Internet Engineering Task Force
INTERNET DRAFT

Arifumi Matsumoto
Masahiro Kozuka
Kenji Fujikawa
Yasuo Okabe
Kyoto University

7 Oct 2003

TCP Multi-Home Options

<draft-arifumi-tcp-mh-00.txt>

Status of this Memo

Abstract

In the existing TCP, only one local and one remote address is used through a TCP session, even when a client or a server is located under multi-homed site and has multiple IP addresses. When a network outage occurs and the access-line associated with the local and remote addresses is down, the TCP session itself gets lost even if another access-line is alive. TCP MH option makes it possible to handle multiple local and remote address pairs in one TCP session and to survive network outages by finding out an alternative network path. Our path transition mechanism is simple, fast, lightweight and as secure as existing TCP.

1. Introduction

Multihoming nodes that connected to the global network through multiple up-stream access-lines are expected to have multiple addresses given by each ISP. The existing TCP, however, is not designed to manipulate multiple addresses in one TCP session. When a network outage occurs and the access-line associated with the local and remote addresses is down, the TCP session itself gets lost.

These new TCP options specified in this document enable a host to get benefit from multi-home in a end-to-end multi-homing[E2E] manner.
By introducing these simple options, TCP becomes much more reliable and powerful without loss of security and without dependency on IPsec.
In this model, both end nodes exchange their addresses using these options. TCP manages all possible network ''paths'', which means a quartet of local and remote IP addresses and ports, and switches from one to another rapidly when a path becomes unavailable. A session can even switch from IPv4 address to IPv6 address and vice versa.

These processes resemble SCTP's[SCTP] multi-homing method.
And in fact, this paper is based on and resembles SCTP's new multi-homing method[ADDIP].
In this paper we try to prove that TCP can be improved and can support multi-homing relatively easily.
This kind of multi-home solution can be rapidly deployed and we believe our solution is of great advantage.

2. MH-Permitted Option

This two-byte option may be sent in a SYN by extended TCP that can recognize (and presumably process) the MH options once the connection is opened. It MUST NOT be sent on non-SYN segments.

```
  MH-Permitted Option:
  Kind: 22
  0                   1
  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 | Kind = 0x16   | Length = 2    |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

3. Address Configuration Options (MH-Add/Delete)

The MH-Add and MH-Delete options are to be used to convey local address information from the sender to the receiver over an established TCP connection. These options MUST be acknowledged by MH-Ack or MH-Non-Ack as described in the next section.

MH-Serial : 16 bits (unsigned integer)

This value represents a Serial Number for MH-Add and MH-Delete options. The valid range of Serial Number is from 0 to 65535 (2**16 -1). Serial Number wrap back to 0 after reaching 65535.

```
 MH-Add-IPv4 Option:
 Kind: 23
 Length: 8
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Kind = 0x17   | Length = 8    |           MH-Serial           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       IPv4 Address                            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

 MH-Delete-IPv4 Option:
 Kind: 24
 Length: 8
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Kind = 0x18   | Length = 8    |           MH-Serial           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       IPv4 Address                            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

 MH-Add-IPv6 Option:
 Kind: 25
 Length: 20
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Kind = 0x19   | Length = 20   |           MH-Serial           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
|                       IPv6 Address                            |
|                                                               |
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

```
     MH-Delete-IPv6 Option:
     Kind: 26
     Length: 20
     0                   1                   2                   3
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    | Kind = 0x1a   | Length = 20   |           MH-Serial           |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                                                               |
    |                       IPv6 Address                            |
    |                                                               |
    |                                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

4. Address Configuration Acknowledgment Options(MH-Ack/Non-Ack)

These options are used by the receiver of MH-Add or MH-Delete Option to acknowledge or reject the address presented by the remote peer.

MH-Serial : 16 bits (unsigned integer)

This value represents a Serial Number for the received MH-Add or MH-Delete option that is acknowledged or rejected. This value is copied from the received MH-Add or MH-Delete option.

```
     MH-Ack Option:
     Kind: 27
     Length: 4
     0                   1                   2                   3
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    | Kind = 0x1b   | Length = 4    |           MH-Serial           |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

     MH-Non-Ack Option:
     Kind: 28
     Length: 4
     0                   1                   2                   3
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    | Kind = 0x1c   | Length = 4    |           MH-Serial           |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

5. Procedure

This section will lay out specific procedures for generating and processing of these options defined in Section from 2 to 4.

5.1 MH-Permitted Option Procedures

When an endpoint is able to handle MH options it should do the following:

1) Create an MH-Permitted option as defined in Section 2 and include it in the sending SYN packet.

2) When the incoming SYN packet doesn't contain MH-Permitted option, MUST NOT include MH-Permitted option in the sending SYN packet.

3) MUST NOT include MH-Permitted option other than SYN packet.

4) The local and remote addresses used to successfully exchange MH-Permitted option should be shortly registered as a valid path and thereafter the local address should not be notified to the peer. Both endpoints should label the path as the primary one and send packets through the path as far as path switching is not activated (see section 6 for details).

5.2 Address Configuration Option Procedures

When an endpoint successfully exchanges MH-Permitted option in the connection establishing state and the connection is established, it should do the following:

1) Lookup local addresses and create an MH-Add-IPv4 or MH-Add-IPv6 option defined in Section 3. Here SHOULD NOT include IPv6 link-local address in this option.

2) A serial number should be assigned to the option. It should be a monotonically increasing number. It SHOULD be initialized at the start of the connection to the value ISS modulo 0xffff, stored as ''local serial number'' and every time a new MH-Add or MH-Delete option is created it is incremented by one after assigning the serial number to the newly created option.

3) If no MH-Add nor MH-Delete option is outstanding (un-acknowledged) with the remote peer, prepare the option to be piggybacked within the next sending packet. If there is no data and outgoing packet, MH option should not be sent in the no data packet. Two or more

address configuration options SHOULD NOT be sent in one packet.

4) In the case of MH-Add option, the sender MUST NOT register or use the local address or associate the address with the existing TCP session until the acknowledgment(MH-Ack) to the option is received. The same is true for MH-Delete option, and the sender MUST NOT un-register the local address before the arrival acknowledgement option.

5) If the RTO timer expires, the endpoint should retransmit the same outstanding MH option last sent.

5.2.1 Congestion Control of Address Configuration Options

One and only one MH-Add or MH-Delete option MAY be in transit and unacknowledged at any time. If a sender, after sending an MH option, decides it needs to transfer another MH option, it MUST wait until the appropriate MH-Ack/Non-Ack option returns before sending a subsequent MH option. Note this restriction binds each side, so at any time two MH option may be in-transit on any given connection (one sent from each endpoint).

5.3 Upon reception of an Address Configuration Option

When an endpoint receives an MH option from the remote peer,

1) Compare the value of the serial number to the value the endpoint stored as the ''peer serial number''. This value MUST be initialized to the ISS modulo 0xffff at the establishment state.

2) If the value found in the serial number is equal to the (peer serial number + 1), the endpoint MUST process the MH option and perform the following action.

2-1) For MH-Add options, register all the paths those can be created using the address included in the option. For MH-Delete options un-register all the paths using the specified address.

2-2) However, the endpoint MUST NOT un-register paths too soon so that it can receive a packet through deleting paths for a certain amount of time, at least for 2 or 3 times RTT, for security reasons. (This is discussed a bit more in security considerations section).

2-3) Soon after the reception and process of the option the endpoint SHOULD build a response MH-Ack message with the serial

number contained in the received MH option so that the message can be piggybacked in the next outgoing packet. If the address family specified in a option is not supported by the endpoint or the address in a MH-Delete option has not notified before, the endpoint SHOULD build a MH-Nack option.

2-4) Update the peer serial number to the value found in the serial number field.

3) If the value found in the serial number is equal to the value stored in the peer serial number, the endpoint should build the same response option last sent and should not update the peer serial number.

4) If the value found in the serial number is not equal to these values, send RST segment to the remote peer and MUST ABORT the connection shortly for security reasons.

5.4 Upon reception of acknowledgement Option

When an endpoint receives an MH-Ack or MH-Nack option from the peer,

1) Compare the value of the serial number to that of the stored outstanding address configuration option last sent.

2) If there is a stored outstanding address configuration option and the serial number is equal to that of the stored option, the endpoint perform the following.

2-1) If the incoming option is MH-Ack for MH-Add option last sent, register all the paths that can be created using the local address specified in the outstanding option. In the same way, if the incoming option is MH-Ack for MH-Delete option last sent un-register all the paths using that address.

2-2) If the incoming option is MH-Nack, do nothing and should not repeat sending the same option at least until the peer processes some other address configuration options correctly.

2-3) Update the ''local serial number'' to the value found in the serial number field + 1.

3) If there is no outstanding data and the serial number is equal to the (local serial number - 1), just discard the option.

4) In other case, send RST segment to the remote peer and MUST ABORT the connection shortly for security reasons.

6 Path Transition (Setting of the primary path)

6.1 In connection establishment state

A TCP client endpoint should do the following to establish a session if the socket is not bound to a certain address other than INADDR_ANY.

1) In the case of active open, TCP layer is told to connect to one remote address specified by the upper layer, which is usually the user land. Here, one local address, corresponding to the given remote address, is chosen by source address selection mechanism and a SYN packet are sent through that path.

2) If the SYN packet is retransmitted several times and not acknowledged, the endpoint should switch the source address to another one if available. As stated above, IPv6 link-local address should not be used here. IPv6 link-local address can only be selected in the 1) state.

3) The endpoint should not stop connecting until all the existing paths failed.

4) When a connection is established label the path used for connection establishment as a primary one and afterward should send packets through that path. The endpoint should not notify the address that is used for successful connection establishment to the peer.

6.2 After the connection is established

In order to make the both endpoint use the same path and to survive a network outage, the endpoint do the followings:

1) When a sending data retransmits several times,

1-1) The endpoint should choose another registered and available path. and label the path as ''temporary path''. Use the temporary path if exists to send a packet.

1-2) Switch the temporary path to another one if data retransmission repeats several times. Continue changing the temporary path as far as an acknowledgement packet doesn't arrive and session is not aborted by a timer.

2) Or when a packet is received through not a primary path, label the path as a temporary path and send a packet through the temporary

path. Use the temporary path if exists to send a packet.

3) If a ICMP error, such as host un-reach, is received through not a primary path, the endpoint should not shutdown TCP session itself but just try to use another path. In the same way, when the endpoint received TCP RST segment through an unused path, the endpoint should try to use another one. When a RST segment comes from a used path, through which a packet is successfully received at least one, the endpoint should regarded it as a message from the valid remote peer and should abort TCP session.

4) If a packet arrives at the existing temporary path, label the path as the primary one.

7. Security Considerations

This kind of Add/Delete of IP address option seemingly introduces an additional mechanism to hijack existing TCP connection. But some measures are taken in this specification so as not make TCP more vulnerable than ever.

As for man in the middle attack, existing TCP protocol is known to be vulnerable and this is not improved nor degraded in this specification except that the connection can be taken to another place in the network.

As for an attack by wiretapping host, which is often the case when shared media like wireless LAN is used, an endpoint can not easily be connected to a faked host. This is owing to delayed MH-Delete execution (2-2 in section 5.3) , stringent serial number management and address configuration only through legitimate path.

8. References

[E2E] M. Ohta, ``The Architecture of End to End Multihoming,'' Internet-draft, IETF (Nov 2002), draft-ohta-e2e-multihoming-03.txt.

[RFC2960] R. Stewart, Q. Xie, K. Morneault, C. Sharp, H. Schwarzbauer, T. Taylor, I. Rytina, M. Kalla, L. Zhang, V. Paxson, ``Stream Control Transmission Protocol,`` RFC2960, IETF (Oct 2000).

[ADDIP]
R. Stewart, at el, ``Stream Control Transmission Protocol (SCTP) Dynamic Address Reconfiguration,'' Internet-draft, IETF (Feb 2003), draft-ietf-tsvwg-addip-sctp-07.txt. (Work In Progress)

9. Authors' Addresses

Arifumi Matsumoto
Graduate School of Informatics
Kyoto University
Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501 JAPAN
Tel: +81 75-753-7468
Fax: +81 75-753-7472
Email: arifumi@net.ist.i.kyoto-u.ac.jp

Masahiro Kozuka
Faculty of Law, Kyoto University
Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501 JAPAN
Tel: +81 75-753-7468
Fax: +81 75-753-7472
Email: ma-kun@kozuka.jp

Kenji Fujikawa
Graduate School of Informatics
Kyoto University
Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501 JAPAN
Tel: +81 75-753-5387
Fax: +81 75-753-4961
Email: fujikawa@real-internet.org

Yasuo Okabe
Academic Center for Computing and Media Studies
Kyoto University
Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501 JAPAN
Tel: +81 75-753-7458
Fax: +81 75-751-0482
Email: okabe@i.kyoto-u.ac.jp