Workgroup: IDR
Internet-Draft:
draft-spaghetti-idr-bgp-sendholdtimer-04
Updates: 4271 (if approved)
Published: 14 April 2022
Intended Status: Standards Track
Expires: 16 October 2022
Authors: J. Snijders B. Cartwright-Cox
Fastly

Border Gateway Protocol 4 (BGP-4) Send Hold Timer

Abstract

This document defines the SendHoldTimer session attribute for the Border Gateway Protocol (BGP) Finite State Machine (FSM). Implementation of a SendHoldTimer should help overcome situations where BGP sessions are not terminated after it has become detectable for the local system that the remote system is not processing BGP messages. For robustness, this document specifies that the local system should close BGP connections and not solely rely on the remote system for session tear down when BGP timers have expired. This document updates RFC4271.

Requirements Language

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].

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at https://datatracker.ietf.org/drafts/current/.

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."

This Internet-Draft will expire on 16 October 2022.

Table of Contents

1. Introduction

This document defines the SendHoldTimer session attribute for the Border Gateway Protocol (BGP) [RFC4271] Finite State Machine (FSM) defined in section 8.

Failure to terminate a 'stuck' BGP session can result in Denial Of Service, the subsequent failure to generate and deliver BGP WITHDRAW messages to other BGP peers of the local system is detrimental to all participants of the inter-domain routing system. This phenomena is theorised to have contributed to IP traffic backholing events in global Internet routing system [bgpzombies].

This specification intends to improve this situation by requiring sessions to be terminated if the local system has detected that the remote system cannot possibly have received any BGP messages for the duration of the SendHoldTimer. Through codification of the

aforementioned requirement, operators will benefit from consistent behavior across different BGP implementations.

BGP speakers following this specification do not exclusively rely on remote systems robustly closing connections, but will also locally close connections.

2. Example of a problematic scenario - RFC EDITOR: REMOVE BEFORE PUBLICATION

A malfunctioning or overwhelmed peer may cause data on the BGP socket in the local system to back up, and the current RFC specification will not cause the session to be torn down. For example, as BGP runs over TCP [RFC0793] it is possible for hosts in the ESTABLISHED state to encounter a BGP peer that is advertising a TCP Receive Window (RCV.WND) of size zero and thus preventing the local system from sending KEEPALIVE, CEASE, WITHDRAW, UPDATE, or other critical messages across the wire. At the moment of writing, most BGP implementations appear unable to handle this situation in a robust fashion.

Generally BGP implementation have no visibility into lower-layer subsystems such as TCP or the peer's current Receive Window. Therefor this document banks on BGP implementations being able to detect an inability to push more data to the remote peer, at which point the SendHoldTimer starts.

3. Specification of the Send Hold Timer

BGP speakers are implemented following a conceptual model "BGP Finite State Machine" (FSM), which is outlined in section 8 of [RFC4271]. This specification updates the BGP FSM as following:

3.1. Session Attributes

The following mandatory session attributes are added to paragraph 6 of Section 8, before "The state session attribute indicates the current state of the BGP FSM":

9) SendHoldTimer

10) SendHoldTime (an initial value of 4 minutes is recommended)

3.2. SendHoldTimer_Expires Event Definition

Section 8.1.3 [RFC4271] is extended as following:

Event XX: SendHoldTimer_Expires
Definition : An event generated when the SendHoldTimer expires.
Status: Mandatory

If the SendHoldTimer_Expires (Event XX), the local system:

- logs a message with the BGP Error Notification Code "Send Hold Timer Expired",
- releases all BGP resources,
- sets the ConnectRetryTimer to zero,
- drops the TCP connection,
- increments the ConnectRetryCounter,
- (optionally) performs peer oscillation damping if the DampPeerOscillations attribute is set to TRUE, and
- changes its state to Idle.

If the DelayOpenTimer_Expires event (Event 12) occurs in the Connect state, the local system:

- sends an OPEN message to its peer,
- sets the HoldTimer to a large value, and
- sets the SendHoldTimer to a large value, and
- changes its state to OpenSent.

If the DelayOpen attribute is set to FALSE, the local system:

- stops the ConnectRetryTimer (if running) and sets the ConnectRetryTimer to zero,
- completes BGP initialization
- sends an OPEN message to its peer,
- sets the HoldTimer to a large value, and
- sets the SendHoldTimer to a large value, and
- changes its state to OpenSent.

A HoldTimer value of 4 minutes is suggested.

A SendHoldTimer value of 4 minutes is suggested.

4. Send Hold Timer Expired Error Handling

If a system does not send and receive successive KEEPALIVE, UPDATE, and/or NOTIFICATION messages within the period specified in the Send Hold Time, then the BGP connection is closed and a log message is emitted.

5. Implementation status - RFC EDITOR: REMOVE BEFORE PUBLICATION

This section records the status of known implementations of the protocol defined by this specification at the time of posting of this Internet-Draft, and is based on a proposal described in RFC 7942. The description of implementations in this section is intended to assist the IETF in its decision processes in progressing drafts to RFCs. Please note that the listing of any individual implementation here does not imply endorsement by the IETF. Furthermore, no effort has been spent to verify the information presented here that was supplied by IETF contributors. This is not intended as, and must not be construed to be, a catalog of available implementations or their features. Readers are advised to note that other implementations may exist.

According to RFC 7942, "this will allow reviewers and working groups to assign due consideration to documents that have the benefit of running code, which may serve as evidence of valuable experimentation and feedback that have made the implemented protocols more mature. It is up to the individual working groups to use this information as they see fit".

*OpenBGPD [openbgpd]

6. Acknowledgements

The authors would like to thank William McCall and Theo de Raadt for their helpful review of this document.

7. Security Considerations

This specification addresses the vulnerability of a BGP speaker to a potential attack whereby a BGP peer can pretend to be unable to process BGP messages and in doing so create a scenario where the local system is poisoned with stale routing information.

There are three detrimental aspects to the problem of not robustly handling 'stuck' peers:

*Failure to send BGP messages to a peer implies the peer is operating based on stale routing information.

* Failure to disconnect from a 'stuck' peer hinders the local system's ability to construct a non-stale local Routing Information Base (RIB).

* Failure to disconnect from a 'stuck' peer hinders the local system's ability to inform other BGP peers with current network reachability information.

In other respects, this specification does not change BGP's security characteristics.

8. IANA Considerations

This document requests IANA to assign a value named "Send Hold Timer Expired" in the "BGP Error (Notification) Codes" sub-registry under the "Border Gateway Protocol (BGP) Parameters" registry.

9. References

9.1. Normative References

[RFC0793] Postel, J., "Transmission Control Protocol", STD 7, RFC 793, DOI 10.17487/RFC0793, September 1981, <https://www.rfc-editor.org/info/rfc793>.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, DOI 10.17487/RFC4271, January 2006, <https://www.rfc-editor.org/info/rfc4271>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

9.2. Informative References

[bgpzombies] Fontugne, R., "BGP Zombies", April 2019, <https://labs.ripe.net/author/romain_fontugne/bgp-zombies/>.

[openbgpd] Jeker, C., "bgpd send side hold timer", December 2020, <https://marc.info/?l=openbsd-tech&m=160820754925261&w=2>.

Authors' Addresses

Job Snijders
Fastly
Amsterdam
Netherlands

Email: job@fastly.com

Ben Cartwright-Cox
London
United Kingdom

Email: ben@benjojo.co.uk