

BANANA
Internet Draft
Intended Category: Informational

M. Zhang
Huawei
N. Leymann
C. Heidemann
Deutsche Telekom AG
M. Cullen
Painless Security
March 21, 2016

Expires: September 22, 2016

Flow Control for Bonding Tunnels
draft-zhang-banana-tcp-in-bonding-tunnels-00.txt

Abstract

Tunnel bonding enables Bandwidth Aggregation across multiple Internet access links. From the practice of deployment, flow control is a desirable feature of the bonding tunnel. This document explores the way to equip bonding tunnel with flow control mechanism.

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at
<http://www.ietf.org/lid-abstracts.html>

The list of Internet-Draft Shadow Directories can be accessed at
<http://www.ietf.org/shadow.html>

Copyright and License Notice

Copyright (c) 2016 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents

(<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	2
2. Acronyms and Terminology	3
3. A Network Based Sliding Window	3
4. Bonding De-Activation	4
5. Security Considerations	5
6. IANA Considerations	5
7. References	5
7.1. Normative References	5
7.2. Informative References	5
Author's Addresses	6

1. Introduction

Bandwidth aggregation capabilities across multiple Internet access links can significantly improve end users' Quality of Experience. Lots of service providers who possess multiple access networks are about to deploy or have already deployed the bandwidth aggregation capabilities. There are various enabling techniques arising to address bandwidth aggregation issues [Problem]. Among them, tunneling techniques are promising since the network-based requirement can be fulfilled easily.

To achieve bandwidth aggregation, two tunnels are to be bonded together to form a single logical tunnel in between the access and aggregation equipments. When the bandwidth of the primary tunnel fills, the secondary tunnel can be used to accommodate the excessive load, while end users are supposed not to be aware of the shifting.

Flow control is a desirable feature for bandwidth aggregation. In this document, TCP-like sliding window mechanism is integrated into the tunnel to handle the packet delay, loss and network congestion. However, considering the overhead that may be introduced into the tunnel, the transportation utilities used here will not cover a full set of TCP functions.

In the practical deployment, the quality of the two bonded tunnels

may become un-comparable. The bonding mode will be deactivated and user's traffic will be carried using the primary tunnel only. From calculating of the Adaptive Time-Out in the flow control, operators can judge whether the bonding mode should be deactivated.

2. Acronyms and Terminology

Hybrid Access: The bonding of multiple access connections based on heterogeneous technologies (e.g., DSL and LTE) that are provided by the same carrier.

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC 2119](#) [RFC2119].

3. A Network Based Sliding Window

This document equips bonded tunnels with the well-known sliding window mechanism, which used to be a well-know TCP utility, to handle packet loss, packet delay and congestions introduced by the bonding tunnel. Compared to the sliding windows that are implemented on end-systems, this document specifies a network based sliding window mechanism.

In order to realize the sliding window mechanism, sequence number and acknowledgement number are required. Note that the sequence number specified in this document is used to number packets sent on each individual tunnel. It is different from the sequence number that is used for packet reordering for the entire bonding tunnel [GREbond]. However, the two sequence numbers maybe concatenated to use the 4-octet Sequence Number field of the GRE header [GREbond]. The sequence number for the entire bonding tunnel uses the higher order two octets while the sequence number of the individual tunnel uses the lower order two octets.

Based on this sequence number and acknowledgement number, retransmission could be realized. However, considering the heavy overhead that might be caused, HCPE and HAG will not perform retransmission. Retransmission might be realized in accompanied TCP proxies or accelerators, but it is out the scope of this document. The individual tunnels will merely adopt sliding window mechanism to moderate the rate at which the data is being transmitted. To support such a mechanism, the HCPE and HAG devices are required to implement a sending window.

Sequence numbers and acknowledgement numbers are maintained per individual tunnels. Sequence numbers have to be piggybacked on data packets while acknowledgement numbers can be sent separately from

data packets. Each bonding tunnel is treated as a constant long-live session. Sliding window protocol of [RFC2637] will be used as the basis. The following changes to RFC 2637 are required.

(Section 4.2.4. Window Overflow) When a receiver's window overflows with too many incoming packets, the receiver immediately starts to digest the packets in the window and put them into the reordering buffer, even if there are missing packets. This process continues until excess packets are accommodated. When the secondary transmission window fills, no more traffic should be distributed to the secondary tunnel. Packets will enter the primary transmission window.

(Section 4.2.5. Multi-packet Acknowledgment) Each packet is acknowledged as it enters the reordering buffer of the bonding tunnel. When the receiving window overflows, packets are digested even if there are missing packets with lower sequence number. At that point, those digested packets will be acknowledged as well.

(Section 4.3. Out-of-sequence Packets) Out of sequence packets are not dropped. Instead, they will be digested and enter the reordering buffer of the bonding tunnel. For these out-of-sequence packets, no acknowledgement will be sent.

(Section 4.4.1. Calculating Adaptive Acknowledgment Time-Out) When a time-out occurs for the secondary tunnel, the sending device will not stop distributing packets onto the secondary tunnel as long as a good throughput is promised (i.e., the transmission window does not fill), unless the difference of the ATOs of the two tunnels exceeds the configured MaxDiffTimeOut. In the case that the difference of ATOs exceeds MaxDiffTimeOut or the secondary transmission window fills, no packets will be distributed to the secondary tunnel. Packets remained in the secondary transmission window will be moved to the primary transmission window.

4. Bonding De-Activation

In the practical deployment, the secondary connection might suffer from large latency, jitter and high packet loss rate. In the bonding tunnel, packets are distributed onto both primary and secondary tunnels at one end and then recombined again in a buffer at the receiver. If the secondary GRE tunnel suffers, the entire bonding tunnel will suffer as well due to the recombination function. For example, assume packet 1 is distributed onto the secondary GRE tunnel while packet 2 is distributed onto the primary GRE tunnel. If packet 1 is delayed by 100 ms, even packet 2 arrives at the recombination buffer at the first ms, it has to remain in the buffer awaiting for 99 ms for packet 1.

Deployment experiences show that the TCP throughput of the bonding tunnel might be greatly reduced due to the downgrade or disruption of the secondary tunnel. Hence, monitoring the performance of the secondary tunnel is required. Based on the monitoring, when the performance of the two links are comparable operators will remain the bonding mode open. Otherwise, when the monitored performance parameters do not meet their predefined requirements, the bonding mode will be de-activated so that the HCPE and HAG only use the primary connection. As specified in [Section 3](#), the calculation of the Adaptive Acknowledgment Time-Out provides such kind of monitoring.

5. Security Considerations

Unless the payload data and control messages carried in the tunnels are cryptographically protected, they can be captured and read or modified. Attackers may make up bogus sequence numbers and acknowledge numbers to cause replay or deny of service attacks.

6. IANA Considerations

No new registration is required to be allocated from IANA. RFC editor, please remove this section before its publication.

7. References

7.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), DOI 10.17487/RFC2119, March 1997, <<http://www.rfc-editor.org/info/rfc2119>>.
- [RFC2637] Hamzeh, K., Pall, G., Verthein, W., Taarud, J., Little, W., and G. Zorn, "Point-to-Point Tunneling Protocol (PPTP)", [RFC 2637](#), DOI 10.17487/RFC2637, July 1999, <<http://www.rfc-editor.org/info/rfc2637>>.
- [GREbond] Leymann, N., Heidemann, C. , et al, "GRE Tunnel Bonding", [draft-zhang-gre-tunnel-bonding](#), Work in Progress.

7.2. Informative References

- [Problem] Cullen, M., et al, "Problem Statement: Bandwidth Aggregation for Internet Access", Work in Progress, [draft-zhang-banana-problem-statement](#), Work in Progress.

Author's Addresses

Mingui Zhang
Huawei Technologies
No. 156 Beiqing Rd., Haidian District
Beijing 100095
China

Email: zhangmingui@huawei.com

Nicolai Leymann
Deutsche Telekom AG
Winterfeldtstrasse 21-27
Berlin 10781
Germany

Phone: +49-170-2275345
EMail: n.leymann@telekom.de

Cornelius Heidemann
Deutsche Telekom AG
Heinrich-Hertz-Strasse 3-7
Darmstadt 64295
Germany

Phone: +4961515812721
Email: heidemannc@telekom.de

Margaret Cullen
Painless Security
14 Summer St. Suite 202
Malden, MA 02148
United States

Email: margaret@painless-security.com