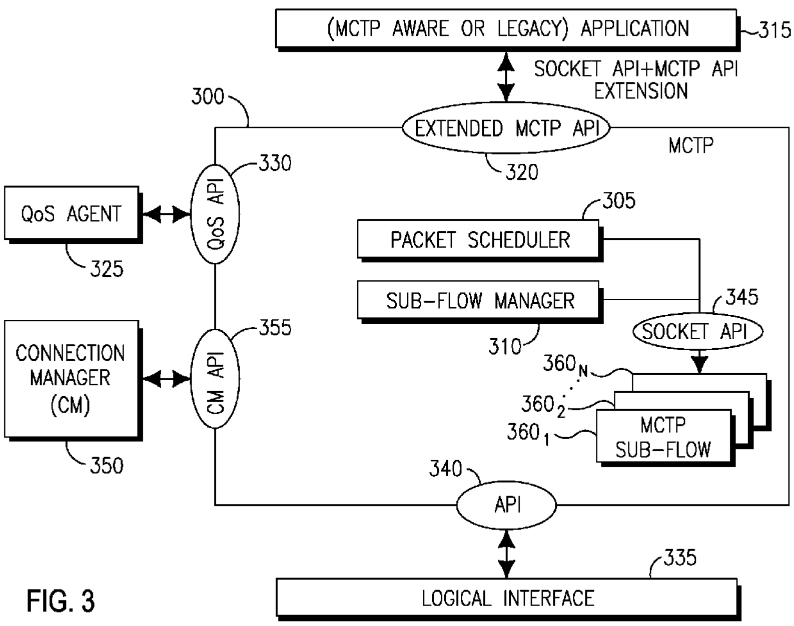

FIG. 3

(57) **Abstract**: A method is described for establishing an end-to-end multipath connection between two peers using of a multi-connection transport protocol (MCTP) architecture. The MCTP architecture may include a packet scheduler, a queue configured to store a plurality of packets received from the packet scheduler in a particular sequence, a sub-flow manager configured to distribute the packets among a plurality of sub-flows, wherein at least two of the sub-flows are synchronized, and a synchronized congestion controller (SCC) configured to control congestion of the synchronized sub-flows. A first transmission control protocol (TCP) positive acknowledgement (ACK) may be generated when a first TCP packet is received over one of the synchronized sub-flows. Duplicate TCP packets received after the first TCP packet is received may be discarded. ACKs received after the first ACK may be discarded if they correspond to a TCP packet that was already acknowledged.

WO 2012/106032 A1

METHOD AND APPARATUS FOR ESTABLISHING AN END-TO-END MULTIPATH CONNECTION USING A MULTI-CONNECTION TRANSPORT PROTOCOL

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 61/439,245 filed February 3, 2011, the contents of which are hereby incorporated by reference herein.

FIELD OF INVENTION

[0002] This application is related to communication protocols.

BACKGROUND

[0003] In computer networking, a transport layer may provide end-to-end communication services for applications within a layered architecture of network components and protocols. The transport layer may also provide convenient services, such as connection-oriented data stream support, flow control and multiplexing. Well-known transport protocols include the transmission control protocol (TCP), the user datagram protocol (UDP) and the stream control transmission protocol (SCTP).

[0004] Multi-interface devices may be capable of simultaneously using multiple transport paths for a single application connection. A multipath TCP (MPTCP) protocol may allow this by connecting multiple TCP sessions simultaneously. Each TCP session may then be referred to as an MPTCP sub-flow. It has been proposed that each sub-flow of an MPTCP session may behave as a fully valid independent TCP flow in order to retain the congestion control associated with a TCP, (widely viewed as a key contributor to the relative stability of Internet protocol (IP)-based communication over the Internet). However, an MPTCP may be configured to send different data on different sub-flows, although duplicate transmissions are desired over such sub-flows, (e.g., due to quality of service (QoS) needs). Thus, sending a single positive acknowledgement (ACK) may not be sufficient for all sub-links because, with a fully preserved TCP operation, each TCP link may expect its own ACK.

SUMMARY

[0005] A method is described for establishing an end-to-end multipath connection between two peers using of a multi-connection transport protocol (MCTP) architecture. An MCTP connection is created of multiple transport sub-flows based on the TCP, UDP or any other transport layer protocol. The MCTP architecture may include a packet scheduler, a queue configured to store a plurality of packets received from the packet scheduler in a particular sequence, a sub-flow manager configured to distribute the packets among a plurality of sub-flows, wherein at least two of the sub-flows are synchronized, and a synchronized congestion controller (SCC) configured to control congestion of the synchronized sub-flows. A first transmission control protocol (TCP) positive acknowledgement (ACK) may be generated when a first TCP packet is received over one of the synchronized sub-flows. Duplicate TCP packets received after the first TCP packet is received may be discarded. ACKs received after the first ACK may be discarded if they correspond to a TCP packet that was already acknowledged.

BRIEF DESCRIPTION OF THE DRAWINGS

[0006] A more detailed understanding may be had from the following description, given by way of example in conjunction with the accompanying drawings wherein:

[0007] Figure 1A shows an example communications system in which one or more described embodiments may be implemented;

[0008] Figure 1B shows an example wireless transmit/receive unit (WTRU) that may be used within the communications system shown in Figure 1A;

[0009] Figure 2 shows an example of a multipoint transmission control protocol (MPTCP) stack interfacing with an application via a socket interface;

[0010] Figure 3 shows an example of internal architecture of a multi-connection transport protocol (MCTP) and its interaction with other functions in a device through various application programming interfaces (APIs);

[0011] Figure 4 shows a MCTP architecture with different sub-flow types; and

[0012] Figures 5A and 5B illustrate MCTP packet redirection.

DETAILED DESCRIPTION

[0013] Figure 1A shows an example communications system 100 in which one or more described embodiments may be implemented. The communications system 100 may be a multiple access system that provides content, such as voice, data, video, messaging, broadcast, and the like, to multiple wireless users. The communications system 100 may enable multiple wireless users to access such content through the sharing of system resources, including wireless bandwidth. For example, the communications systems 100 may employ one or more channel access methods, such as code division multiple access (CDMA), time division multiple access (TDMA), frequency division multiple access (FDMA), orthogonal FDMA (OFDMA), single-carrier FDMA (SC-FDMA), and the like.

[0014] As shown in Figure 1A, the communications system 100 may include wired or wireless subscriber communication units 102a, 102b, 102c, 102d, a radio access network (RAN) 104, a core network (CN) 106, a public switched telephone network (PSTN) 108, the Internet 110, and other networks 112, though it will be appreciated that the described embodiments contemplate any number of subscriber communication units, base stations, networks, and/or network elements. Each of the subscriber communication units 102a, 102b, 102c, 102d may be any type of device configured to operate and/or communicate in a wired or wireless environment. By way of example, the subscriber communication units 102a, 102b, 102c, 102d may be configured to transmit and/or receive wired or wireless signals and may include user equipment (UE), a mobile station, a fixed or mobile subscriber unit, a pager, a cellular telephone, a personal digital assistant (PDA), a smartphone, a laptop, a notebook, a personal computer, a tablet computer, a wireless sensor, consumer electronics, and the like.

[0015] The communications system 100 may also include a base station 114a and a base station 114b. Each of the base stations 114a, 114b may be any type of device configured to interface with at least one of the subscriber communication units 102a, 102b, 102c, 102d to facilitate access to one or more communication networks, such as the CN 106, the Internet 110, and/or the other networks 112. By way of example, the base stations 114a, 114b may be a base transceiver station (BTS), a Node-B, an evolved Node-B (eNB), a Home Node-B (HNB), a Home eNB (HeNB), a site controller, an access point (AP), a wireless router, and the like. While the base stations 114a, 114b are each depicted as a single element, it will be appreciated that the base stations 114a, 114b may include any number of interconnected base stations and/or network elements.

[0016] The base station 114a may be part of the RAN 104, which may also include other base stations and/or network elements (not shown), such as a base station controller (BSC), a radio network controller (RNC), relay nodes, and the like. The base station 114a and/or the base station 114b may be configured to transmit and/or receive wireless signals within a particular geographic region, which may be referred to as a cell (not shown). The cell may further be divided into cell sectors. For example, the cell associated with the base station 114a may be divided into three sectors. Thus, in one embodiment, the base station 114a may include three transceivers, i.e., one for each sector of the cell. In another embodiment, the base station 114a may employ multiple-input multiple-output (MIMO) technology and, therefore, may utilize multiple transceivers for each sector of the cell.

[0017] The base station 114b in Figure 1A may be a wireless router, HNB, HeNB, or AP, for example, and may utilize any suitable RAT for facilitating wireless connectivity in a localized area, such as a place of business, a home, a vehicle, a campus, and the like. In one embodiment, the base station 114b and the subscriber communication units 102c, 102d may implement a radio technology such as IEEE 802.11 to establish a wireless local area network (WLAN). In another embodiment, the base station 114b and the subscriber communication units 102c, 102d may implement a radio technology such as IEEE 802.15 to establish a wireless personal area network (WPAN). In yet another embodiment, the base station 114b and the subscriber communication units 102c, 102d may utilize a cellular-based RAT, (e.g., WCDMA, CDMA2000, GSM, LTE, LTE-A, and the like), to establish a picocell or femtocell. As shown in Figure 1A, the base station 114b may have a direct connection to the Internet 110. Thus, the base station 114b may not be required to access the Internet 110 via the CN 106.

[0018] The RAN 104 may be in communication with the CN 106, which may be any type of network configured to provide voice, data, applications, and/or voice over Internet protocol (VoIP) services to one or more of the subscriber communication units 102a, 102b, 102c, 102d. For example, the CN 106 may provide call control, billing services, mobile location-based services, pre-paid calling, Internet connectivity, video distribution, and the like, and/or perform high-level security functions, such as user authentication. Although not shown in Figure 1A, it will be appreciated that the RAN 104 and/or the CN 106 may be in direct or indirect communication with other RANs that employ the same RAT as the RAN 104 or a different RAT. For example, in addition to being connected to the RAN 104, which may be utilizing an E-UTRA radio technology, the CN 106 may also be in communication with another RAN (not shown) employing a GSM radio technology.

[0019] The CN 106 may also serve as a gateway for the subscriber communication units 102a, 102b, 102c, 102d to access the PSTN 108, the Internet 110, and/or other networks 112. The PSTN 108 may include circuit-switched telephone networks that provide plain old telephone service (POTS). The Internet 110 may include a global system of interconnected computer networks and devices that use common communication protocols, such as the transmission control protocol (TCP), user datagram protocol (UDP) and the

Internet protocol (IP) in the TCP/IP suite. The networks 112 may include wired or wireless communications networks owned and/or operated by other service providers. For example, the networks 112 may include another CN connected to one or more RANs, which may employ the same RAT as the RAN 104 or a different RAT.

[0020] Some or all of the subscriber communication units 102a, 102b, 102c, 102d in the communications system 100 may include multi-mode capabilities, i.e., the subscriber communication units 102a, 102b, 102c, 102d may include multiple transceivers for communicating with different wired or wireless networks over different wired or wireless links. For example, the subscriber communication unit 102c shown in Figure 1A may be configured to communicate with the base station 114a, which may employ a cellular-based radio technology, and with the base station 114b, which may employ an IEEE 802 radio technology. Alternatively, one or more of the subscriber communication units 102 may be a laptop having a wired Ethernet connection plus a broadband wireless connection to provide multiple sub-flows,

[0021] Figure 1B shows an example subscriber communication unit 102 that may be used within the communications system 100 shown in Figure 1A. As shown in Figure 1B, the subscriber communication unit 102 may include a processor 118, a transceiver 120, a wired or wireless interface 122, a speaker/microphone 124, a keypad 126, a display/touchpad 128, a non-removable memory 130, a removable memory 132, a power source 134, a global positioning system (GPS) chipset 136, and peripherals 138. It will be appreciated that the subscriber communication unit 102 may include any sub-combination of the foregoing elements while remaining consistent with an embodiment.

[0022] The processor 118 may be a general purpose processor, a special purpose processor, a conventional processor, a digital signal processor (DSP), a microprocessor, one or more microprocessors in association with a DSP core, a controller, a microcontroller, an application specific integrated circuit (ASIC), a field programmable gate array (FPGA) circuit, an integrated circuit (IC), a state machine, and the like. The processor 118 may perform signal coding, data processing, power control, input/output processing, and/or any other functionality that enables the subscriber communication unit 102 to operate in a wired or wireless environment. The processor 118 may be coupled to the transceiver 120, which may be coupled to the interface 122. While Figure 1B depicts the processor 118 and the transceiver 120 as separate components, the processor 118 and the transceiver 120 may be integrated together in an electronic package or chip.

[0023] The interface 122 may be configured to transmit signals to, or receive signals from, a base station (e.g., the base station 114a) over an air interface 116. For example, in one embodiment, the interface 122 may be an antenna configured to transmit and/or receive RF signals. In another embodiment, the interface 122 may be an emitter/detector configured to transmit and/or receive IR, UV, or visible light signals, for example. In yet another embodiment, the interface 122 may be configured to transmit and receive both RF and light signals. The interface 122 may be configured to transmit and/or receive any combination of wired or wireless signals.

[0024] The processor 118 of the subscriber communication unit 102 may be coupled to, and may receive user input data from, the speaker/microphone 124, the keypad 126, and/or the display/touchpad 128 (e.g., a liquid crystal display (LCD) display unit or organic light-emitting diode (OLED) display unit). The processor 118 may also output user data to the speaker/microphone 124, the keypad 126, and/or the display/touchpad 128. In addition, the processor 118 may access information from, and store data in, any type of suitable memory, such as the non-removable memory 130 and/or the removable memory 132. The non-removable memory 130 may include random-access memory (RAM), read-only memory (ROM), a hard disk, or any other type of memory storage device. The removable memory 132 may include a subscriber identity module (SIM) card, a memory stick, a secure digital (SD) memory card, and the like. In other embodiments, the processor 118 may access information from, and store data in, memory that is not physically located on the subscriber communication unit 102, such as on a server or a home computer (not shown).

[0025] The processor 118 may further be coupled to other peripherals 138, which may include one or more software and/or hardware modules that provide additional features, functionality and/or wired or wireless connectivity. For example, the peripherals 138 may include an accelerometer, an e-compass, a satellite transceiver, a digital camera (for photographs or video), a universal serial bus (USB) port, a vibration device, a television transceiver, a hands free headset, a Bluetooth® module, a frequency modulated (FM) radio unit, a digital music player, a media player, a video game player module, an Internet browser, and the like.

[0026] Figure 2 shows an example of a multipoint transmission control protocol (MPTCP) stack 205 interfacing with an application 210 via a socket interface 215. As defined in MPTCP architecture, two levels of sequence numbering (SN) including an MPTCP packet SN and a TCP SN are used, allowing the MPTCP to perform "connection-level segmentation and reassembly" in addition to the segmentation and reassembly provided by each TCP sub-flow. A mapping may be provided to a receiver so that it may correctly re-assemble the connection-level segments. Additionally, MPTCP provides two layers of acknowledgement, (overall MPTCP as well as for each TCP flow). Finally, in addition to the congestion control utilized by each TCP sub-flow, a global congestion control algorithm may be used for the overall MPTCP connection. Thus, the MPTCP may be seen as a new protocol stack layer (205) which resides on top of multiple TCP instances (220) bound to an IP layer interface (225), which may be a "logical" construct whenever layer 4 (TCP or UPD) and IP are jointly implemented.

[0027] A multi-connection transport protocol (MCTP) is described herein that allows for an end-to-end multipath connection between two peers. Unlike MPTCP sub-flows, MCTP sub-flows may be capable of being either TCP or UDP connections and may be optimized to better interact with the MCTP. An

MCTP scheduler may be provided for optimizing the use of connection sub-flows.

[0028] MCTP functions may include MCTP connection management, (i.e., MCTP peer-to-peer negotiation, connection initiation by opening a first sub-flow, sub-flows addition and deletion to the MCTP session and MCTP connection termination, address management), and MCTP data management, (i.e., sending data received from the application to the MCTP peer and reception of data received through the sub-flows and transmission up to the application).

[0029] The purpose of an MCTP is to improve usage of multiple available sub-flows. Therefore, an optimized MCTP data scheduler may be required to optimize the use of the bandwidth of the various sub-flows and provide the best strategy for data transmission and reception. The MCTP may optimize packets scheduling received from an application over the different sub-flows. A scheduling policy may take into account the bandwidth availability, the round trip time (RTT), the favorite link, and the like. Further, when MCTP operates over a single source and destination interface, it may be identical to TCP, (i.e., provides backwards compatibility to TCP).

[0030] Figure 3 shows an example of internal architecture of an MCTP 300 and its interaction with other functions in a device through various application programming interfaces (APIs). The MTCP 300 may include a packet scheduler 305, (i.e., a packet scheduling function), and a sub-flow manager 310, (i.e., a sub-flow management function). The MTCP 300 may be incorporated into a wired or wireless subscriber communication unit, a base station, a computer or any other type of communication device, network entity or node. The MCTP 300 may interact with an application 315 that is either MCTP-aware or non-MCTP-aware via an extended MCTP API 320. An MCTP-aware application may use socket API plus an MCTP API extension, while a non-MCTP-aware may use a Legacy socket API.

[0031] The MCTP 300 may also interact with a QoS agent via a QoS API 330, through which application QoS requirements may be sent directly by the application 315 to the MCTP 300 at initial connection. During the connection, the QoS agent 325 may provide dynamic and application-specific QoS information to the MCTP 300 so that the MCTP 300 may optimize its scheduling and/or sub-flow management. The MCTP 300 may interact with IP-layer functionality via an API 340. The MCTP 300 may also interact with a connection manager (CM) 350 via a CM API 355. The CM 350 may provide some policy configuration parameters, such as preferred IP address, policies provided by the operator, quality of the link, and the like.

[0032] The packet scheduler 305 and the selection of MCTP sub-flows 360_1, 360_2, ..., 360_N are described herein. The MCTP-aware application 315 may pass data to the MCTP 300 and provide additional information about the expected QoS for the data sent if it has a higher priority, such as control data versus user data, if a data splitting partition is implemented, (i.e., the application splits the data based on the relative importance or other criteria related to the corresponding traffic stream). This additional information may be provided by, for example, adding a new API to the socket API. The packet scheduler 305 may also receive QoS information input from the CM 350, to learn the throughput, latency, and reliability per connection available. This QoS information may allow the packet scheduler 305 to compare the different connections available and to pick the most appropriate based on the expected QoS.

[0033] For example, in case one connection has the highest throughput, and if the application's requested QoS is focused on throughput, it may be uniquely chosen as a single sub-flow MCTP connection, with no need of an additional sub-flow. On the other hand, if minimizing the latency is the overriding goal, then sending over all connections may be the best alternative, (and waiting for the first ACK). If one connection is clearly the best, the delta may be minimal and that connection may be chosen.

[0034] Figure 4 shows an MCTP architecture with different sub-flow types. An MCTP overlay layer 405 may maintain its own data queue and an MTCP scheduler 415 may distribute data 420 to sub-flows according to established rules. Three sub-flows are shown. On the left is a single independent sub-flow 425, with a fully functional TCP 430, with its own congestion control algorithm. On the right, two synchronized sub-flows 430 are shown. Each of these employs its own L4 underlay protocol 435 fed by a common MCTP synchronized congestion controller (MCTP-SCC) 440. The MCTP-SCC 440 and each L4 underlay protocol 435 may transmit the same data over both sub-flows at the same time, (although the same may not be true for reception).

[0035] In one configuration, the MCTP-SCC 440 may be configured to operate as closely to a TCP as possible. In this case, the L4 underlay protocol 435 may be a pass-through layer, (i.e., it does nothing). The MCTP-SCC 440 may operate using TCP congestion control, acknowledgement algorithms, and the like, and may utilize a TCP header, (i.e., no new headers are required), TCP sequencing, and the like. However, as TCP is currently not defined to operate over multiple streams, certain changes need to be made.

[0036] The same TCP packet may be sent over two sub-flows 430. In particular, the TCP header may be the same in both cases. Because of this, sequence numbers are naturally synchronized over both sub-flows 430. An ACK may be generated whenever a TCP packet is received over any one of the sub-flows 430. Any duplicate packets that are received after the reception of the first TCP packet may be discarded. Similarly, an ACK may be included in the TCP response sent on all sub-flows. After the first ACK is received, the rest are discarded, (because TCP ACKs are part of a data packet, this occurs naturally when the duplicate packet itself is discarded). TCP flow control may be utilized.

[0037] Features of this configuration may include each sub-flow operating as a completely valid independent TCP flow, except that some packets may be acknowledged without an ACK coming through on that flow. An ACK may be sent on each sub-flow. However, it may be discarded if the corresponding packet is already acknowledged.

[0038] Further, a flow-control window size may be set based on an ACK which is used to acknowledge the data. Flow-control window sizes in duplicate (discarded) ACKs may be ignored. A synchronization connection establishment mechanism and the TCP header per packet may not need to change and new headers may not be required.

[0039] However, MCTP connection setup messaging may need to include additional information indicating that a particular TCP sub-flow may be setup as a set of synchronized sub-flows. This may provide additional signaling to the MCTP.

[0040] The MCTP setup signaling may indicate that certain sub-flows may be setup as synchronized UDP connections.

[0041] Congestion control for MCTP-SCC is also described herein. A potential problem with MCTP-SCC is the potential to continue sending packets, (new data or retransmissions), over interfaces and into networks that are congested, thus negatively impacting the congestion. A TCP may handle this problem through its congestion control algorithm. However, with MCTP-SCC signaling, this may be dominated by the least congested link or links. Because these links may tend to increase the rate of outgoing traffic over all synchronized interfaces, congestion on congested interfaces may increase.

[0042] In the case where an MCTP-SCC is defined using TCP, there may be a constraint in what may be performed while still making each sub-flow appear to be TCP-like, including in its congestion control behavior. Instead of simply discarding "late ACKs", these may be used for congestion control as would normally be performed by TCP. Consequently, only the first received ACK may be used to schedule re-transmission and maintain re-transmission buffers across all sub-flows. However, in computing the RTT value for a particular sub-flow, the ACK for that particular sub-flow may be used.

[0043] To keep data synchronized across all re-transmissions, the following process may be used. First, the least congested sub-flow may be determined, (e.g., the sub-flow having the smallest $\beta\times$RTT value, where β is a parameter of TCP flow control, each TCP sub-flow having its own β and RTT). Whenever the time since the last transmission exceeds a transmission threshold on the least congested sub-flow, a retransmission is made on that sub-flow. The exact same data may be transmitted on all other sub-links for which the time since the last transmission exceeds a transmission threshold associated with that particular sublink, (i.e., the β×RTT of that particular sub-flow). This process may preserve the congestion control properties of TCP over each sub-link, while keeping all sub-links perfectly synchronized.

[0044] An alternative implementation of MCTP-SCC may be to use a UDP underlay as described herein. Usage of the UDP as an L4 underlay protocol may allow any number of flow control and acknowledgement mechanisms to be employed by an MCTP-SCC, without the restriction of being compatible with TCP flow control. Thus, while this approach may add additional overhead (UDP) headers, it may offer more flexibility in the design of the MCTP.

[0045] In the case of MCTP-SCC using a UDP underlay protocol, there may be more freedom of operation and several methods to address the congestion problem are described herein. In one approach, a terminal may monitor incoming traffic on an MCTP-SCC. Because the terminal may presume that its peer terminal sends traffic on all links, a long-term averaged packet arrival rate may be similar across all links. If one of the links is observed as lacking in packet arrivals, this may indicate congestion. Such congestion may be communicated to the peer terminal via a control message (measurement communication). One way to do so would be to augment the ACK to indicate which source IP addresses are being acknowledged so that, for each ACK, the transmitter knows which outgoing interfaces provided the better performance.

[0046] In another approach, (which may work under an assumption that congestion is symmetric), a receiver chooses to not transmit over the interface where it consistently received very little data, or to not transmit to an IP address from which it receives very little data. The latter approach, (to select to not transmit from or to some IP addresses), may suffer from a collapse issue whereby, once a source-destination IP address pair is unused by both peers, there is no way to detect that the particular sub-flow is no longer congested. Thus, a softer variant of this approach may work better whereby a sender monitors traffic arriving at an interface (the sender's own IP address) and from an interface (destination IP address). A solution where the complete pair (own IP address, destination IP address) is a sub-flow is considered. The terminal may maintain a list of all possible sub-flows and, for each sub-flow, it assigns a quality value.

[0047] Every packet may be transmitted over a sub-flow with the best quality value. For sub-flows with lower quality values, packets may be transmitted with a likelihood that is dependent on the ratio of the quality of that sub-flow and the best sub-flow. That is, if q^* is the best sub-flow quality and q is the quality of a given sub-flow, then the likelihood of a sub-flow being transmitted depends on some function $f(q/q^*)$, where f() is a monotonically, (but not strictly monotonically), increasing function on the interval [0,1] and $f(1) = 1$, (needed to ensure that the best sub-flow always gets used) and $f(0) \geq 0$.

[0048] Moreover, the value $f(q/q^*)$ may determine a deterministic ratio that a sub-flow is used to transmit a packet. For example, for $f = 0.2$, every 2 out of 5 packets may be sent. Alternatively, the value $f(q/q^*)$ may determine the probability that a sub-flow may be used to transmit a packet.

[0049] The quality, q, of a sub-flow may be updated as follows. Whenever a packet is received on a given sub-flow first, its quality, q, may go up and the quality of every other sub-flow may be reduced. The value by which the quality goes up or down may be dependent on history. For example, the value by which sub-flow quality goes up may depend on the number of consecutive packets received first on that sub-flow. The value by which sub-flow quality goes down may depend on the number of consecutive packet receptions where the given sub-flow was not a winner. The value by which sub-flow quality goes down may depend on how long it takes for a particular packet to arrive after it has been received over another sub-flow, (and the packet may be discarded). Since some packets may be lost, an upper limit may be needed.

[0050] Finally, a sub-flow may be defined by the pair of IP addresses (own/peer). However, this definition may be reduced to only one's own IP address (to/from any peer IP address), or any peer IP address (to/from any one's own IP address). The operational definition may be the same, although different parameters may have to be chosen to make the system work well.

[0051] MCTP ACK strategy is described herein. An acknowledgement mechanism may be implemented at an MCTP session level, at a sub-flow level, (e.g., TCP already supports its own ACK mechanism), or a combined way to improve robustness or to prevent misleading behavior by middle-boxes such as firewalls, NATs (native address translators), and the like. Specifically, some middle-boxes may pro-actively ACK sub-flows packets, without ensuring that the data reached the receiver. This is a particular problem for MCTP (and MPTCP), as such a "fake" ACK on one sub-flow may potentially impact operation on other sub-flows.

[0052] In a situation where MCTP packets are duplicated over 2 sub-flows, to increase, for instance, reliability, as soon as the packet is acknowledged in a first link, it may be discarded in the transmission (Tx) buffer of the second link, even if it was not acknowledged by the sub-flow acknowledgement mechanism. Thus, just as in the case of synchronized links, packets may be removed from all MCTP sub-links without an explicit ACK being received on that link. In the same manner, the sequence number expected to be received in the next incoming packet may be shared between all MCTP sub-links, (thus avoiding any retransmission requests based on missing sequence numbers). For example, if packets with sequence numbers 1, 2 and 6 are received, normal TCP processing may drop the packet with sequence number 6 and may request retransmission starting from 3).

[0053] Packet redirection in the case of sub-flows closing is described herein. Assuming that MCTP has its own sequence numbering, it may send the packet over different sub-flows based on each sub-flow's performance capabilities. When a sub-flow closes, some MCTP packets may still remain in its buffer. The MCTP acknowledgement mechanism may eventually realize that some unacknowledged packets may have to be resent.

[0054] To optimize this process, the closing sub-flow may report back to an MCTP scheduler which packets have not been sent, so that the MCTP scheduler may move those packets back in other active sub-flows. For example, Figures 5A and 5B illustrate MCTP packet redirection. In Figure 5A, packets 1-7 stored in a queue 505 are processed by an MCTP scheduler 510 such that packets 1, 3, 4 and 6 are placed in a first queue 515 and sent over a first sub-flow 520, and packets 2, 5 and 7 are placed in a second queue 525 and sent over a second sub-flow 530. As shown in Figure 5B, if sub-flow 530 is closed, the packets 2, 5 and 7 may be inserted in the queue 515 with the correct sequence to be sent over the first sub-flow 520, thus preserving the integrity of the batch ACK process of TCP. Alternatively, packets may be dropped from the closed sub-flow and, eventually, the retransmission may recover. Further, packets may be added to the top of the queue, rather than inserting them in order.

[0055] Embodiments

1. A method of establishing an end-to-end multipath connection between two peers using a multi-connection transport protocol (MCTP) architecture, the method comprising:

storing a plurality of packets received from a packet scheduler in a queue in a particular sequence; and

distributing the packets among a plurality of sub-flows, wherein at least two of the sub-flows are synchronized.

2. The method of embodiment 1 further comprising:

controlling congestion of the synchronized sub-flows.

3. The method as in any one of embodiments 1-2 wherein the MCTP optimizes the scheduling of packets received from an application over different sub-flows.

4. The method as in any one of embodiments 1-3 wherein each of the synchronized sub-flows employs a layer 4 (L4) underlay protocol.

5. The method as in any one of embodiments 1-4 further comprising:

sending transmission control protocol (TCP) packets having the same header over each of the synchronized sub-flows to provide naturally synchronized sequence numbers.

6. The method of embodiment 5 further comprising:

generating a first TCP positive acknowledgement (ACK) when a first TCP packet is received over one of the synchronized sub-flows;

discarding duplicate TCP packets received after the first TCP packet is received; and

discarding ACKs received after the first ACK if it corresponds to a TCP packet that was already acknowledged.

7. The method as in any one of embodiments 1-6 wherein each of the synchronized sub-flows employs a user datagram protocol (UDP) underlay.

8. The method as in any one of embodiments 1-7 further comprising:

determining that a transmission threshold has been reached for a specific sub-flow; and

retransmitting data on the specific sub-flow.

9. The method as in any one of embodiments 1-8 wherein each packet is sent over a selected one of the sub-flows based on the performance of the selected sub-flow.

10. The method of embodiment 9 wherein when a sub-flow closes, packets stored in a queue associated with the closed sub-flow are inserted into a queue associated with a queue associated with another one of the sub-flows in a correct sequential order.

11. The method as in any one of embodiments 1-10 wherein each packet is transmitted over a sub-flow that has a best quality value.

12. A multi-connection transport protocol (MCTP) architecture comprising:

a packet scheduler;

a queue configured to store a plurality of packets received from the packet scheduler in a particular sequence; and

a sub-flow manager configured to distribute the packets among a plurality of sub-flows, wherein at least two of the sub-flows are synchronized.

13. The MCTP architecture of embodiment 12 further comprising:

a synchronized congestion controller (SCC) configured to control congestion of the synchronized sub-flows.

14. The MCTP architecture as in any one of embodiments 12-13 wherein the MCTP architecture is used to establish an end-to-end multipath connection between two peers and optimize the scheduling of packets received from an application over different sub-flows.

15. The MCTP architecture as in any one of embodiments 12-14 wherein each of the synchronized sub-flows employs a layer 4 (L4) underlay protocol.

16. The MCTP architecture as in any one of embodiments 12-15 wherein transmission control protocol (TCP) packets having the same header are sent over each of the synchronized sub-flows to provide naturally synchronized sequence numbers.

17. The MCTP architecture of embodiment 16 wherein a first TCP positive acknowledgement (ACK) is generated when a first TCP packet is received over one of the synchronized sub-flows, duplicate TCP packets received after the first TCP packet is received are discarded, and ACKs received after the first ACK are discarded if they correspond to a TCP packet that was already acknowledged.

18. The MCTP architecture as in any one of embodiments 12-17 wherein each of the synchronized sub-flows employs a user datagram protocol (UDP) underlay.

19. The MCTP architecture as in any one of embodiments 12-18 wherein if it is determined that a transmission threshold has been reached for a specific sub-flow, data is retransmitted on the specific sub-flow.

20. The MCTP architecture as in any one of embodiments 12-19 wherein each packet is sent over a selected one of the sub-flows based on the performance of the selected sub-flow.

21. The MCTP architecture of embodiment 20 wherein when a sub-flow closes, packets stored in a queue associated with the closed sub-flow are inserted into a queue associated with a queue associated with another one of the sub-flows in a correct sequential order.

22. The MCTP architecture as in any one of embodiments 12-21 wherein each packet is transmitted over a sub-flow that has a best quality value.

[0056] Although features and elements are described above in particular combinations, one of ordinary skill in the art will appreciate that each feature or element may be used alone or in combination with any of the other features and elements. In addition, the embodiments described herein may be implemented in a computer program, software, or firmware incorporated in a computer-readable medium for execution by a computer or processor. Examples of computer-readable media include electronic signals, (transmitted over wired or wireless connections), and computer-readable storage media. Examples of computer-readable storage media include, but are not limited to, a read only memory (ROM), a random access memory (RAM), a register, a cache memory, a semiconductor memory device, a magnetic media, (e.g., an internal hard disc or a removable disc), a magneto-optical media, and an optical media such as a compact disc (CD) or a digital versatile disc (DVD). A processor in association with software may be used to implement a radio frequency transceiver for use in a WTRU, UE, terminal, base station, Node-B, eNB, HNB, HeNB, AP, RNC, wireless router or any host computer.

* * *

CLAIMS

What is claimed is:

1. A method of establishing an end-to-end multipath connection between two peers using a multi-connection transport protocol (MCTP) architecture, the method comprising:

storing a plurality of packets received from a packet scheduler in a queue in a particular sequence;

distributing the packets among a plurality of sub-flows, wherein at least two of the sub-flows are synchronized; and

controlling congestion of the synchronized sub-flows.

2. The method of claim 1 wherein the MCTP optimizes the scheduling of packets received from an application over different sub-flows.

3. The method of claim 1 wherein each of the synchronized sub-flows employs a layer 4 (L4) underlay protocol.

4. The method of claim 1 further comprising:

sending transmission control protocol (TCP) packets having the same header over each of the synchronized sub-flows to provide naturally synchronized sequence numbers.

5. The method of claim 4 further comprising:

generating a first TCP positive acknowledgement (ACK) when a first TCP packet is received over one of the synchronized sub-flows;

discarding duplicate TCP packets received after the first TCP packet is received; and

discarding ACKs received after the first ACK if it corresponds to a TCP packet that was already acknowledged.

6. The method of claim 1 wherein each of the synchronized sub-flows employs a user datagram protocol (UDP) underlay.

7. The method of claim 1 further comprising:

determining that a transmission threshold has been reached for a specific sub-flow; and

retransmitting data on the specific sub-flow.

8. The method of claim 1 wherein each packet is sent over a selected one of the sub-flows based on the performance of the selected sub-flow.

9. The method of claim 8 wherein when a sub-flow closes, packets stored in a queue associated with the closed sub-flow are inserted into a queue associated with a queue associated with another one of the sub-flows in a correct sequential order.

10. The method of claim 1 wherein each packet is transmitted over a sub-flow that has a best quality value.

11. A multi-connection transport protocol (MCTP) architecture comprising:

a packet scheduler;

a queue configured to store a plurality of packets received from the packet scheduler in a particular sequence;

a sub-flow manager configured to distribute the packets among a plurality of sub-flows, wherein at least two of the sub-flows are synchronized; and

a synchronized congestion controller (SCC) configured to control congestion of the synchronized sub-flows.

12. The MCTP architecture of claim 11 wherein the MCTP architecture is used to establish an end-to-end multipath connection between two peers and optimize the scheduling of packets received from an application over different sub-flows.

13. The MCTP architecture of claim 11 wherein each of the synchronized sub-flows employs a layer 4 (L4) underlay protocol.

14. The MCTP architecture of claim 11 wherein transmission control protocol (TCP) packets having the same header are sent over each of the synchronized sub-flows to provide naturally synchronized sequence numbers.

15. The MCTP architecture of claim 14 wherein a first TCP positive acknowledgement (ACK) is generated when a first TCP packet is received over one of the synchronized sub-flows, duplicate TCP packets received after the first TCP packet is received are discarded, and ACKs received after the first ACK are discarded if they correspond to a TCP packet that was already acknowledged.

16. The MCTP architecture of claim 11 wherein each of the synchronized sub-flows employs a user datagram protocol (UDP) underlay.

17. The MCTP architecture of claim 11 wherein if it is determined that a transmission threshold has been reached for a specific sub-flow, data is retransmitted on the specific sub-flow.

18. The MCTP architecture of claim 11 wherein each packet is sent over a selected one of the sub-flows based on the performance of the selected sub-flow.

19. The MCTP architecture of claim 18 wherein when a sub-flow closes, packets stored in a queue associated with the closed sub-flow are inserted into a queue associated with a queue associated with another one of the sub-flows in a correct sequential order.

20. The MCTP architecture of claim 11 wherein each packet is transmitted over a sub-flow that has a best quality value.

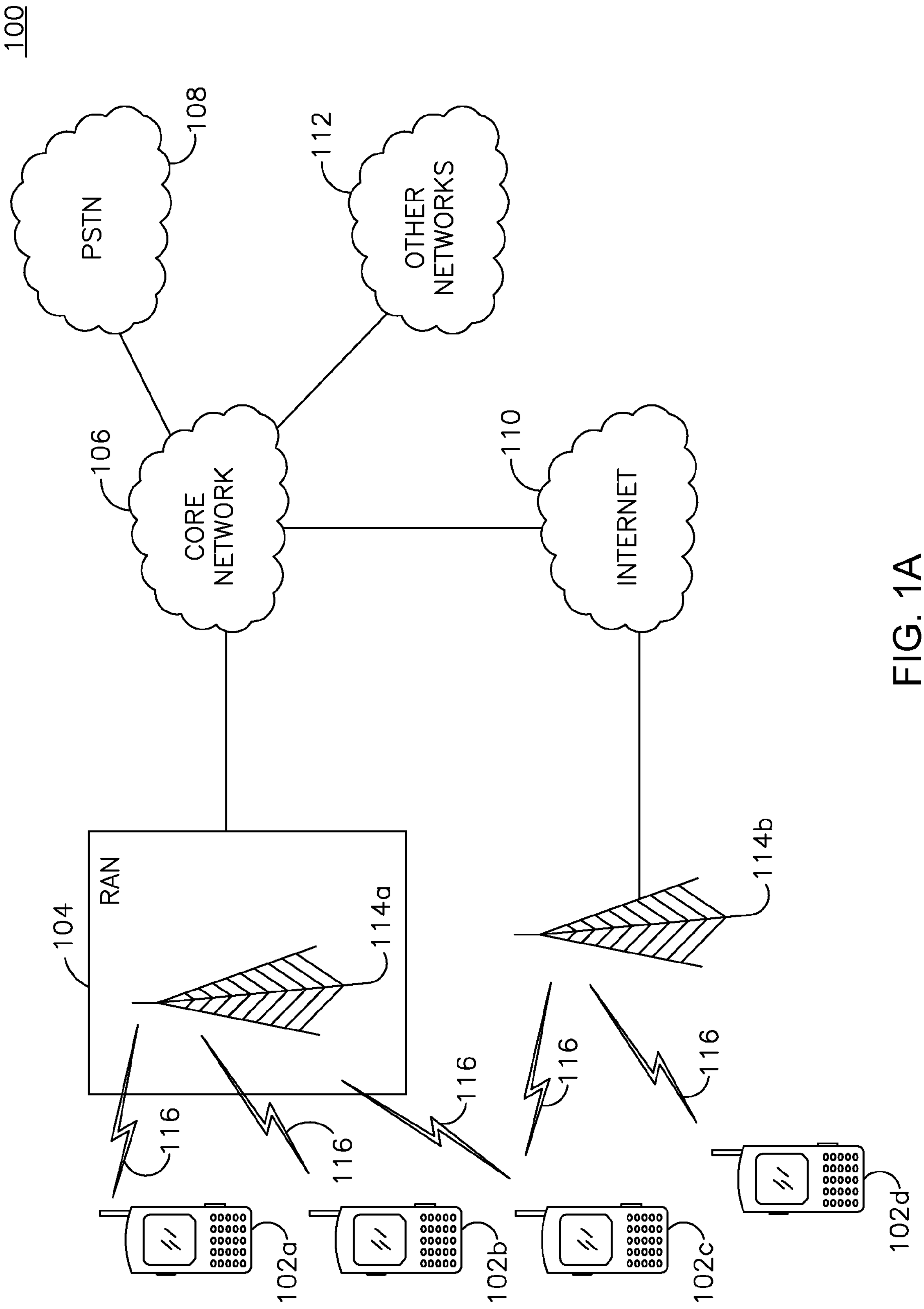
100
PSTN
108
OTHER NETWORKS
112
CORE NETWORK
106
INTERNET
110
RAN
104
114a
114b
116
102a
102b
102c
102d

FIG. 1A

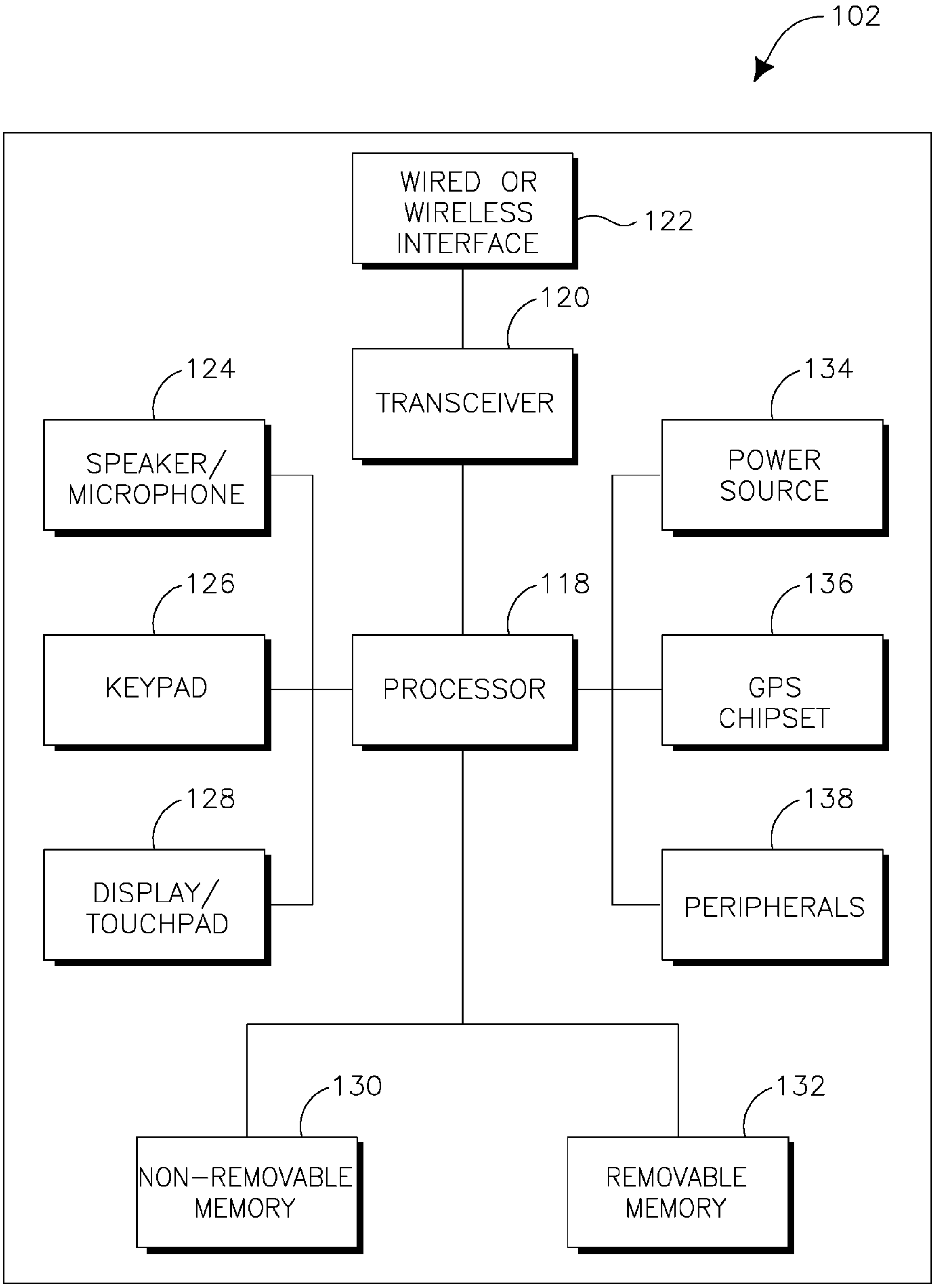
102
WIRED OR WIRELESS INTERFACE
122
120
TRANSCEIVER
124
SPEAKER/ MICROPHONE
134
POWER SOURCE
126
KEYPAD
118
PROCESSOR
136
GPS CHIPSET
128
DISPLAY/ TOUCHPAD
138
PERIPHERALS
130
NON-REMOVABLE MEMORY
132
REMOVABLE MEMORY

FIG. 1B

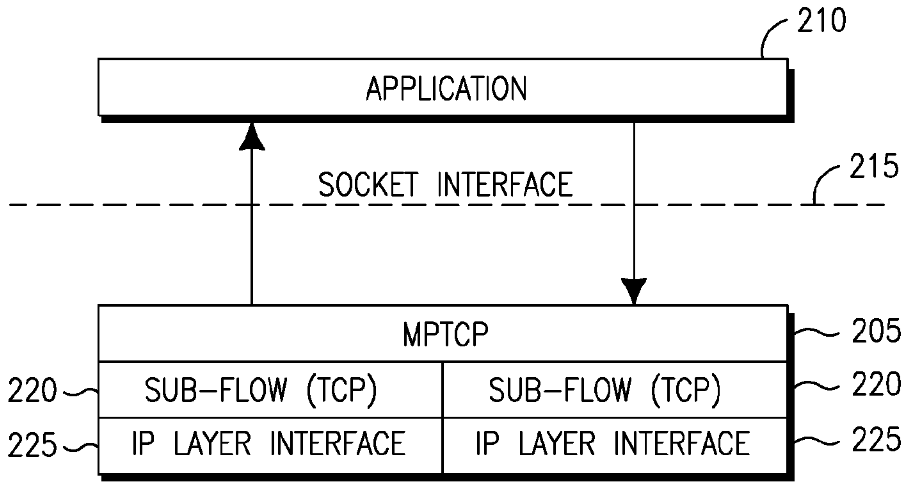
210
APPLICATION
SOCKET INTERFACE
215
MPTCP
205
220
SUB-FLOW (TCP)
SUB-FLOW (TCP)
220
225
IP LAYER INTERFACE
IP LAYER INTERFACE
225

FIG. 2

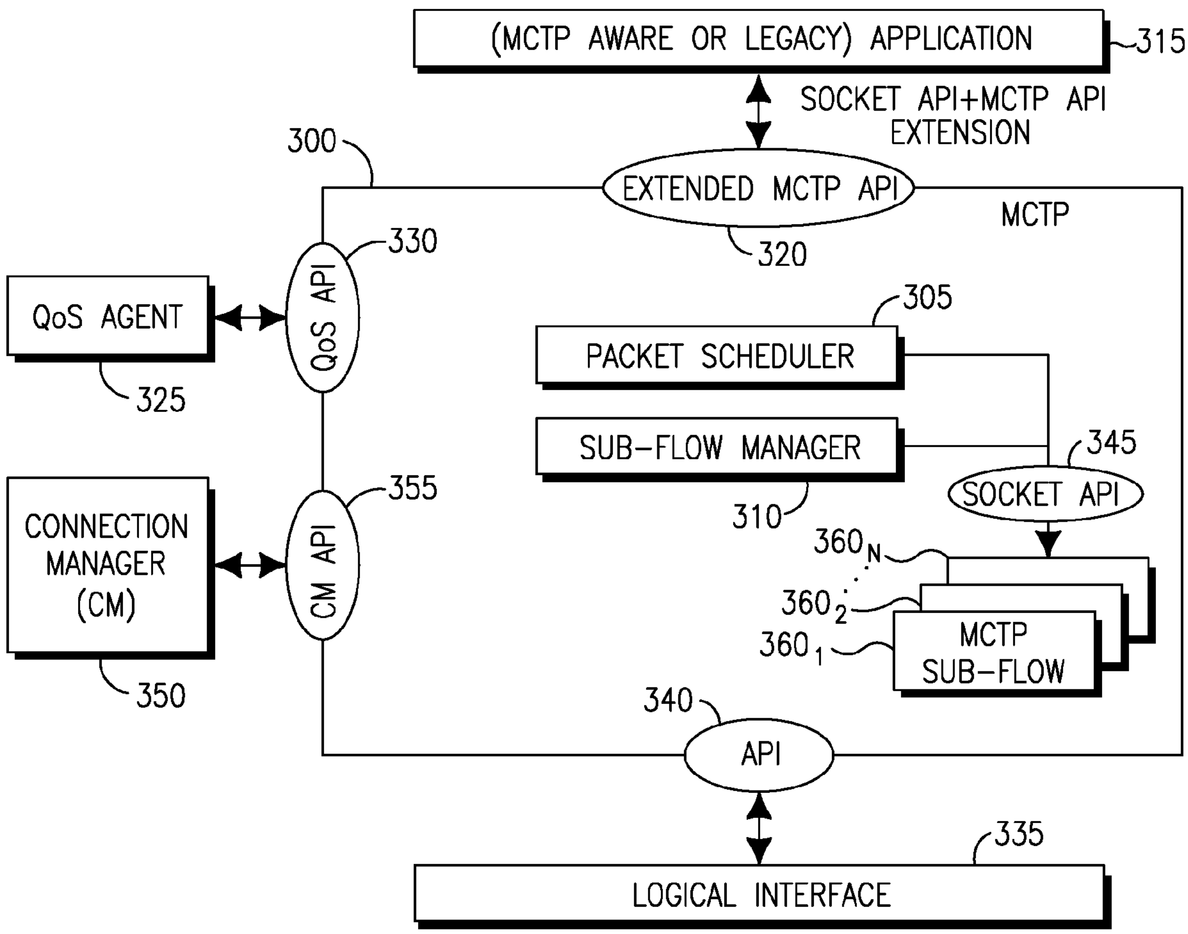
(MCTP AWARE OR LEGACY) APPLICATION
315
SOCKET API+MCTP API EXTENSION
300
EXTENDED MCTP API
MCTP
320
330
QoS API
QoS AGENT
325
305
PACKET SCHEDULER
SUB-FLOW MANAGER
345
310
SOCKET API
355
CM API
CONNECTION MANAGER (CM)
350
360N
360 2
360 1
MCTP SUB-FLOW
340
API
335
LOGICAL INTERFACE

FIG. 3

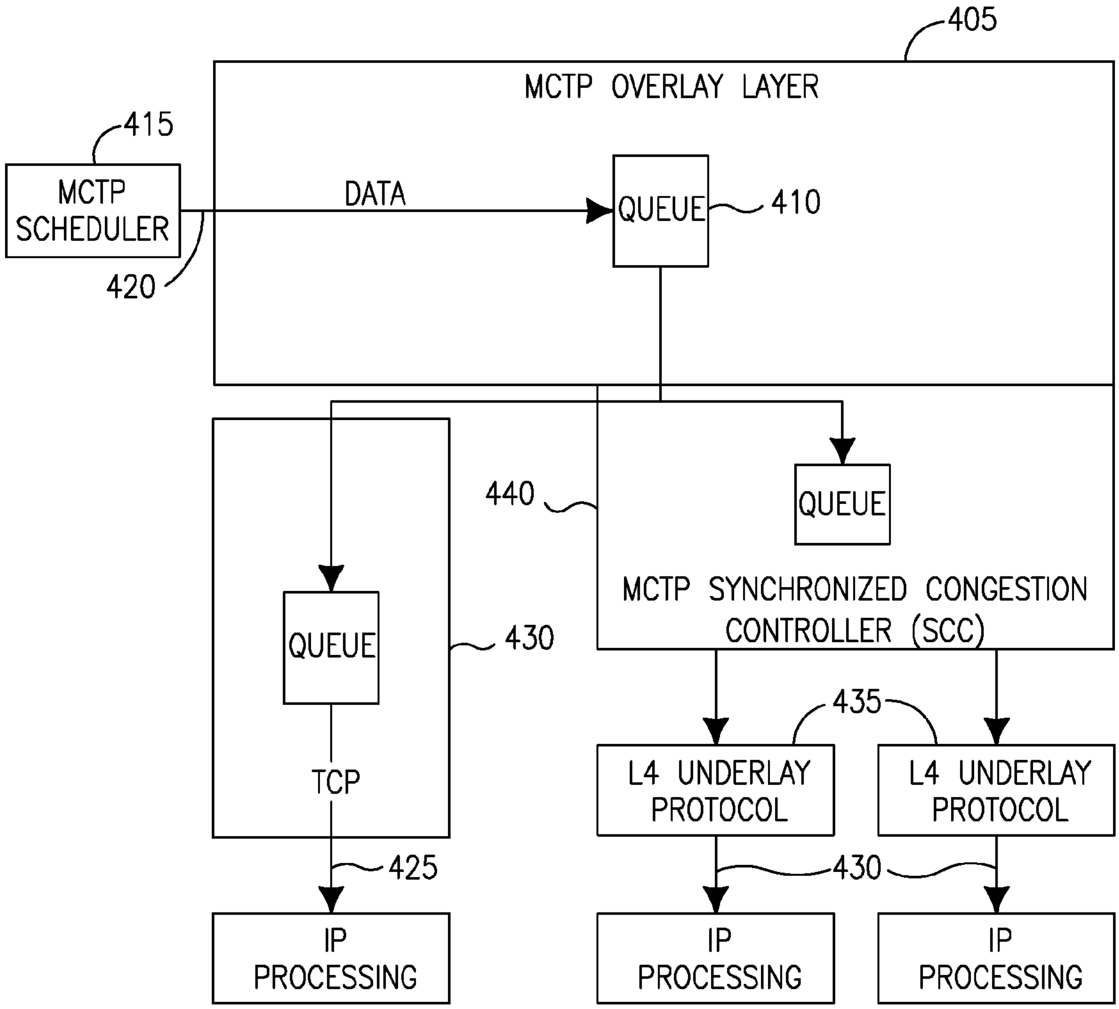
405
MCTP OVERLAY LAYER
415
MCTP SCHEDULER
DATA
QUEUE
410
420
440
QUEUE
MCTP SYNCHRONIZED CONGESTION CONTROLLER (SCC)
QUEUE
430
TCP
435
L4 UNDERLAY PROTOCOL
L4 UNDERLAY PROTOCOL
425
430
IP PROCESSING
IP PROCESSING
IP PROCESSING

FIG. 4

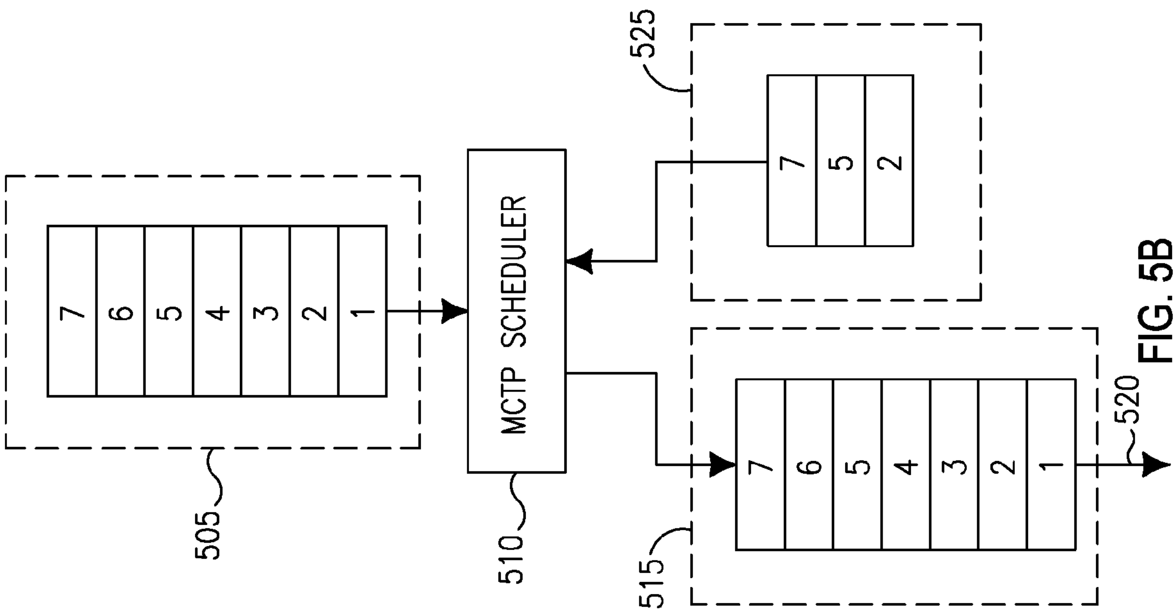
505
510
515
520
525
MCTP SCHEDULER
7
6
5
4
3
2
1

FIG. 5B

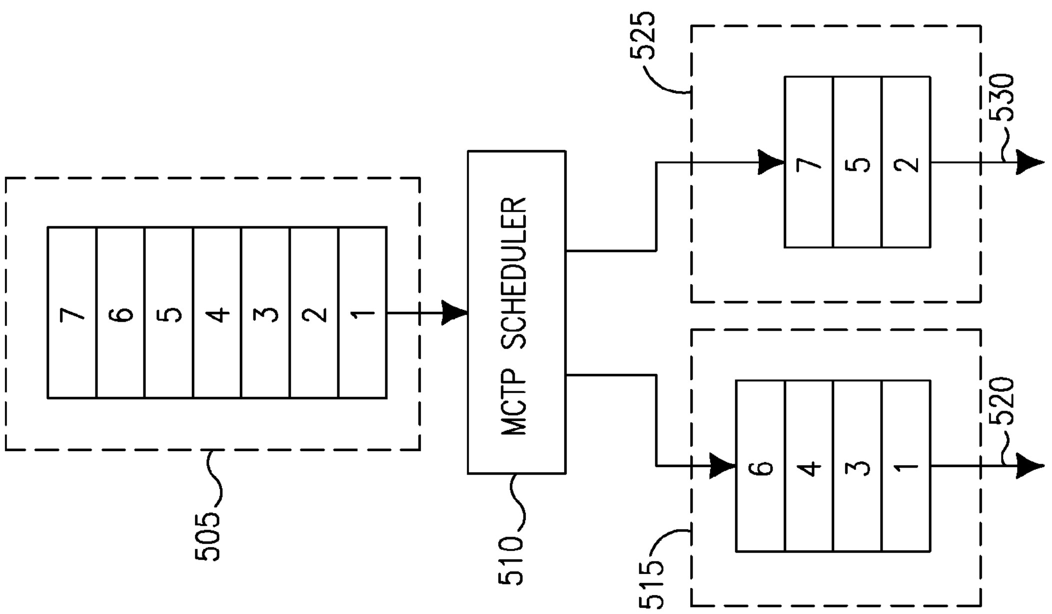
505
510
515
520
525
530
MCTP SCHEDULER
7
6
5
4
3
2
1

FIG. 5A

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2011/064592

A. CLASSIFICATION OF SUBJECT MATTER
INV. H04L29/06 H04L12/56
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
H04L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, COMPENDEX

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	FORD ROKE MANOR RESEARCH C RAICIU M HANDLEY UNIVERSITY COLLEGE LONDON S BARRE UNIVERSITE CATHOLIQUE DE LOUVAIN J IYENGAR FRANKLIN: "Architectural Guidelines for Multipath TCP Development; draft-ietf-mptcp-architecture-05.txt", ARCHITECTURAL GUIDELINES FOR MULTIPATH TCP DEVELOPMENT; DRAFT-IETF-MPTCP-ARCHITECTURE-05.TXT, INTERNET ENGINEERING TASK FORCE, IETF; STANDARDWORKINGDRAFT, INTERNET SOCIETY (ISOC) 4, RUE DES FALAISES CH- 1205 GENEVA, SWITZERLAND, no. 5, 21 January 2011 (2011-01-21), pages 1-29, XP015073614, [retrieved on 2011-01-21]	1-3,6, 8-13,16, 18-20
Y	page 13 - page 14; figure 7 figure 6 page 15 - page 16	4,5,7, 14,15,17

-/--

[X] Further documents are listed in the continuation of Box C.

[ ] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
5 March 2012	14/03/2012
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Authorized officer Tsuchiya, Kuni

3

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	----- ROJVIBOONCHAI K ET AL: "AN EVALUATION OF MULTI-PATH TRANSMISSION CONTROL PROTOCOL (M/TCP) WITH ROBUST ACKNOWLEDGEMENT SCHEMES", IEICE TRANSACTIONS ON COMMUNICATIONS, COMMUNICATIONS SOCIETY, TOKYO, JP, vol. E87-B, no. 9, 1 September 2004 (2004-09-01), pages 2699-2707, XP001209303, ISSN: 0916-8516	1,2,6, 11,12,16
Y	figure 2 3.1 and 3.2; page 2700 - page 2701 -----	4,5,7, 14,15,17

3