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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 January 2014 (03.01.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/001773 A1

(51) International Patent Classification:
H04L 12/26 (2006.01)

(21) International Application Number:
PCT/GB2013/051652

(22) International Filing Date:
24 June 2013 (24.06.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
1211323.9 26 June 2012 (26.06.2012) GB

(71) Applicant: **BAE SYSTEMS PLC** [GB/GB]; 6 Carlton Gardens, London SW1Y 5AD (GB).

(72) Inventors: **JARVIS, Richard Thomas**; BAE SYSTEMS Detica, Surrey Research Park, Guildford, Surrey GU2 7YP (GB). **FURLEY, Paul Michael**; BAE SYSTEMS Detica, Surrey Research Park, Guildford, Surrey GU2 7YP (GB). **KEENE, Henri William**; BAE SYSTEMS Detica, Surrey Research Park, Guildford, Surrey GU2 7YP (GB).

(74) Agent: **BAE SYSTEMS PLC, GROUP IP DEPT**; PO Box 87, Farnborough Aerospace Centre, Farnborough, Hampshire GU14 6YU (GB).

(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, BN, 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, 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, PA, 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.

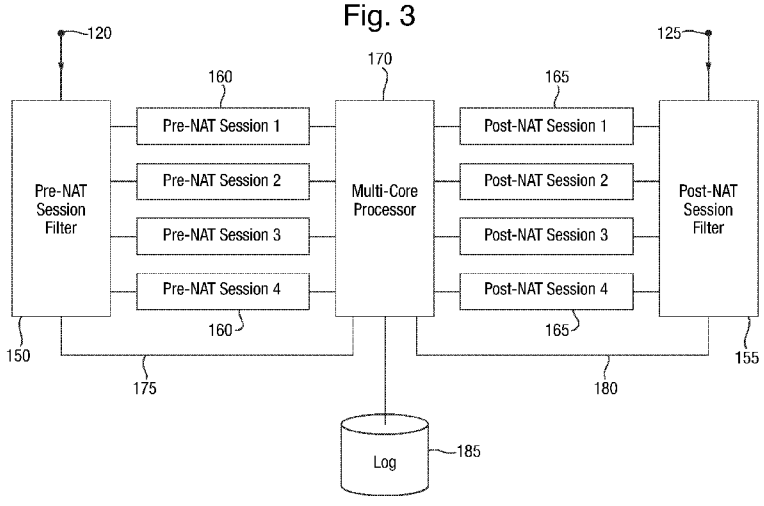
(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, 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, KM, ML, MR, NE, SN, TD, TG).

Published:

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

(54) **Title**: RESOLUTION OF ADDRESS TRANSLATIONS

Fig. 3



(57) **Abstract**: A method and apparatus are provided for mapping IP communications sessions, e.g. TCP or UDP sessions, before and after IP source address and port number translation by a network Address Translation/Port Address Translation (NAT/PAT) device. Data packets outgoing from a network may be captured before and after the NAT/PAT device and processed in a passive correlation device according to the present invention to map pre-NAT and post-NAT IP quint data fields.

WO 2014/001773 A1

RESOLUTION OF ADDRESS TRANSLATIONS

This invention relates to address translation and in particular, but not exclusively, to the detection of address translations across a point of interconnection between networks, for example between two Internet Protocol (IP) networks having incompatible addressing schemes.

Network Address Translation (NAT)/Port Address Translation (PAT), as described for example in RFC 2663, "IP Network Address Translator (NAT) Terminology and Considerations", by P. Srisuresh and M. Holdrege, August 1999, published by the Internet Society, has been devised as a technique for mapping Internet Protocol (IP) addresses from those used in one network to those used in another, for example between a private addressing scheme used within a corporate network and a global addressing scheme as used for the public Internet. Such a technique enables a private addressing scheme to be hidden from a public network behind a single IP address or a small number of IP addresses. This is achieved by means of a device that is able to make appropriate alterations to network addressing information in the headers of IP packets outgoing from and incoming to a network and to maintain a translation table so that returning packets within a particular communications session can be correctly routed to the originating IP address.

The use of NAT/PAT devices can cause problems for compliance systems, for example, in that an examination of IP packets arriving at a network destination or being carried over a public network will not necessarily provide a unique identifier for the originator. Moreover, the only source of information on the mapping to an originator is a transient record created within a respective NAT/PAT device, a record that exists, typically, only for the duration of a particular session, e.g. a Transmission Control Protocol (TCP) session, a User Datagram Protocol (UDP) session or an Internet Control Message Protocol (ICMP) query session. The NAT/PAT devices themselves are not typically available for remote interrogation and the high data volumes handled make longer term storage of mappings infeasible.

From a first aspect, the present invention resides in an apparatus for mapping data packets undergoing changes to their addressing information between a first point and a second point in a network, the apparatus comprising:

first data capture means for capturing data packets at a first tap point in a communications path, prior to address translation;

second data capture means for capturing data packets at a second tap point in a communications path, following address translation; and

passive correlating means for detecting mappings between data packets captured by the second data capture means and data packets captured by the first data capture means and for outputting a detected mapping between addressing information before and after translation.

In a preferred embodiment, the first and second data capture means comprise means for associating data packets within a same identified data stream and for organising associated data packets into processing queues, establishing a different processing queue for each identified data stream. In this way, the correlating means may be arranged to read packets from each processing stream in a parallel processing arrangement.

Preferably, the first and second data capture means are arranged to associate data packets in a given data stream by determining a hash value for a predetermined combination of data fields in each data packet and maintaining a hash table mapping each distinct determined hash value to those packets having the same determined hash value.

The present invention finds particular application in the mapping of network addresses within data streams comprising TCP or UDP sessions over an IP network. In particular, the apparatus of the present invention may be deployed to capture data from points either side of a NAT/PAT device and to passively resolve mappings between pre-NAT and post-NAT source 'IP quint' data fields.

Preferred embodiments of the present invention will now be described, by way of example only, with reference to the accompanying drawings of which:

Figure 1 shows an example of a communications path between a user of a mobile communications device and an internet service being accessed from that device;

Figure 2 shows a simplified communications arrangement with a deployed correlation device according to preferred embodiments of the present invention;

Figure 3 shows preferred functional elements in a correlation device according to the present invention; and

Figure 4 shows an example of an output by the correlation device of the present invention.

In a typical implementation of a Network Address Translation (NAT) device that a mobile communications service provider (CSP) may deploy, each subscriber to the mobile CSP's services will be allocated a unique private IP address in the 10.0.0.0/8 range, assigned by means of a *RADIUS/DHCP* or an analogous accounting system. A NAT/PAT device is deployed at the border between the mobile CSP's network and the wider Internet. Figure 1 shows a typical communications path between a mobile communications device of a subscriber to the CSP's services wishing to access an internet-based service. Corresponding network architectures may be envisaged in which, for example, users in a corporate fixed line network wish to access the same internet-based service.

Referring to Figure 1, a subscriber 10 to a CSP's mobile communications services is able to use a mobile communications device 15, for example a so-called "smart phone" or other mobile communications device, to access an internet-based service 20 over the CSP's private General Packet Radio Service (GPRS) or equivalent mobile communications network and the public internet 25. In the CSP's private network, the mobile communication device 15 communicates wirelessly with one or more local based stations 30 and a communications path is established through a serving GPRS support node (SGSN) 35 and its corresponding gateway GPRS support node (GGSN) 40 to a NAT/PAT device 45 deployed at the boundary between the CSP's network and the public internet 20.

At a notional point 50 in the communication path between the GGSN 40 and the NAT/PAT device 45, IP data packets being carried over the CSP's network in a typical TCP/UDP communications session, established between the subscriber's device 15 and the internet-based service 20, carry a private source IP address allocated within the CSP's network in respect of that subscriber's device 15 and a source TCP/UDP port number. In this particular example, the NAT/PAT device 45 is arranged to hide the private IP addressing scheme used within the CSP's network by substituting the private source IP address carried in the headers of all outgoing IP data packets with one common public routeable source IP address (or one of a small number of common public source IP addresses) for all internet communications sessions with subscribers in that particular CSP's network. The NAT/PAT device 45 also allocates a unique TCP/UDP source port number for that particular session and substitutes the original source port number with the newly allocated port number. The modified packets are then forwarded through a firewall/gateway device 60 to the internet 25. At the same time, the NAT/PAT device 45 maintains a record in the form of a state table of the IP address and port allocations it has made so that when IP packets inbound to the network are received within a particular session, the destination IP address and port number of each received packet may be translated back to the original private source IP address for that session and routed to the originating mobile subscriber's device 15.

To external hosts, traffic originating from a mobile CSP's network using NAT, as would be seen for example at a notional point 65 outgoing from the NAT/PAT device 45, or at a point 70 upon arrival at the internet-based service 20, will appear to come from the comparatively small IP address range that the NAT assigns to outbound traffic. An individual external NAT IP address may have been shared by many subscribers within the CSP's network and it is therefore not possible to uniquely identify an originating mobile subscriber by their IP address alone. Additional information on session mapping between internal and external IP addresses must be captured to be able to identify a particular mobile subscriber's session with any level of certainty.

An ability to identify sessions owned by particular subscribers at a later date is a valuable asset to Law Enforcement (LE). Seized hardware from

servers hosting illegal material may conceivably contain logs of source IP addresses. Where those source IP addresses originate in networks with NAT devices in place, the source IP address logs enable LE to identify only the originating network. The individual subscriber on that network cannot be identified. If NAT logs have not been retained at the originating network, for example by the mobile CSP, then it may be impossible to link the external IP address captured in the source IP address log to an individual mobile subscriber. Preferred embodiments of the present invention are arranged to capture the information necessary to make such links. Such information would also be of use in a Lawful Interception system where traffic being monitored is captured on the public side of a NAT device, after address translation. Real-time mapping information of particular target sessions, captured by the present invention, may be sent to a respective monitoring device so that it may identify and collect data for the correct sessions.

Whereas, in some circumstances, it may be possible to obtain live data feeds from NAT/PAT devices, including syslog or netflow data, the present invention offers an entirely passive solution to the problem of linking source IP addressing to NAT/PAT allocated addressing. A preferred embodiment will now be described with reference to Figure 2.

Referring to Figure 2, in a simplification of the communications path of Figure 1, a user's terminal equipment 100 is shown communicating with a remote server 105 over a communications path that includes a NAT/PAT device 110 for performing source IP address translations. A passive correlation device 115 according to the present invention is deployed to monitor data packets at first and second tap points 120, 125 in the communication path, the first tap point 120 for monitoring outgoing data packets before they enter the NAT/PAT device 110 and the second tap point 125 for monitoring data packets emerging from the NAT/PAT device 110.

The correlation device 115 is arranged to implement a preferred correlation technique, to be described in detail below, for processing the monitored pre-NAT and post-NAT data packets and for deriving address and port mappings made by the NAT/PAT device 110, substantially in real-time. Such mappings may then be made available more or less rapidly for use in

numerous applications requiring the identification of a true source address in a particular IP session being observed at some point downstream from the originator's network.

A TCP/UDP session is identifiable by five data fields in a TCP or UDP packet header: the source IP address and port number; the destination IP address and port number; and the IP Protocol. This combination of IP addressing information is known as an IP quint, or IPQ. However, a NAT/PAT device 110 replaces the source IP address and port number, as described above, for various reasons, retaining a record of the mapping between originating and translated source data so that returning packets within the same TCP or UDP session may be delivered to their originator. Therefore, in order for the correlation device 115 of the present invention to associate a data packet entering the NAT/PAT device 110 with one leaving it, an analysis on only the destination IP address, port number and IP protocol fields may not always enable a one-to-one correspondence to be established, for example if two users in a mobile network access the same web site at substantially the same time.

Preferred steps in a simplified top-level process, as may be performed by the correlation device 115, for deriving the mappings made by the NAT/PAT device 110 may be summarised as follows:

1) at a first, pre-NAT tap point, capture outgoing data packets and associate captured packets within the same TCP/UDP session, identifiable from a comparison of their pre-NAT IPQs, and queue data packets from each identified session in a different processing queue;

2) at a second, post-NAT tap point, capture outgoing data packets and associate those with the same IPQs, as for the pre-NAT capture, queued using a different queue for packets with each distinct IPQ;

3) on receipt of a first packet from the second tap point, analyse one or more packets in each pre-NAT session queue looking for a match on at least the destination IP address, destination port number and IP protocol fields;

4) if only one pre-NAT session queue includes packets with the matching destination IP address and port number and IP protocol fields, then a unique mapping has been captured between the pre-NAT IPQ and the post-NAT IPQ;

5) if more than one pre-NAT session queue includes matching packets, then a comparison to a greater depth is required, for example comparing other fields within the IP header of pre-NAT queued packets, not altered by the NAT/PAT device 110, with the corresponding fields in the received post-NAT packet until a unique pre-NAT IP session is matched to the post-NAT packet. However, it may be decided to report an ambiguity, listing what is likely to be a relatively small number of alternative pre-NAT session IPQs rather than continue to expend processing effort once the number of possible matches drops below a certain threshold number.

On detecting a match between a pre-NAT and post-NAT IPQ, the association is recorded by the correlation device 115 in a log. Logged associations are output periodically to receiving applications with whatever frequency is required.

Various implementation innovations are provided to enable the correlation device 115 to carry out the above processing steps at a speed sufficient to match the rate of address translation by the NAT/PAT device 110. Such innovations will now be described with reference to Figure 3 which shows a simplified functional block diagram of a preferred correlation device 115. The preferred functionality of the correlation device will be described mainly in the context of processing data packets relating to TCP/UDP sessions. However, it would be apparent to a person of ordinary skill in the data communications field how to apply the principles described to the processing of packets relating to other protocols.

Referring to Figure 3, a pre-NAT session filter 150 is provided to receive outgoing IP packets captured at the first tap point 120 (as shown in Figure 2) immediately before entering a NAT/PAT device (not shown in Figure 3). A similarly functioning post-NAT session filter 155 is provided to receive outgoing IP packets emerging from the NAT/PAT device, captured at the tap point 125. The pre-NAT session filter 150 is arranged to examine at least the IPQs of the received IP packets and to associate those packets in the same TCP/UDP session, as distinguished by IPQ, organising packets from each distinct session into different pre-NAT session queues 160, four different session queues 160 being shown in Figure 3. Similarly, the post-NAT session filter 155 examines at

least the IPQs of received packets and organises packets with each distinct IPQ into different post-NAT session queues 165. While IP packets captured at the second tap point 125 may be associated by IPQ, they do not necessarily correspond to different sessions, as for the pre-NAT IP packets, due to the changes to source IP address and port number made by the NAT/PAT device. However, in a preferred embodiment, each of the session filters 150, 155 may be arranged to queue IP packets on the basis of the combination of IPQ and IP Identification fields so as to increase the chances of distinguishing post-NAT packets in different sessions with only a small additional processing overhead.

A multi-core processor 170 is provided to perform the main correlation functions of the correlation device 115. In the specific example implementation shown in Figure 3, the processor 170 comprises four processor cores with each processor core arranged to execute, in a separate processing thread, an instance of the correlation functionality. Each of the executing correlation threads is arranged to receive queued IP packets from a different post-NAT session queue 165, in parallel, and to look for a matching session from amongst the queued IP packets in the pre-NAT session queues 160. The processor 170 is arranged to receive and to output queue management control signals over notional control signal paths 175 and 180 to the pre-NAT and post-NAT session filters respectively. In particular, when a pre-NAT session (160) has been matched to a packet in a post-NAT session queue 165, the processor signals to the post-NAT session filter 165 to cease capture of IP packets in the matched session (post-NAT IPQ + IP Identification field). Similarly, the processor 170 signals to the pre-NAT session filter to cease capture of IP packets relating to the matched pre-NAT session. Each session filter 150, 155 responds by clearing the respective queues 160, 165 of packets and begins to queue IP packets with a newly identified IPQ + IP Identification field, captured at the respective tap points 120, 125. In this way, the loading on the processor's correlation functionality is reduced as far as possible.

Details of the matched pre-NAT and post-NAT IPQs are output by the processor 170 to a log 185 which may comprise volatile memory or a persistent storage device, accessible to other processes for reporting of matched pre-NAT to post-NAT IPQs to external systems.

To identify IP packets in distinct sessions, each of the session filters 150, 155 executes a hashing function on the IPQ + IP Identification fields read from each received packet and maintains respective hash maps relating each distinct hash value to a session queue (160, 165) identifier. This provides a very rapid way for packets in the same session to be organised into different session queues 160, 165. In practice, the conceptual session queues 160, 165 illustrated in Figure 3 may comprise no more than a list of pointers in the respective hash maps to packets from the same session, the packets being otherwise held in a common buffer. The processor 170 is arranged to process packets from each queue in parallel using multiple instances of the correlation functionality, each instance running on a distinct CPU core of the multi-core processor 170. The preferred queuing arrangement ensures that all packets for a given session or stream (i.e. TCP, UDP or otherwise) are handled by the same processing thread. Furthermore, as all packets for a given session are handled by the same instance of correlation functionality, there is no need for inter-thread communication.

In practice, the processor 170 aims to correlate a single packet from a post-NAT session queue 165, e.g. a TCP packet with the ACK flag set, with a single packet from one of the pre-NAT session queues 160. The detection of a particular type of packet, such as one with the ACK flag set, by the post-NAT session filter 155 may trigger the processor 170 to begin a correlation function using that packet. If successful this represents the most rapid correlation of pre-NAT and post-NAT sessions, likely to be performed in the example implementation above in less than 100ms.

If there are a number of substantially simultaneous sessions established with a common web site by multiple different subscribers to a network, then it is possible that the processor 170 will not be able to establish a one-to-one correlation of pre-NAT and post-NAT sessions using IPQ + IP Identification alone. The session filters 150, 155 are arranged to record timing information to a high level of accuracy for each observed session, recording the time at which each session is first identified in received data and the time of closure (time last seen) of each identified session - being the time at which a correlation is found or the time of a timeout (corresponding to the timeout period applied to sessions

by a NAT/PAT device). By considering the relative timing of sessions, the correlator may be able to resolve an ambiguity. However, if timing is not sufficient to resolve an ambiguity, further fields in the pre-NAT and post-NAT IP packets may need to be examined and compared. In the specific example of a TCP/IP packet, in a preferred order the following additional fields may be examined by the processor 170:

TCP SEQ Number
TCP ACK Number
TCP Options
TCP Flags
Payload

Certain protocols, such as Voice-over-IP (VoIP) and the File Transfer Protocol (FTP) are known to "disobey" the OSI layer model by referencing lower layers, for example by embedding addressing information within the payload of respective IP packets. In the event that an examination of packet payload is to be used to match sessions in the present invention, then in such cases the payload cannot be used directly in the correlation functionality. This is because a typical NAT/PAT device is provided with "application aware" functionality to alter the payload of such packets to replicate, in the payload, the change in addressing that it performs in the IP header fields. However, if required, the processor 170 or the session filters 150, 155 may be arranged to perform a similar alteration to the payload of packets such as VoIP or FTP packets so that the changes made by a NAT/PAT device can be taken into account when correlating sessions. RFC 3027 (Protocol Complications with the IP Network Address Translator), published in January 2001, lists some of the protocols requiring application awareness and the complications arising from NAT.

During normal operation, the processor 170 is arranged to generate four types of message, as follows:

- START – indicates an unambiguous correlation, sent at the point of the correlation.
- END – corresponds to a START message, this indicates that a session that was being analysed has timed out.

- MATCH – indicates a unique correlation between pre-NAT and post-NAT sessions.
- AMBIG – indicates one (of two or more) possible match for a post-NAT IPQ. Multiple AMBIG messages are generated, one for each possible pre-NAT IPQ. This message is sent following an IPQ timeout.
- FAIL – indicates a failure to match a post-NAT IPQ, sent after an IPQ timeout. This may be due to system faults, such as packet loss into the correlation device or bit errors caused by the NAT device or correlation device receiver, or by logical defects in the correlation device. Logical faults may include packet decoding, e.g. unhandled application-aware protocols requiring payload rewriting. If detected, correlation failures are reported to indicate that investigation may be required.

For the purposes of reporting to external systems, a preferred output from the processor 170 when correlations between pre-NAT and post-NAT sessions are found will now be described with reference to Figure 4.

Referring to Figure 4, the output fields preferably include the following:

Match type – whether unique("Match") or ambiguous ("Non-Unique");

Date and start time of each session (to an accuracy of 10^{-8} seconds);

Private Source IP Address;

Private Source Port;

Public Source IP Address;

Public Source Port;

Destination IP Address;

Destination Port; and

IP Protocol.

If required, a separate process may be executed by the correlation device 115 to monitor the output log 185 of the processor 170 and to report the

output data to other interested applications, or to field enquiries by remote applications.

Preferably the first and second tap points 120, 125 are immediately adjacent, in network terms, before and after the NAT/PAT device 110 so that no further changes to the data packets between the first tap point 120 and the second tap point 125 would need to be taken into account by the correlation device, beyond those made by the NAT/PAT device 110 itself. However, in principle, the first tap point 120 may be located further into the network on the user's side, necessitating a certain amount of pre-processing of data by the correlation device 115 in order to extract the IP packets being conveyed within other network-specific protocols. Similarly, the second tap point 125 may be located anywhere in the communications path between the NAT/PAT device 110 and the remote server 105, according to the particular communications sessions that need to be monitored. However, to enable all the outgoing traffic from a network to be captured, or to enable a required rate of data capture to be achieved, it may be necessary to locate the second tap point 125 close to the NAT/PAT device 110 or to the correlation device 115 itself.

In the example of a mobile CSP network, if the only available data is from a first tap point 120 located at the edge of the CSP's core packet network, then the pre-NAT session filter 150 may be required to carry out addition pre-processing steps before organising the captured data into IP session queues, for example to:

Filter only the Gn traffic, identified based on the GTP UDP port;

Remove GTP encapsulation;

Filter only simplex handset traffic – this may be performed to allow only those packets with a source IP in the handset subnets and a destination IP on the internet;

Filter out packets mid-session – on the basis that mid-session TCP packets would be dropped by a NAT/PAT device for not having a valid mapping.

Use of a passive correlation technique to reverse engineer the mapping process performed by the NAT/PAT device 110 has the advantage of being

entirely vendor agnostic, but it requires significant computational capability to implement the techniques described above. In a typical application, the present invention is required to be able to correlate IP packets flowing outbound from a network through a NAT/PAT device 110 at a rate of 10 GBits/s. To carry out its correlation functionality, the correlation device 115 is required to be able to receive packet data both ingoing to and outgoing from the NAT/PAT device, each at a rate of 10 GBits/s, and to process these data packets substantially in real-time. However, to capture both the upstream and downstream sides of a duplex connection, the passive correlation device 115 device would need to process data at 40GBit/s for a 10GBit/s NAT/PAT device 110. In a preferred implementation, a successful correlation device has been implemented using an HP DL380 G7 server with 10 GB of RAM, two 150GB Operating System disks and one Packet Capture Card with 2 x 10Gbit interfaces.

Whereas the present invention has been described in relation to the correlation of sessions across NAT/PAT devices, there are other situations in which certain fields in data packets may be altered for a number of reasons, including anonymisation, as would be apparent to a person of ordinary skill in the relevant art.. The correlation device 115 of the present invention may be used in a passive solution to correlate data packets captured at different points within a communications path in order to detect such alterations and to correlate data before and after such alterations have been made. For example, the correlation device 115 may be deployed to capture data packets either side of an anonymising proxy in order to reverse the anonymisation being performed by the proxy.

CLAIMS

1. An apparatus for mapping data packets undergoing changes to their addressing information between a first point and a second point in a network, the apparatus comprising:

first data capture means for capturing data packets at a first tap point in a communications path, prior to address translation;

second data capture means for capturing data packets at a second tap point in a communications path, following address translation; and

passive correlating means for detecting mappings between data packets captured by the second data capture means and data packets captured by the first data capture means and for outputting a detected mapping between addressing information before and after translation.

2. The apparatus according to claim 1, wherein the first and second data capture means comprise means for associating data packets within a same identified data stream and for organising associated data packets into processing queues, establishing a different processing queue for each identified data stream.

3. The apparatus according to claim 2, wherein the correlating means are arranged to read packets from each processing stream in a parallel processing arrangement.

4. The apparatus according to claim 2 or claim 3, wherein the first and second data capture means are arranged to associate data packets in a given data stream by determining a hash value for a predetermined combination of data fields in each data packet and maintaining a hash table mapping each distinct determined hash value to those packets having the same determined hash value.

5. The apparatus according to any one of the preceding claims, wherein the data streams comprise TCP or UDP sessions over an IP network.

6. An apparatus, substantially as described herein with reference to and as shown in the accompanying drawings.

Fig. 1

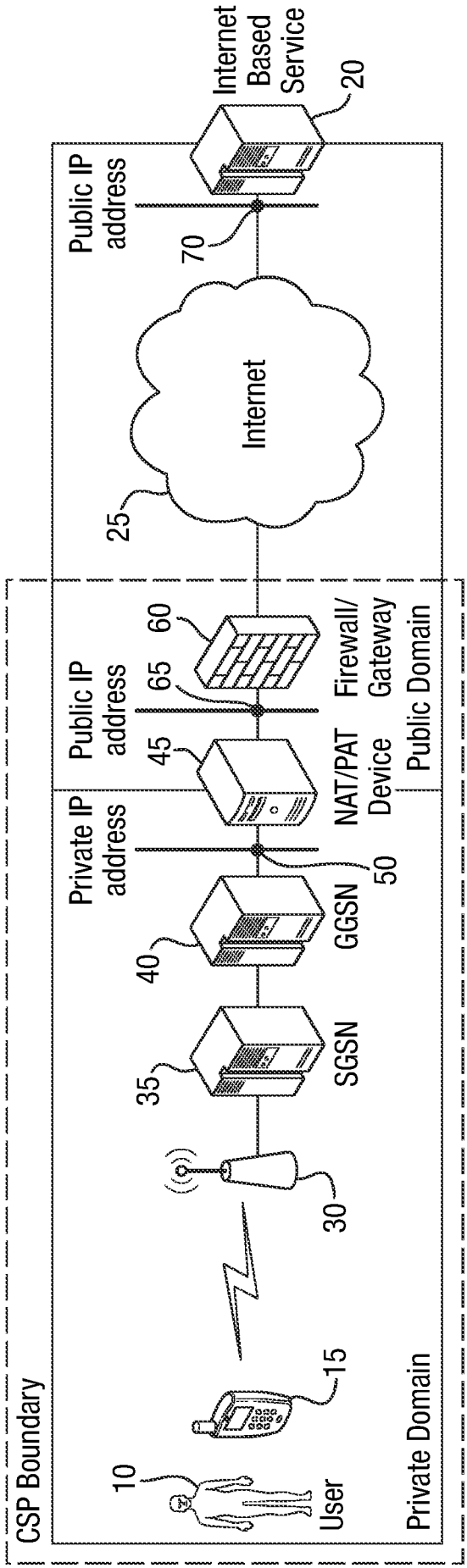

CSP Boundary
User
10
15
30
35
SGSN
40
GGSN
50
Private IP address
45
NAT/PAT Device
Public IP address
65
60
Firewall/ Gateway
Public Domain
Private Domain
25
Internet
70
Public IP address
Internet Based Service
20

Fig. 2

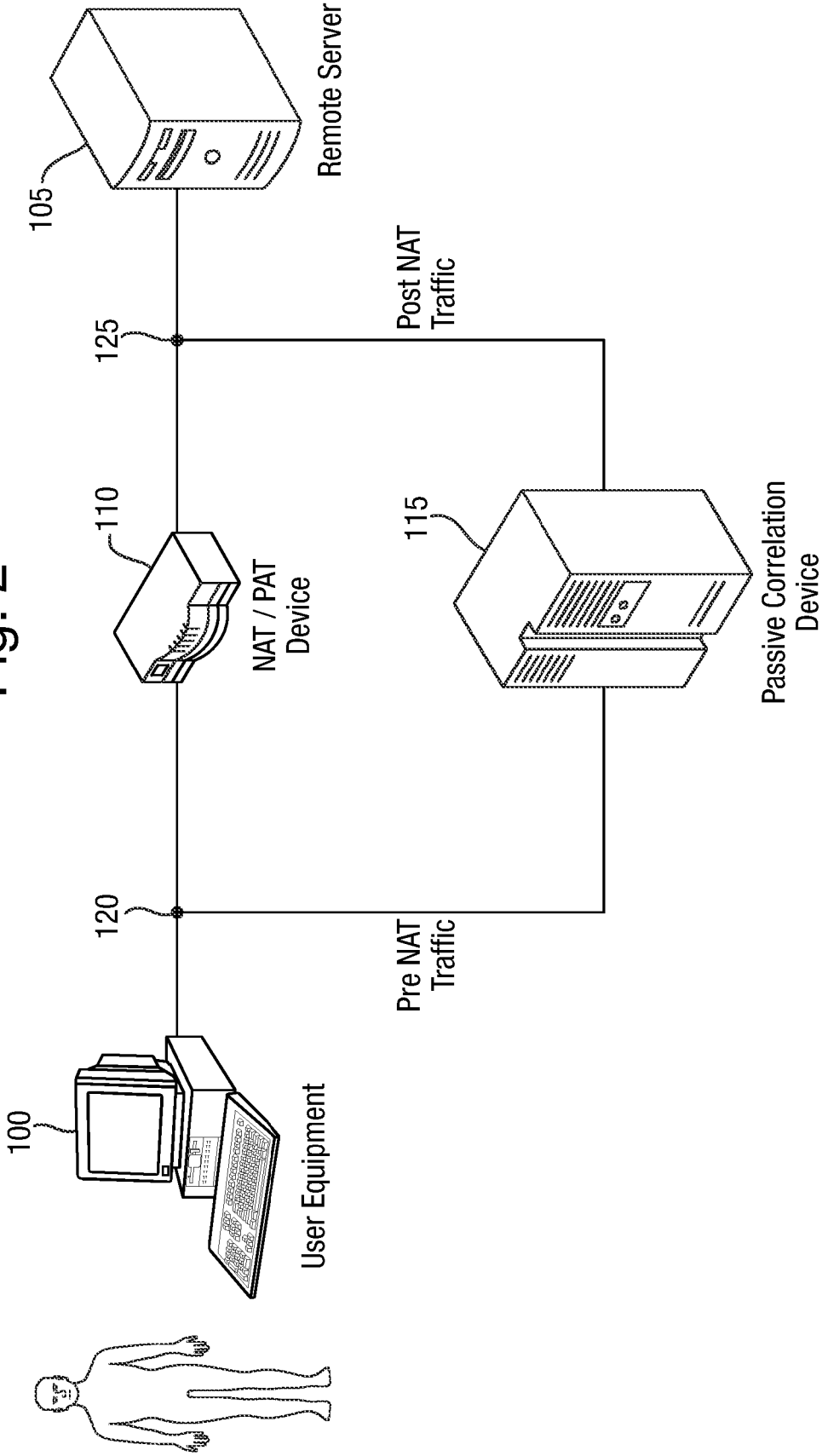

Remote Server
105
Post NAT Traffic
125
110
NAT / PAT Device
115
Passive Correlation Device
120
Pre NAT Traffic
100
User Equipment

Fig. 3

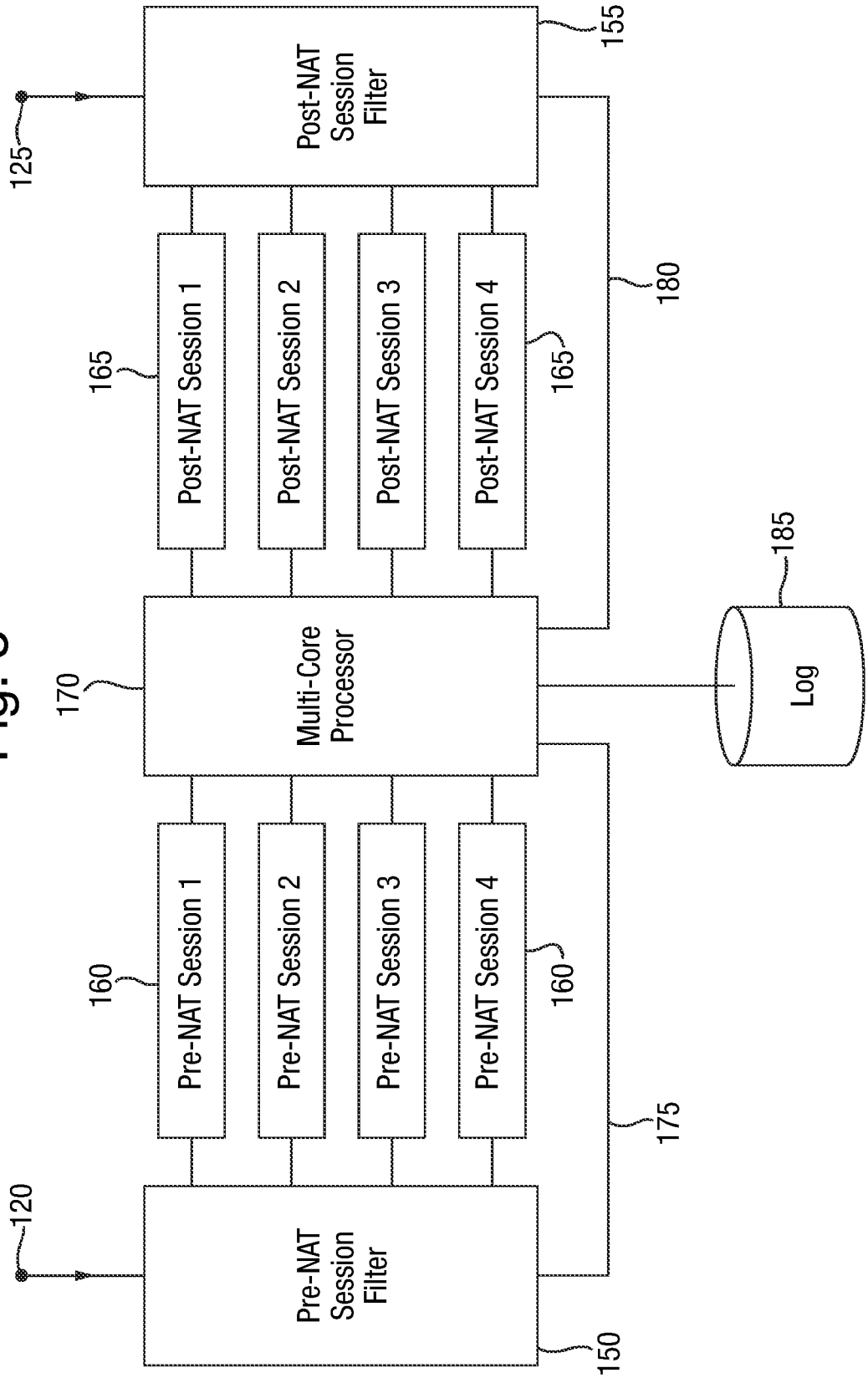

125
Post-NAT Session Filter
155
165
Post-NAT Session 1
Post-NAT Session 2
Post-NAT Session 3
Post-NAT Session 4
165
180
185
170
Multi-Core Processor
Log
160
Pre-NAT Session 1
Pre-NAT Session 2
Pre-NAT Session 3
Pre-NAT Session 4
160
175
120
Pre-NAT Session Filter
150

Fig. 4

Match Type	Date	Private Source IP Address	Private Source Port	Public Source IP	Public Source Port	Destination IP	Destination Port	IP Protocol
Match	26/06/2012 16:00:13.55675315	10.0.114.231	2720	130.8.23.107	368	5.221.178.245	80	TCP
Match	26/06/2012 16:00:14:21690385	10.0.34.11	4480	130.8.23.138	50404	212.209.18.193	80	TCP
Match	26/06/2012 16:00:19.10860559	10.0.169.219	3731	130.8.23.31	62313	65.238.155.87	80	TCP
Match	26/06/2012 16:00:28.28067842	10.0.194.94	4796	130.8.23.156	61253	123.92.193.198	443	TCP
Match	26/06/2012 16:00:30.31997897	10.0.23.164	1400	130.8.23.154	21400	61.131.110.20	80	TCP
Match	26/06/2012 16:00:37.10058305	10.0.223.122	2611	130.8.23.143	39459	198.107.145.34	80	TCP
Match	26/06/2012 16:00:38.64921215	10.0.220.186	198	130.8.23.62	45794	63.155.242.21	80	TCP
Match	26/06/2012 16:00:48.24493048	10.0.235.123	2455	130.8.23.111	45677	133.88.227.114	22	TCP
Match	26/06/2012 16:00:56.14362394	10.0.70.27	3254	130.8.23.88	20180	70.191.84.239	80	TCP
Non-Unique	26/06/2012 16:01:38.39723540	10.0.37.33	44142	130.8.23.99	55872	155.90.217.169	80	TCP
Non-Unique	26/06/2012 16:01:38.39723540	10.0.37.208	417	130.8.23.99	1232	155.90.217.169	80	TCP
Match	26/06/2012 16:01:39.83463934	10.0.37.33	44142	130.8.23.99	55872	155.90.217.169	80	TCP
Match	26/06/2012 16:02:11.40205411	10.0.208.37	2659	130.8.23.122	11325	62.193.75.192	80	TCP
Match	26/06/2012 16:02:16.01704805	10.0.31.63	4341	130.8.23.245	61352	71.79.113.121	25	TCP
Match	26/06/2012 16:02:25.08806653	10.0.85.130	2130	130.8.23.48	22256	126.134.128.118	80	TCP
Match	26/06/2012 16:02:26.06962204	10.0.164.94	2198	130.8.23.75	46757	233.68.220.83	80	TCP

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2013/051652

A. CLASSIFICATION OF SUBJECT MATTER
INV. H04L12/26
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 practicable, search terms used)
EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	EP 2 482 522 A1 (ROKE MANOR RESEARCH [GB]) 1 August 2012 (2012-08-01) abstract paragraphs [0011], [0012], [0016], [0017], [0021] - [0026]; claims 1,5,6 -----	1-6
X	US 2011/145391 A1 (IVERSHEN ALEKSEY G [US]) 16 June 2011 (2011-06-16) abstract paragraphs [0006], [0011], [0024] - [0031]; claims 1,4,7; figure 1 ----- -/--	1-6

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

[X] 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 application or patent 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
2 September 2013	09/09/2013

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Goller, Wolfgang

1

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2013/051652

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CHAPPELL LAURA: "Chapter 1:The world of network analysis", INTERNET CITATION, 29 April 2010 (2010-04-29), pages 1-23, XP002674675, ISBN: 978-1-893939-99-8 Retrieved from the Internet: URL:http://cdn.ttgtmedia.com/searchNetworking/downloads/chapter1_wiresharkbook.pdf [retrieved on 2012-04-23] page 13, line 9 - page 13, line 16	1,6
X	YINJIE CHEN ET AL: "Identifying mobiles hiding behind wireless routers", INFOCOM, 2011 PROCEEDINGS IEEE, IEEE, 10 April 2011 (2011-04-10), pages 2651-2659, XP031953481, DOI: 10.1109/INFCOM.2011.5935093 ISBN: 978-1-4244-9919-9 paragraph [0IIB]	1,6
A	US 2003/223367 A1 (SHAY A DAVID [US] ET AL) 4 December 2003 (2003-12-04) abstract paragraphs [0004], [0007], [0008], [0018]	1-6

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/GB2013/051652

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 2482522 A1	01-08-2012	EP 2482522 A1	01-08-2012
		GB 2487818 A	08-08-2012
		US 2012218999 A1	30-08-2012
US 2011145391 A1	16-06-2011	NONE	
US 2003223367 A1	04-12-2003	AU 2003230764 A1	13-10-2003
		US 2003223367 A1	04-12-2003
		WO 03084137 A2	09-10-2003