Workgroup: Transport Area Working Group
Internet-Draft: draft-ietf-tsvwg-nqb-15
Updates: rfc8325 (if approved)
Published: 11 January 2023
Intended Status: Standards Track
Expires: 15 July 2023
Authors: G. White CableLabs
T. Fossati ARM

A Non-Queue-Building Per-Hop Behavior (NQB PHB) for Differentiated Services

Abstract

This document specifies properties and characteristics of a Non-Queue-Building Per-Hop Behavior (NQB PHB). The purpose of this NQB PHB is to provide a separate queue that enables smooth, low-data-rate, application-limited traffic flows, which would ordinarily share a queue with bursty and capacity-seeking traffic, to avoid the latency, latency variation and loss caused by such traffic. This PHB is implemented without prioritization and can be implemented without rate policing, making it suitable for environments where the use of these features is restricted. The NQB PHB has been developed primarily for use by access network segments, where queuing delays and queuing loss caused by Queue-Building protocols are manifested, but its use is not limited to such segments. In particular, applications to cable broadband links, Wi-Fi links, and mobile network radio and core segments are discussed. This document recommends a specific Differentiated Services Code Point (DSCP) to identify Non-Queue-Building flows.

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 15 July 2023.

Copyright Notice

Table of Contents

1. Introduction

This document defines a Differentiated Services per-hop behavior (PHB) called "Non-Queue-Building Per-Hop Behavior" (NQB PHB), which isolates traffic flows that are relatively low data rate and that do not themselves materially contribute to queueing delay and loss, allowing them to avoid the queuing delays and losses caused by other traffic. Such Non-Queue-Building flows (for example: interactive voice, low-data-rate online gaming, machine-to-machine applications) are application limited flows that are distinguished from the high-data-rate traffic flows that are typically managed by an end-to-end congestion control algorithm.

Most packets carried by broadband access networks are managed by an end-to-end congestion control algorithm, such as Reno, Cubic or BBR. These congestion control algorithms attempt to seek the available capacity of the end-to-end path (which can frequently be the access network link capacity), and in doing so generally overshoot the available capacity, causing a queue to build-up at the bottleneck link. This queue build-up results in queuing delay (variable latency) and possibly packet loss that can affect all the applications that are sharing the bottleneck link. Moreover, many bottleneck links implement a relatively deep buffer (100 ms or more) in order to enable traditional congestion-controlled applications to effectively use the link, which exacerbates the latency and latency variation experienced.

In contrast to traditional congestion-controlled applications, there are a variety of relatively low data rate applications that do not materially contribute to queueing delay and loss but are nonetheless subjected to it by sharing the same bottleneck link in the access network. Many of these applications can be sensitive to latency or latency variation, as well as packet loss, and thus produce a poor quality of experience in such conditions.

Active Queue Management (AQM) mechanisms (such as PIE [RFC8033], DOCSIS-PIE [RFC8034], or CoDel [RFC8289]) can improve the quality of experience for latency sensitive applications, but there are practical limits to the amount of improvement that can be achieved without impacting the throughput of capacity-seeking applications. For example, AQMs generally allow a significant amount of queue depth variation to accommodate the behaviors of congestion control algorithms such as Reno and Cubic. If the AQM attempted to control the queue much more tightly, applications using those algorithms would not perform well. Alternatively, flow queueing systems, such as fq_codel [RFC8290] can be employed to isolate flows from one another, but these are not appropriate for all bottleneck links, due to complexity or other reasons.

The NQB PHB supports differentiating between these two classes of traffic in bottleneck links and queuing them separately so that both classes can deliver satisfactory quality of experience for their applications. In particular, the NQB PHB provides a shallow-buffered, best-effort service as a complement to a default deep-buffered best-effort service.

To be clear, a network implementing the NQB PHB solely provides isolation for traffic classified as behaving in conformance with the NQB DSCP (and optionally enforces that behavior). It is the NQB senders' behavior itself which results in low latency and low loss.

2. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

3. Context

3.1. Non-Queue-Building Behavior

There are many applications that send traffic at relatively low data rates and/or in a fairly smooth and consistent manner such that they are highly unlikely to exceed the available capacity of the network path between source and sink. These applications might themselves only cause very small, transient queues to form in network buffers, but nonetheless they can be subjected to packet delay and delay variation as a result of sharing a network buffer with applications that tend to cause large and/or standing queues to form. Many of these applications are negatively affected by excessive packet delay and delay variation. Such applications are ideal candidates to be queued separately from the applications that are the cause of queue build-up, latency and loss.

In contrast, Queue-Building (QB) flows include those that use TCP or QUIC, with Cubic, Reno or other TCP congestion control algorithms that probe for the link capacity and induce latency and loss as a result. Other types of QB flows include those that send at a high burst rate even if the long-term average data rate is much lower.

3.2. Relationship to the Diffserv Architecture

The IETF has defined the Differentiated Services architecture [RFC2475] with the intention that it allows traffic to be marked in a manner that conveys the performance requirements of that traffic either quantitatively or in a relative sense (i.e. priority). The architecture defines the use of the Diffserv field [RFC2474] for

this purpose, and numerous RFCs have been written that describe recommended interpretations of the values (Diffserv Code Points) of the field, and standardized treatments (traffic conditioning and per-hop-behaviors) that can be implemented to satisfy the performance requirements of traffic so marked.

While this architecture is powerful and flexible enough to be configured to meet the performance requirements of a variety of applications and traffic categories, or to achieve differentiated service offerings, it has proven problematic to enable its use for these purposes end-to-end across the Internet.

This difficulty is in part due to the fact that meeting the performance requirements of an application in an end-to-end context involves all the networks in the path agreeing on what those requirements are and sharing an interest in meeting them. In many cases this is made more difficult since the performance "requirements" are not strict ones (e.g., applications will degrade in some manner as loss/latency/jitter increase), so the importance of meeting them for any particular application in some cases involves a judgment as to the value of avoiding some amount of degradation in quality for that application in exchange for an increase in the degradation of another application.

Further, in many cases the implementation of Diffserv PHBs has historically involved prioritization of service classes with respect to one another, which sets up the zero-sum game alluded to in the previous paragraph, and results in the need to limit access to higher priority classes via mechanisms such as access control, admission control, traffic conditioning and rate policing, and/or to meter and bill for carriage of such traffic. These mechanisms can be difficult or impossible to implement in an end-to-end context.

Finally, some jurisdictions impose regulations that limit the ability of networks to provide differentiation of services, in large part based on the belief that doing so necessarily involves prioritization or privileged access to bandwidth, and thus a benefit to one class of traffic always comes at the expense of another.

In contrast, the NQB PHB has been designed with the goal that it avoids many of these issues, and thus could conceivably be deployed end-to-end across the Internet. The intent of the NQB DSCP is that it signals verifiable behavior that permits the sender to request differentiated treatment. Also, the NQB traffic is to be given a separate queue with priority equal to default traffic and given no reserved bandwidth other than the bandwidth that it shares with default traffic. As a result, the NQB PHB does not aim to meet specific application performance requirements. Instead, the goal of the NQB PHB is to provide statistically better loss, latency, and

jitter performance for traffic that is itself only an insignificant contributor to those degradations. The PHB is also designed to minimize any incentives for a sender to mismark its traffic, since neither higher priority nor reserved bandwidth are being offered. These attributes eliminate many of the trade-offs that underlie the handling of differentiated service classes in the Diffserv architecture as it has traditionally been defined. They also significantly simplify access control and admission control functions, reducing them to simple verification of behavior.

3.3. Relationship to L4S

The NQB DSCP and PHB described in this draft have been defined to operate independently of the experimental L4S Architecture [I-D.ietf-tsvwg-l4s-arch]. Nonetheless, the NQB-marked traffic flows are intended to be compatible with [I-D.ietf-tsvwg-l4s-arch], with the result being that NQB traffic and L4S traffic can share the low-latency queue in an L4S DualQ node [I-D.ietf-tsvwg-aqm-dualq-coupled]. Compliance with the DualQ Coupled AQM requirements (Section 2.5 of [I-D.ietf-tsvwg-aqm-dualq-coupled]) is considered sufficient to support the NQB PHB requirement of fair allocation of bandwidth between the QB and NQB queues (Section 5). Note that these requirements in turn require compliance with all the requirements in Section 5 of [I-D.ietf-tsvwg-ecn-l4s-id].

Applications that comply with both the NQB sender requirements in Section 4.1 and the L4S "Prague" requirements in Section 4 of [I-D.ietf-tsvwg-ecn-l4s-id] could mark their packets both with the NQB DSCP and with the ECT(1) value. Packets marked with both codepoints SHOULD NOT be subject to less stringent policing than they would with either codepoint alone.

4. DSCP Marking of NQB Traffic

4.1. Non-Queue-Building Sender Requirements

Flows that are eligible to be marked with the NQP DSCP are typically UDP flows that send traffic at a low data rate relative to typical network path capacities. Current examples include many online games, voice chat, DNS lookups, and real-time IoT analytics data. Here the data rate is limited by the application itself rather than by network capacity - these applications send at most a few packets per RTT or a data rate of no more than about 1 percent of the "typical" network path capacity. In today's network, where access network data rates are typically on the order of 100 Mbps, this implies 1 Mbps as an upper limit. In addition, these applications send their traffic in a smooth (i.e. paced) manner, where the number of bytes sent in

any time interval "T" is less than or equal to R * T + 1500 bytes, where "R" is the maximum rate described above.

Note that, while such flows ordinarily don't implement a traditional congestion control mechanism, they nonetheless are expected to comply with existing guidance for safe deployment on the Internet, for example the requirements in [RFC8085] and Section 2 of [RFC3551] (also see the circuit breaker limits in Section 4.3 of [RFC8083] and the description of inelastic pseudowires in Section 4 of [RFC7893]). To be clear, the description of NQB-marked flows in this document should not be interpreted as suggesting that such flows are in any way exempt from this responsibility.

Applications that align with the description of behavior in the preceding paragraphs in this section SHOULD identify themselves to the network using a Diffserv Code Point (DSCP) of 45 (decimal) so that their packets can be queued separately from QB flows. The choice of the value 45 is motivated in part by the desire to achieve separate queuing in existing WiFi networks (see Section 8.3) and by the desire to make implementation of the PHB simpler in network gear that has the ability to classify traffic based on ranges of DSCP value (see Section 4.2 for further discussion). In networks where another (e.g., a local-use) codepoint is designated for NQB traffic, or where specialized PHBs are available that can meet specific application requirements (e.g., a guaranteed-latency path for voice traffic), it could be preferred to use another DSCP. In end systems where the choice of using DSCP 45 is not available to the application, the CS5 DSCP (40 decimal) could be used as a fallback. See Section 4.2 for rationale as to why this choice could be fruitful.

If the application's traffic exceeds the rate equation provided in the first paragraph of this section, the application SHOULD NOT mark its traffic with the NQB DSCP. In such a case, the application could instead consider implementing a low latency congestion control mechanism as described in [I-D.ietf-tsvwg-ecn-l4s-id]. At the time of writing, it is believed that 1 Mbps is a reasonable upper bound on instantaneous traffic rate for an NQB-marked application, but this value is of course subject to the context in which the application is expected to be deployed.

An application that marks its traffic as NQB but happens to exceed the available path capacity (even on an instantaneous basis) runs the risk of being subjected to a traffic protection algorithm (see Section 5.2), which could result in the excess traffic being discarded or queued separately as default traffic (and thus potentially delivered out of order). As a result, applications that aren't clearly beneath the threshold described above would need to weigh the risk of additional loss or out-of-order delivery against

the expected latency benefits of NQB treatment in determining whether to mark their packets as NQB.

The sender requirements outlined in this section are all related to observable attributes of the packet stream, which makes it possible for network elements (including nodes implementing the PHB) to monitor for inappropriate usage of the DSCP, and re-mark traffic that does not comply. This functionality, when implemented as part of the PHB is described in Section 5.2.

4.2. Aggregation of the NQB DSCP with other Diffserv PHBs

It is RECOMMENDED that networks and nodes that do not support the NQB PHB be configured to treat NQB-marked traffic the same as traffic marked "Default". It is additionally RECOMMENDED that such networks and nodes simply classify the NQB DSCP into the same treatment aggregate as Default traffic, or encapsulate the NQB-marked packet, rather than re-marking NQB traffic as Default. This preservation of the NQB marking enables hops further along the path to provide the NQB PHB successfully.

In backbone and core network switches (particularly if shallow-buffered), as well as in nodes that do not typically experience congestion, treating NQB-marked traffic the same as Default might be sufficient to preserve loss/latency/jitter performance for NQB traffic. In other nodes, treating NQB-marked traffic as Default could result in degradation of loss/latency/jitter performance but is recommended nonetheless in order to preserve the incentives described in Section 5. An alternative, in controlled environments where there is no risk of mismarking of traffic, would be to aggregate NQB-marked traffic with real-time, latency sensitive traffic. Similarly, networks and nodes that aggregate service classes as discussed in [RFC5127] and [RFC8100] might not be able to provide a PDB/PHB that meets the requirements of this document. In these cases it is RECOMMENDED that NQB-marked traffic be aggregated into the Elastic Treatment Aggregate (for [RFC5127] networks) or the Default / Elastic Treatment Aggregate (for [RFC8100] networks), although in some cases a network operator might instead choose to aggregate NQB traffic into the (Bulk) Real-Time Treatment Aggregate. Either approach comes with trade-offs: when the aggregated traffic encounters a bottleneck, aggregating with Default/Elastic traffic could result in a degradation of loss/latency/jitter performance for NQB traffic, while aggregating with Real-Time (assuming such traffic is provided a prioritized PHB) risks creating an incentive for mismarking of non-compliant traffic as NQB (except in controlled environments). In either case, the NQB DSCP SHOULD be preserved (possibly via encapsulation) in order to limit the negative impact that such networks would have on end-to-end performance for NQB traffic. This aligns with recommendations in [RFC5127].

Nodes that support the NQB PHB may choose to aggregate other service classes into the NQB queue. This is particularly useful in cases where specialized PHBs for these other service classes are not provided. Candidate service classes for this aggregation would include those that carry inelastic traffic that has low to very-low tolerance for loss, latency and/or jitter as discussed in [RFC4594]. These could include Telephony (EF/VA), Signaling (CS5), Real-Time Interactive (CS4) and Broadcast Video (CS3). Or, in some networks, equipment limitations may necessitate aggregating all traffic marked with DSCPs 40-47 (i.e., whose three MSBs are 0b101). As noted in Section 4.1, the choice of the value 45 is motivated in part by the desire to make this aggregation simpler in network equipment that can classify packets via comparing the DSCP value to a range of configured values.

4.3. End-to-end usage and DSCP Re-marking

In contrast to some existing standard PHBs, many of which are typically only meaningful within a Diffserv Domain (e.g., an AS or an enterprise network), this PHB is expected to be used end-to-end across the Internet, wherever suitable operator agreements apply. Under the [RFC2474] model, this requires that the corresponding DSCP is recognized by all operators and mapped across their boundaries accordingly.

If NQB support is extended across a DiffServ domain boundary, the interconnected networks agreeing to support NQB SHOULD use the value 45 for NQB at network interconnection, unless a different DSCP is explicitly documented in the TCA (Traffic Conditioning Agreement, see [RFC2475]) for that interconnection. Similar to the handling of DSCPs for other PHBs (and as discussed in [RFC2475]), networks can re-mark NQB traffic to a DSCP other than 45 for internal usage. To ensure reliable end-to-end NQB PHB treatment, the appropriate NQB DSCP should be restored when forwarding to another network.

In order to enable interoperability with WiFi equipment as described in Section 8.3.1, networks SHOULD ensure NQB traffic is marked DSCP 45 prior to a customer access link, subject to the safeguards described below and in that section.

4.3.1. Unmanaged Networks

As discussed in Section 4 of [RFC2475], there may be cases where a network operator is delivering traffic into a network outside of their control, where there is no knowledge of the traffic management capabilities of the downstream domain, and no agreement in place (e.g., a residential ISP delivering traffic to a customer's home network that may contain a legacy WiFi AP). In such cases, the network operator should presume that the existing network equipment

in the downstream domain does not support the NQB PHB and might instead prioritize traffic marked with the NQB DSCP. In these cases, the network operator SHOULD take precautions to prevent issues that could be caused by excessive NQB traffic and/or traffic mismarked as NQB.

Network equipment that is intended to deliver traffic into such unmanaged networks (e.g., an access network gateway for a residential ISP) SHOULD by default ensure that NQB traffic is re-marked with a DSCP that is unlikely to result in prioritized treatment in the downstream domain, such as DSCP 0 (Default). Such equipment SHOULD support the ability to configure the re-marking, so that (when it is appropriate) traffic can be delivered as NQB-marked. As an alternative to re-marking, the operator could deploy a traffic protection (see Section 5.2) or a shaping/policing function on NQB-marked traffic that minimizes the potential for negative impacts on Default traffic. It should be noted that a traffic protection function as defined in this document might only provide protection from issues occuring in subsequent network hops if the device implementing the traffic protection function is the bottleneck link on the path, so it might not be a solution for all situations. In the case that a traffic policing function or a rate shaping function is applied to the aggregate of NQB traffic destined to such a downstream domain, the policer/shaper rate SHOULD be set to either 5% of the interconnection data rate, or 5% of the typical rate for such interconnections, whichever is greater, with excess traffic being either dropped or re-marked and classified for Default forwarding. A traffic policing function SHOULD allow approximately 100 ms of burst tolerance (e.g. a token bucket depth equal to 100 ms multiplied by the policer rate). A traffic shaping function SHOULD allow approximately 10 ms of burst tolerance, and approximately 50 ms of buffering.

The recommendation to limit NQB traffic to 5% in these situations is based on an expectation of support for at least 5 simultaneous NQB streams, and SHOULD be adjusted according to local network policy.

4.4. The NQB DSCP and Tunnels

[RFC2983] discusses tunnel models that support Diffserv. It describes a "uniform model" in which the inner DSCP is copied to the outer header at encapsulation, and the outer DSCP is copied to the inner header at decapsulation. It also describes a "pipe model" in which the outer DSCP is not copied to the inner header at decapsulation. Both models can be used in conjunction with the NQB PHB. In the case of the pipe model, any DSCP manipulation (re-marking) of the outer header by intermediate nodes would be discarded at tunnel egress, potentially improving the possibility of achieving NQB treatment in subsequent nodes.

As is discussed in [RFC2983], tunnel protocols that are sensitive to reordering can result in undesirable interactions if multiple DSCP PHBs are signaled for traffic within a tunnel instance. This is true for NQB-marked traffic as well. If a tunnel contains a mix of QB and NQB traffic, and this is reflected in the outer DSCP in a network that supports the NQB PHB, it would be necessary to avoid a reordering-sensitive tunnel protocol.

5. Non-Queue-Building PHB Requirements

It is important that incentives are aligned correctly, i.e., that there is a benefit to the application in marking its packets correctly, and a disadvantage (or at least no benefit) to an application in intentionally mismarking its traffic. Thus, a useful property of nodes (i.e. network switches and routers) that support separate queues for NQB and QB flows is that for flows consistent with the NQB sender requirements in Section 4.1, the NQB queue would likely be a better choice than the QB queue; and for flows inconsistent with those requirements, the QB queue would likely be a better choice than the NQB queue (this is discussed further in this section and Section 12). By adhering to these principles, there is no incentive for senders to mismark their traffic as NQB. As mentioned previously, the NQB designation and marking is intended to convey verifiable traffic behavior, as opposed to simply a desire for differentiated treatment. As a result, any mismarking can be identified by the network.

5.1. Primary Requirements

A node supporting the NQB PHB makes no guarantees on latency or data rate for NQB-marked flows, but instead aims to provide an upper-bound to queuing delay for as many such marked flows as it can and shed load when needed.

A node supporting the NQB PHB MUST provide a queue for Non-Queue-Building traffic separate from any queue used for Queue-Building traffic.

A node supporting the NQB PHB SHOULD NOT rate limit or rate police the aggregate of NQB traffic separately from Queue-Building traffic of equivalent importance. An exception to this recommendation is discussed in Section 4.3.1.

The NQB queue SHOULD be given equivalent forwarding preference compared to Queue-Building traffic of equivalent importance. The node SHOULD provide a scheduler that allows QB and NQB traffic of equivalent importance to share the link in a fair manner, e.g., a deficit round-robin scheduler with equal weights. Compliance with

these recommendations helps to ensure that there are no incentives for QB traffic to be mismarked as NQB.

A node supporting the NQB PHB SHOULD treat traffic marked as Default (DSCP=0) as QB traffic having equivalent importance to the NQB-marked traffic. A node supporting the NQB DSCP MUST support the ability to configure the classification criteria that are used to identify QB and NQB traffic of equivalent importance.

The NQB queue SHOULD have a buffer size that is significantly smaller than the buffer provided for QB traffic. It is expected that most QB traffic is engineered to work well when the network provides a relatively deep buffer (e.g., on the order of tens or hundreds of ms) in nodes where support for the NQB PHB is advantageous (i.e., bottleneck nodes). Providing a similarly deep buffer for the NQB queue would be at cross purposes to providing very low queueing delay and would erode the incentives for QB traffic to be marked correctly. An NQB buffer size less than or equal to 10 ms is RECOMMENDED.

In some cases, existing network gear has been deployed that cannot readily be upgraded or configured to support the PHB requirements. This equipment might however be capable of loosely supporting an NQB service - see Section 8.3.1 for details and an example where this is particularly important. A similar approach might prove necessary in other network environments.

5.2. Traffic Protection

It is possible that due to an implementation error or misconfiguration, a QB flow would end up getting mismarked as NQB, or vice versa. In the case of a low data rate flow that isn't marked as NQB and therefore ends up in the QB queue, it would only impact its own quality of service, and so it seems to be of lesser concern. However, a QB flow that is mismarked as NQB would cause queuing delays and/or loss for all the other flows that are sharing the NQB queue.

To prevent this situation from harming the performance of the flows that are in compliance with the requirements in Section 4.1, network elements that support the NQB PHB SHOULD support a "traffic protection" function that can identify flows that are inconsistent with the sender requirements in Section 4.1, and either re-mark those flows/packets as Default and reclassify them to the QB queue or discard the offending traffic. Such a function SHOULD be implemented in an objective and verifiable manner, basing its decisions upon the behavior of the flow rather than on application-layer constructs. While it is possible to utilize a per-flow rate policer to perform this function, it is RECOMMENDED that traffic

protection algorithms base their decisions on the detection of actual queuing, as opposed to simply packet arrival rate or data rate. It could be advantageous for a traffic protection function to employ hysteresis to prevent borderline flows from being reclassified capriciously.

The traffic protection function described here requires that the network element maintain some sort of flow state. The traffic protection function MUST be designed to fail gracefully in the case that the flow state is exhausted.

One example traffic protection algorithm can be found in [I-D.briscoe-docsis-q-protection]. This algorithm maintains per-flow state for up to 32 simultaneous "queue-building" flows, and shared state for any additional flows in excess of that number. [NOTE (to be removed by RFC-Editor): This ISE submission draft is approved for publication as an RFC, the NQB draft should be held for publication until the queue protection RFC can be referenced.]

There are some situations where such a function is potentially not necessary. For example, a network element designed for use in controlled environments (e.g., enterprise LAN). Additionally, some networks might prefer to police the application of the NQB DSCP at the ingress edge, so that per-hop traffic protection is not needed.

5.3. Guidance for Very Low-Rate Links

The NQB sender requirements in Section 4.1 place responsibility in the hands of the application developer to determine the likelihood that the application's sending behavior could result in a queue forming along the path. These requirements rely on application developers having a reasonable sense for the network context in which their application is to be deployed. Even so, there will undoubtedly be networks that contain links having a data rate that is below the lower end of what is considered "typical", and some of these links could even be below the instantaneous sending rate of some NQB-marked applications.

To limit the consequences of this scenario, operators of such networks SHOULD utilize a traffic protection function that is more tolerant of burstiness (i.e., a temporary queue). Alternatively, operators of such networks MAY choose to disable NQB support (and thus aggregate NQB-marked traffic with Default traffic) on these low-speed links. For links that are less than ten percent of "typical" path rates, it is RECOMMENDED that NQB support be disabled and for NQB-marked traffic to thus be carried using the default PHB.

6. Impact on Higher Layer Protocols

Network elements that support the NQB PHB and that support traffic protection as discussed in the previous section introduce the possibility that flows classified into the NQB queue could experience out of order delivery or packet loss if their behavior is not consistent with NQB. This is particularly true if the traffic protection algorithm makes decisions on a packet-by-packet basis. In this scenario, a flow that is (mis)marked as NQB and that causes a queue to form in this bottleneck link could see some of its packets forwarded by the NQB queue, and some of them either discarded or redirected to the QB queue. In the case of redirection, depending on the queueing latency and scheduling within the network element, this could result in packets being delivered out of order. As a result, the use of the NQB DSCP by a higher layer protocol carries some risk that an increased amount of out of order delivery or packet loss will be experienced. This characteristic provides one disincentive for mismarking of traffic.

7. Configuration and Management

As required above, nodes supporting the NQB PHB provide for the configuration of classifiers that can be used to differentiate between QB and NQB traffic of equivalent importance. The default for such classifiers is recommended to be the assigned NQB DSCP (to identify NQB traffic) and the Default (0) DSCP (to identify QB traffic).

8. Example Use Cases

8.1. DOCSIS Access Networks

Residential cable broadband Internet services are commonly configured with a single bottleneck link (the access network link) upon which the service definition is applied. The service definition, typically an upstream/downstream data rate tuple, is implemented as a configured pair of rate shapers that are applied to the user's traffic. In such networks, the quality of service that each application receives, and as a result, the quality of experience that it generates for the user is influenced by the characteristics of the access network link.

To support the NQB PHB, cable broadband services MUST be configured to provide a separate queue for NQB-marked traffic. The NQB queue MUST be configured to share the service's rate shaped bandwidth with the queue for QB traffic.

8.2. Mobile Networks

Historically, 3GPP mobile networks have utilised "bearers" to encapsulate each user's user plane traffic through the radio and core networks. A "dedicated bearer" can be allocated a Quality of Service (QoS) to apply any prioritisation to its flows at queues and radio schedulers. Typically, an LTE operator provides a dedicated bearer for IMS VoLTE (Voice over LTE) traffic, which is prioritised in order to meet regulatory obligations for call completion rates; and a "best effort" default bearer, for Internet traffic. The "best effort" bearer provides no guarantees, and hence its buffering characteristics are not compatible with low-latency traffic. The 5G radio and core systems offer more flexibility over bearer allocation, meaning bearers can be allocated per traffic type (e.g., loss-tolerant, low-latency etc.) and hence support more suitable treatment of Internet real-time flows.

To support the NQB PHB, the mobile network SHOULD be configured to give User Equipment a dedicated, low-latency, non-GBR, EPS bearer, e.g., one with QCI 7, in addition to the default EPS bearer; or a Data Radio Bearer with 5QI 7 in a 5G system (see Table 5.7.4-1: Standardized 5QI to QoS characteristics mapping in [SA-5G]).

A packet carrying the NQB DSCP SHOULD be routed through the dedicated low-latency EPS bearer. A packet that has no associated NQB marking SHOULD NOT be routed through the dedicated low-latency EPS bearer.

8.3. WiFi Networks

WiFi networking equipment compliant with 802.11e/n/ac/ax [IEEE802-11] generally supports either four or eight transmit queues and four sets of associated Enhanced Multimedia Distributed Control Access (EDCA) parameters (corresponding to the four WiFi Multimedia (WMM) Access Categories) that are used to enable differentiated media access characteristics. As discussed in [RFC8325], most existing WiFi implementations use a default DSCP to User Priority mapping that utilizes the most significant three bits of the Diffserv Field to select "User Priority" which is then mapped to the four WMM Access Categories. [RFC8325] also provides an alternative mapping that more closely aligns with the DSCP recommendations provided by the IETF. In the case of some managed WiFi gear, this mapping can be controlled by the network operator, e.g., via TR-369 [TR-369].

In addition to the requirements provided in other sections of this document, to support the NQB PHB, WiFi equipment (including equipment compliant with [RFC8325]) SHOULD map the NQB codepoint 45

into a separate queue in the same Access Category as the queue that carries default traffic (i.e. the Best Effort Access Category).

8.3.1. Interoperability with Existing WiFi Networks

While some existing WiFi equipment might be capable (in some cases via firmware update) of supporting the NQB PHB requirements, many currently deployed devices cannot be configured in this way. As a result, the remainder of this section discusses interoperability with these existing WiFi networks, as opposed to PHB compliance.

Since this equipment is widely deployed, and the WiFi link is commonly a bottleneck link, the performance of NQB-marked traffic across such links could have a significant impact on the viability and adoption of the NQB DSCP and PHB. Depending on the DSCP used to mark NQB traffic, existing WiFi equipment that uses the default mapping of DSCPs to Access Categories and the default EDCA parameters will support either the NQB PHB requirement for separate queuing of NQB traffic, or the recommendation to treat NQB traffic with priority equal to Default traffic, but not both.

The DSCP value 45 is recommended for NQB. This maps NQB to UP_5 using the default mapping, which is in the "Video" Access Category. While this choice of DSCP enables these WiFi systems to support the NQB PHB requirement for separate queuing, existing WiFi devices generally utilize EDCA parameters that result in statistical prioritization of the "Video" Access Category above the "Best Effort" Access Category. In addition this equipment does not support the remaining NQB PHB recommendations in Section 5. The rationale for the choice of DSCP 45 as well as its ramifications, and remedies for its limitations are discussed further below.

The choice of separated queuing rather than equal priority in existing WiFi networks was motivated by the following:

* Separate queuing is necessary in order to provide a benefit for NQB-marked traffic.

* All known WiFi gear has hardware support (albeit generally not exposed for user control) for adjusting the EDCA parameters in order to meet the equal priority recommendation. This is discussed further below.

* NQB-compliant traffic is unlikely to cause more degradation to lower priority Access Categories than the existing recommended Video Access Category traffic types: Broadcast Video, Multimedia Streaming, Multimedia Conferencing from [RFC8325], and AudioVideo, ExcellentEffort from [QOS_TRAFFIC_TYPE].

* Application instances on WiFi client devices are already free to choose any Access Category that they wish, regardless of their sending behavior, without any policing of usage. So, the choice of 45 for NQB creates no new avenues for non-NQB-compliant client applications to exploit the prioritization function in WiFi.

* Several existing client applications that are compatible with the NQB sender requirements already select the Video Access Category, and thus would not see a degradation in performance by transitioning to the NQB DSCP, regardless of whether the network supported the PHB.

* For application traffic that originates outside of the WiFi network, and thus is transmitted by the Access Point, opportunities exist in the upstream network components to police the usage of the NQB codepoint and potentially re-mark traffic that is considered non-compliant, as is recommended in Section 4.3.1. A residential ISP that re-marks the Diffserv field to zero, bleaches all DSCPs and hence would not be impacted by the introduction of traffic marked as NQB. Furthermore, any change to this practice ought to be done alongside the implementation of those recommendations in the current document.

The choice of Video Access Category rather than the Voice Access Category was motivated by the desire to minimize the potential for degradation of Best Effort Access Category traffic. The choice of Video Access Category rather than the Background Access Category was motivated by the much greater potential of degradation to NQB traffic that would be caused by the vast majority of traffic in most WiFi networks, which utilizes the Best Effort Access Category.

If left unchanged, the prioritization of Video Access Category traffic (particularly in the case of traffic originating outside of the WiFi network as mentioned above) could erode the principle of alignment of incentives. In order to preserve the incentives principle for NQB, WiFi systems SHOULD be configured such that the EDCA parameters for the Video Access Category match those of the Best Effort Access Category. These changes can be deployed in managed WiFi systems or those deployed by an ISP and are intended for situations when the vast majority of traffic that would use AC_VI is NQB. In other situations (e.g., consumer-grade WiFi gear deployed by an ISP's customer) this configuration might not be possible, and the requirements and recommendations in Section 4.3.1 would apply.

Similarly, systems that utilize [RFC8325] but that are unable to fully support the PHB requirements, SHOULD map the recommended NQB codepoint 45 (or the locally determined alternative) to UP_5 in the "Video" Access Category.

9. Acknowledgements

Thanks to Diego Lopez, Stuart Cheshire, Brian Carpenter, Bob Briscoe, Greg Skinner, Toke Hoeiland-Joergensen, Luca Muscariello, David Black, Sebastian Moeller, Ruediger Geib, Jerome Henry, Steven Blake, Jonathan Morton, Roland Bless, Kevin Smith, Martin Dolly, and Kyle Rose for their review comments. Thanks also to Gorry Fairhurst, Ana Custura, and Ruediger Geib for their input on selection of appropriate DSCPs.

10. IANA Considerations

This document requests that IANA assign the Differentiated Services Field Codepoint (DSCP) 45 ('0b101101', 0x2D) from the "Differentiated Services Field Codepoints (DSCP)" registry (https://www.iana.org/assignments/dscp-registry/) ("DSCP Pool 3 Codepoints", Codepoint Space xxxx01, Standards Action) as the RECOMMENDED codepoint for Non-Queue-Building behavior.

IANA should update this registry as follows:

*Name: NQB

*Value (Binary): 101101

*Value (Decimal): 45

*Reference: this document

11. Implementation Status

Note to RFC Editor: This section should be removed prior to publication

The NQB PHB is implemented in equipment compliant with the current DOCSIS 3.1 specification, published by CableLabs at: CableLabs Specifications Search.

CableLabs maintains a list of production cable modem devices that are Certified as being compliant to the DOCSIS Specifications, this list is available at https://www.cablelabs.com/wp-content/uploads/2013/10/cert_qual.xlsx. DOCSIS 3.1 modems certified in CW 134 or greater implement the NQB PHB. This includes products from Arcadyan Technology Corporation, Arris, AVM, Castlenet, Commscope, Hitron, Motorola, Netgear, Sagemcom and Vantiva. There are additional production implementations that have not been Certified as compliant to the specification, but which have been tested in non-public Interoperability Events. These implementations are all proprietary, not available as open source.

12. Security Considerations

When the NQB PHB is fully supported in bottleneck links, there is no incentive for a Queue-Building application to mismark its packets as NQB (or vice versa). If a Queue-Building flow were to mismark its packets as NQB, it would be unlikely to receive a benefit by doing so, and it would usually experience a degradation. The nature of the degradation would depend on the specifics of the PHB implementation (and on the presence or absence of a traffic protection function), but could include excessive packet loss, excessive latency variation and/or excessive out-of-order delivery. If a Non-Queue-Building flow were to fail to mark its packets as NQB, it could suffer the latency and loss typical of sharing a queue with capacity seeking traffic.

In order to preserve low latency performance for NQB traffic, networks that support the NQB PHB will need to ensure that mechanisms are in place to prevent malicious NQB-marked traffic from causing excessive queue delays. Section 5.2 recommends the implementation of a traffic protection mechanism to achieve this goal but recognizes that other options might be more desirable in certain situations. The recommendations on traffic protection mechanisms in this document presume that some type of "flow" state be maintained in order to differentiate between flows that are causing queuing delay and those that aren't. Since this flow state is likely finite, this opens up the possibility of flow-state exhaustion attacks. While this document requires that traffic protection mechanisms be designed with this possibility in mind, the outcomes of flow-state exhaustion would depend on the implementation.

Notwithstanding the above, the choice of DSCP for NQB does allow existing WiFi networks to readily (and by default) support some of the PHB requirements, but without a traffic protection function, and (when left in the default state) by giving NQB traffic higher priority than QB traffic. This is not considered to be a compliant implementation of the PHB. These existing WiFi networks currently provide priority to half of the DSCP space, including the NQB DSCP. While the NQB marking could be abused in order to gain priority on such links, the potential presence of traffic protection functions along the path (which apply to the NQB marking alone) would seem to make it less attractive for such abuse than any of the other 31 DSCP values that are provided high priority.

The NQB signal (DSCP) is not integrity protected and could be changed by an on-path attacker. While re-marking DSCPs is permitted for various reasons (some are discussed in this document, others can be found in [RFC2474] and [RFC2475]), if done maliciously, this might negatively affect the QoS of the tampered flow.

13. References

13.1. 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, <https://www.rfc-editor.org/info/rfc2119>.

[RFC2474] Nichols, K., Blake, S., Baker, F., and D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", RFC 2474, DOI 10.17487/RFC2474, December 1998, <https://www.rfc-editor.org/info/rfc2474>.

[RFC2983] Black, D. and RFC Publisher, "Differentiated Services and Tunnels", RFC 2983, DOI 10.17487/RFC2983, October 2000, <https://www.rfc-editor.org/info/rfc2983>.

[RFC8085] Eggert, L., Fairhurst, G., and G. Shepherd, "UDP Usage Guidelines", BCP 145, RFC 8085, DOI 10.17487/RFC8085, March 2017, <https://www.rfc-editor.org/info/rfc8085>.

[RFC8174] Leiba, B. and RFC Publisher, "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>.

[RFC8325] Szigeti, T., Henry, J., and F. Baker, "Mapping Diffserv to IEEE 802.11", RFC 8325, DOI 10.17487/RFC8325, February 2018, <https://www.rfc-editor.org/info/rfc8325>.

13.2. Informative References

[Barik] Barik, R., Welzl, M., Elmokashfi, A., Dreibholz, T., and S. Gjessing, "Can WebRTC QoS Work? A DSCP Measurement Study", ITC 30, September 2018.

[Custura] Custura, A., Venne, A., and G. Fairhurst, "Exploring DSCP modification pathologies in mobile edge networks", TMA , 2017.

[I-D.briscoe-docsis-q-protection] Briscoe, B. and G. White, "The DOCSIS(r) Queue Protection Algorithm to Preserve Low Latency", Work in Progress, Internet-Draft, draft-briscoe-docsis-q-protection-06, 13 May 2022, <https://

www.ietf.org/archive/id/draft-briscoe-docsis-q-protection-06.txt>.

[I-D.ietf-tsvwg-aqm-dualq-coupled] De Schepper, K., Briscoe, B., and G. White, "DualQ Coupled AQMs for Low Latency, Low Loss and Scalable Throughput (L4S)", Work in Progress, Internet-Draft, draft-ietf-tsvwg-aqm-dualq-coupled-25, 29 August 2022, <https://www.ietf.org/archive/id/draft-ietf-tsvwg-aqm-dualq-coupled-25.txt>.

[I-D.ietf-tsvwg-dscp-considerations] Custura, A., Fairhurst, G., and R. Secchi, "Considerations for Assigning a new Recommended DiffServ Codepoint (DSCP)", Work in Progress, Internet-Draft, draft-ietf-tsvwg-dscp-considerations-08, 13 December 2022, <https://www.ietf.org/archive/id/draft-ietf-tsvwg-dscp-considerations-08.txt>.

[I-D.ietf-tsvwg-ecn-l4s-id] De Schepper, K. and B. Briscoe, "Explicit Congestion Notification (ECN) Protocol for Very Low Queuing Delay (L4S)", Work in Progress, Internet-Draft, draft-ietf-tsvwg-ecn-l4s-id-29, 29 August 2022, <https://www.ietf.org/archive/id/draft-ietf-tsvwg-ecn-l4s-id-29.txt>.

[I-D.ietf-tsvwg-l4s-arch] Briscoe, B., De Schepper, K., Bagnulo, M., and G. White, "Low Latency, Low Loss, Scalable Throughput (L4S) Internet Service: Architecture", Work in Progress, Internet-Draft, draft-ietf-tsvwg-l4s-arch-20, 29 August 2022, <https://www.ietf.org/archive/id/draft-ietf-tsvwg-l4s-arch-20.txt>.

[IEEE802-11] IEEE-SA, "IEEE 802.11-2020", IEEE 802, December 2020, <https://standards.ieee.org/standard/802_11-2020.html>.

[QOS_TRAFFIC_TYPE] Microsoft, Corporation, "QOS_TRAFFIC_TYPE enumeration", 2022, <https://learn.microsoft.com/en-us/windows/win32/api/qos2/ne-qos2-qos_traffic_type>.

[RFC2475] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., and W. Weiss, "An Architecture for Differentiated Services", RFC 2475, DOI 10.17487/RFC2475, December 1998, <https://www.rfc-editor.org/info/rfc2475>.

[RFC3551] Schulzrinne, H. and S. Casner, "RTP Profile for Audio and Video Conferences with Minimal Control", STD 65, RFC 3551, DOI 10.17487/RFC3551, July 2003, <https://www.rfc-editor.org/info/rfc3551>.

[RFC4594] Babiarz, J., Chan, K., and F. Baker, "Configuration Guidelines for DiffServ Service Classes", RFC 4594, DOI

10.17487/RFC4594, August 2006, <https://www.rfc-editor.org/info/rfc4594>.

[RFC5127] Chan, K., Babiarz, J., and F. Baker, "Aggregation of Diffserv Service Classes", RFC 5127, DOI 10.17487/RFC5127, February 2008, <https://www.rfc-editor.org/info/rfc5127>.

[RFC7893] Stein, Y., Black, D., and B. Briscoe, "Pseudowire Congestion Considerations", RFC 7893, DOI 10.17487/RFC7893, June 2016, <https://www.rfc-editor.org/info/rfc7893>.

[RFC8033] Pan, R., Natarajan, P., Baker, F., and G. White, "Proportional Integral Controller Enhanced (PIE): A Lightweight Control Scheme to Address the Bufferbloat Problem", RFC 8033, DOI 10.17487/RFC8033, February 2017, <https://www.rfc-editor.org/info/rfc8033>.

[RFC8034] White, G. and R. Pan, "Active Queue Management (AQM) Based on Proportional Integral Controller Enhanced (PIE) for Data-Over-Cable Service Interface Specifications (DOCSIS) Cable Modems", RFC 8034, DOI 10.17487/RFC8034, February 2017, <https://www.rfc-editor.org/info/rfc8034>.

[RFC8083] Perkins, C. and V. Singh, "Multimedia Congestion Control: Circuit Breakers for Unicast RTP Sessions", RFC 8083, DOI 10.17487/RFC8083, March 2017, <https://www.rfc-editor.org/info/rfc8083>.

[RFC8100] Geib, R., Ed. and D. Black, "Diffserv-Interconnection Classes and Practice", RFC 8100, DOI 10.17487/RFC8100, March 2017, <https://www.rfc-editor.org/info/rfc8100>.

[RFC8289] Nichols, K., Jacobson, V., McGregor, A., Ed., and J. Iyengar, Ed., "Controlled Delay Active Queue Management", RFC 8289, DOI 10.17487/RFC8289, January 2018, <https://www.rfc-editor.org/info/rfc8289>.

[RFC8290] Hoeiland-Joergensen, T., McKenney, P., Taht, D., Gettys, J., and E. Dumazet, "The Flow Queue CoDel Packet Scheduler and Active Queue Management Algorithm", RFC 8290, DOI 10.17487/RFC8290, January 2018, <https://www.rfc-editor.org/info/rfc8290>.

[SA-5G] 3GPP, "System Architecture for 5G", TS 23.501, 2019.

[TR-369] Broadband Forum, "The User Services Platform", January 2022, <https://usp.technology/specification/index.html>.

Appendix A. DSCP Re-marking Policies

Some network operators typically bleach (zero out) the Diffserv field on ingress into their network [I-D.ietf-tsvwg-dscp-considerations][Custura][Barik], and in some cases apply their own DSCP for internal usage. Bleaching the NQB DSCP is not expected to cause harm to default traffic, but it will severely limit the ability to provide NQB treatment end-to-end. Reports on existing deployments of DSCP manipulation [Custura] [Barik] categorize the re-marking behaviors into the following six policies: bleach all traffic (set DSCP to zero), set the top three bits (the former Precedence bits) on all traffic to 0b000, 0b001, or 0b010, set the low three bits on all traffic to 0b000, or re-mark all traffic to a particular (non-zero) DSCP value.

Regarding the DSCP value 45, there were no observations of DSCP manipulation reported in which traffic was marked 45 by any of these policies. Thus it appears that these re-marking policies would be unlikely to result in QB traffic being marked as NQB (45). In terms of the fate of NQB-marked traffic that is subjected to one of these policies, the result would be that NQB-marked traffic would be indistinguishable from some subset (possibly all) of other traffic. In the policies where all traffic is re-marked using the same (zero or non-zero) DSCP, the ability for a subsequent network hop to differentiate NQB traffic via DSCP would clearly be lost entirely.

In the policies where the top three bits are overwritten (see Section 4.2 of [I-D.ietf-tsvwg-dscp-considerations]), the NQB value (45) would receive the same marking as would the currently unassigned Pool 3 DSCPs 5,13,21,29,37,53,61, with all of these codepoints getting re-marked to DSCP = 5, 13 or 21 (depending on the overwrite value used). Since none of the DSCPs in the preceding lists are currently assigned by IANA, and they all are reserved for Standards Action, it is believed that they are not widely used currently, but this could vary based on local-usage, and could change in the future. If networks in which this sort of re-marking occurs (or networks downstream) classify the resulting codepoint (i.e. 5, 13, or 21) to the NQB PHB, or re-mark such traffic as 45, they risk treating as NQB other traffic, which was not originally marked as NQB. In addition, as described in Section 6 of [I-D.ietf-tsvwg-dscp-considerations] future assignments of these 0bxxx101 codepoints would need to be made with consideration of the potential that they all are treated as NQB in some networks.

For the policy in which the low three bits are set to 0b000, the NQB (45) value would be re-marked to CS5 and would be indistinguishable from CS5, VA, EF (and the unassigned DSCPs 41, 42, 43). Traffic marked using the existing standardized DSCPs in this list are likely to share the same general properties as NQB traffic (non capacity-

seeking, very low data rate or relatively low and consistent data rate). Similarly, any future recommended usage for DSCPs 41, 42, 43 would likely be somewhat compatible with NQB treatment, assuming that IP Precedence compatibility (see Section 1.5.4 of [RFC4594]) is maintained in the future. Here there might be an opportunity for a node to provide the NQB PHB or the CS5 PHB to CS5-marked traffic and retain some of the benefits of NQB marking. This could be another motivation to (as discussed in Section 4.2) classify CS5-marked traffic into NQB queue.

Authors' Addresses

Greg White
CableLabs

Email: g.white@cablelabs.com

Thomas Fossati
ARM

Email: Thomas.Fossati@arm.com