MQTT Handling of Disallowed Unicode Code Points Version 1.0

Committee Note 01

19 April 2018

Specification URIs

This version:

http://docs.oasis-open.org/mqtt/disallowed-chars/v1.0/cn01/disallowed-chars-v1.0-cn01.pdf (Authoritative)

http://docs.oasis-open.org/mqtt/disallowed-chars/v1.0/cn01/disallowed-chars-v1.0-cn01.html

http://docs.oasis-open.org/mqtt/disallowed-chars/v1.0/cn01/disallowed-chars-v1.0-cn01.docx

Previous version:

N/A

Latest version:

http://docs.oasis-open.org/mqtt/disallowed-chars/v1.0/disallowed-chars-v1.0.pdf (Authoritative)

http://docs.oasis-open.org/mqtt/disallowed-chars/v1.0/disallowed-chars-v1.0.html

http://docs.oasis-open.org/mqtt/disallowed-chars/v1.0/disallowed-chars-v1.0.docx

Technical Committee:

OASIS Message Queuing Telemetry Transport (MQTT) TC

Chairs:

Brian Raymor (brian.raymor@microsoft.com), Microsoft

Richard J Coppen (coppen@uk.ibm.com), IBM

Editors:

Andrew Banks (andrew_banks@uk.ibm.com), IBM

Ed Briggs (edbriggs@microsoft.com), Microsoft

Ken Borgendale (kwb@us.ibm.com), IBM

Rahul Gupta (rahul.gupta@us.ibm.com), IBM

Related work:

This document is related to:

- *MQTT Version 3.1.1 Plus Errata 01*. Edited by Andrew Banks, and Rahul Gupta. 10 December 2015. OASIS Standard Incorporating Approved Errata 01. http://docs.oasis-open.org/mqtt/mqtt/v3.1.1/errata01/os/mqtt-v3.1.1-errata01-os-complete.html.
- ISO/IEC 20922:2016 Preview Information technology -- Message Queuing Telemetry Transport (MQTT) v3.1.1 Edited by Andrew Banks, and Rahul Gupta. Latest version: https://www.iso.org/standard/69466.html
- *MQTT Version 5.0*. Edited by Andrew Banks, Ed Briggs, Ken Borgendale, and Rahul Gupta. Latest version: http://docs.oasis-open.org/mqtt/mqtt/v5.0/mqtt-v5.0.html.

Abstract:

This Committee Note describes identified exposures in the handling of disallowed Unicode code points. Users of MQTT are alerted to the possibility that some combinations of MQTT Clients and Servers might allow properly authorized publishing Clients to cause the disconnection of properly authorized subscribing Clients. We describe how to identify if this risk is present and how to eliminate it.

Status:

This document was last revised or approved by the OASIS Message Queuing Telemetry Transport (MQTT) TC on the above date. The level of approval is also listed above. Check the "Latest version" location noted above for possible later revisions of this document.
Technical Committee (TC) members should send comments on this document to the TC's email list. Others should send comments to the TC's public comment list, after subscribing to it by following the instructions at the "Send A Comment" button on the TC's web page at https://www.oasis-open.org/committees/mqtt/.

Citation format:

When referencing this document the following citation format should be used:

[MQTT-Disallowed-Unicode-v1.0]

MQTT Handling of Disallowed Unicode Code Points Version 1.0. Edited by Andrew Banks, Ed Briggs, Ken Borgendale, and Rahul Gupta. 19 April 2018. OASIS Committee Note 01. http://docs.oasis-open.org/mqtt/disallowed-chars/v1.0/cn01/disallowed-chars-v1.0-cn01.html. Latest version: http://docs.oasis-open.org/mqtt/disallowed-chars/v1.0/disallowed-chars-v1.0.html.

the copyright notice or references to OASIS, except as needed for the purpose of developing any document or deliverable produced by an OASIS Technical Committee (in which case the rules applicable to copyrights, as set forth in the OASIS IPR Policy, must be followed) or as required to translate it into languages other than English.

The limited permissions granted above are perpetual and will not be revoked by OASIS or its successors or assigns.

This document and the information contained herein is provided on an "AS IS" basis and OASIS DISCLAIMS ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY OWNERSHIP RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Table of Contents

1.1 Introduction

The MQTT V3.1.1 specification section 1.5.3 UTF-8 encoded strings and ISO/IEC 20922:2016, describe the set of Unicode Control Codes and Unicode Noncharacters which should not be included in a UTF-8 Encoded String. The specifications do not require a Client or Server implementation to validate that these code points are not used in UTF-8 Encoded Strings, in particular, Topic Names. We refer to these code points as Disallowed Unicode code points in this document.

If the Server does not validate the code points in a UTF-8 encoded string but a subscribing Client does, then a second Client might be able to cause the subscribing Client to disconnect by publishing on a Topic Name that contains a Disallowed Unicode code point. This document recommends some steps that can be taken to prevent this eventuality.

1.2 References (non-normative)

[Unicode] The Unicode Consortium. *The Unicode Standard*,
http://www.unicode.org/versions/latest/

1.3 Considerations for the use of Disallowed Unicode code points

An implementation would normally choose to validate UTF-8 Encoded strings, checking that the Disallowed Unicode code points are not used, so as to avoid implementation difficulties. This includes the use of libraries that are sensitive to these code points, or to protect applications from having to process them.

Validating that these code points are not used removes some security exposures. There are possible security exploits which use control characters in log files to mask entries in the logs or confuse the tools which process log files. The Unicode Noncharacters are commonly used as special markers and allowing them into UTF-8 Encoded Strings could permit such exploits.

1.4 Interactions between Publishers and Subscribers

The publisher of an Application Message normally expects that the Servers will forward the message to subscribers, and that these subscribers are capable of processing the messages.

Here we describe the set of conditions which allow a publishing Client to cause the disconnection of subscribing Clients. Consider a situation where:

- A Client publishes an Application Message using a Topic Name containing one of the Disallowed Unicode code points.
- The publishing Client library allows the Disallowed Unicode code point to be used in a Topic Name rather than rejecting it.
- The publishing Client is authorized to send the publication.
- A subscribing Client is authorized to use a Topic Filter which matches the Topic Name. Note that the Disallowed Unicode code point might occur in a part of the Topic Name matching a wildcard character in the Topic Filter.
- The Server forwards the message to the matching subscriber rather than disconnecting the publisher.

- In this case the subscribing Client might:
  - Disconnect, because it does not allow the use of Disallowed Unicode code points. If the Client reconnects and the message is QoS=1 or QoS=2, the message will be sent again, causing the Client to disconnect again.
  - Accept the Application Message but fail to process it because it contains one of the Disallowed Unicode code points.
  - Successfully process the Application Message.

The potential for Client disconnection might go unnoticed until a publisher uses one of the Disallowed Unicode code points.

1.5 Remedies

If there is a possibility that a Disallowed Unicode code point could be included in a Topic Name delivered to a Client, the solution owner can adopt one of the following suggestions:

1) Change the Server implementation to one that disconnects a publisher which uses a Disallowed Unicode code point in a Topic Name.
2) Restrict the authorization rules for the publisher so that it cannot publish Application Messages using Topic Names which contain Disallowed Unicode code points.
3) Restrict the Topic Filters authorized to subscribers so that a Client cannot use Topic Filters containing Disallowed Unicode code points. If a client is allowed to make a subscription containing a wild card character, ensure that the Server is configured so that publishers cannot make publications where a Disallowed Unicode code point would match the wildcard.
4) Change the Client library used by the subscribers to one that tolerates the use of Disallowed Code points. The client can either process or discard messages with Topic Names that contain Disallowed Unicode code points so long as it continues the protocol.

Appendix A. Acknowledgments

The following individuals have participated in the creation of this specification and are gratefully acknowledged:

Participants:

[Pouyan Sepehrdad, Qualcomm | Non Member]
[Davide Quarta, Qualcomm | Non Member]

Appendix B. Revision History

Revision	Date	Editor	Changes Made
1	13 February 2018	Andrew Banks	Initial draft