

QoS

衛星接続を使用した長距離電話では、会話が、短い間ですが認識できる程度に割り込みされ、不定期に中断されることがあります。このような中断は、ネットワークで送信されるパケットが到着する間隔の時間で、遅延と呼ばれます。音声やビデオなどのネットワーク トラフィックでは、長時間の遅延は許容されません。Quality of Service（QoS）機能を使用すると、重要なトラフィックのプライオリティを高くし、帯域幅の過剰な使用を防ぎ、ネットワーク ボトルネックを管理してパケットのドロップを防止できます。

ここでは、QoS ポリシーの適用方法について説明します。

- QoS について （1 ページ）
- QoS のガイドライン （3 ページ）
- QoS の設定 （4 ページ）
- QoS のモニター （9 ページ）
- QoS の履歴 （11 ページ）

QoS について

常に変化するネットワーク環境では、QoS は 1 回限りの構成ではなく、ネットワーク設計の継続的で不可欠な要素であることを考慮する必要があります。

この項では、ASA で使用できる QoS 機能について説明します。

サポートされている QoS 機能

ASA は、次の QoS の機能をサポートしています。

- ポリシング：分類されたフローがネットワーク帯域幅を大量に使用するのを防ぐため、クラスごとの最大使用帯域幅を制限できます。詳細については、「ポリシング （2 ページ）」を参照してください。
- プライオリティ キューイング：Voice over IP（VoIP）のような遅延を許されない重要なトラフィックについて、トラフィックを低遅延キューイング（LLQ）に指定することで、常に他のトラフィックより先に送信できます。プライオリティ キューイング （2 ページ）を参照してください。

トークン バケットとは

トークン バケットは、フロー内のデータを規制するデバイス（トラフィック ポリサーなど）の管理に使用されます。トークン バケット自体には、廃棄ポリシーまたはプライオリティ ポリシーはありません。むしろ、トークンバケットは、フローによって規制機能が過剰に働く場合に、トークンを廃棄し、送信キューの管理の問題はフローに任せます。

トークン バケットは、転送レートの正式な定義です。トークン バケットには、バースト サイズ、平均レート、時間間隔という 3 つのコンポーネントがあります。平均レートは通常 1 秒間のビット数で表されますが、次のような関係によって、任意の 2 つの値を 3 番目の値から求めることができます。

平均レート = バースト サイズ / 時間間隔

これらの用語の定義は次のとおりです。

- 平均レート：認定情報レート（CIR）とも呼ばれ、単位時間に送信または転送できるデータ量の平均値を指定します。
- バースト サイズ：認定バースト（Bc）サイズとも呼ばれ、スケジューリングに関する問題を発生させることなく単位時間内に送信できるトラフィックの量を、バーストあたりのバイト数で指定します。
- 時間間隔：測定間隔とも呼ばれ、バーストごとの時間を秒単位で指定します。

トークンバケットのたとえで言えば、トークンは特定のレートでバケットに入れられます。バケット自体には指定された容量があります。バケットがいっぱいになると、新しく到着するトークンは廃棄されます。各トークンは、送信元が一定の数のビットをネットワークに送信するための権限です。パケットを送信するため、規制機能はパケットサイズに等しい数のトークンをバケットから削除する必要があります。

パケットを送信するための十分なトークンがバケットにない場合、パケットは、パケットが廃棄されるか、ダウン状態とマークされるまで待機します。バケットがすでにトークンで満たされている場合、着信トークンはオーバーフローし、以降のパケットには使用できません。したがって、いつでも、送信元がネットワークに送信できる最大のバーストは、バケットのサイズにほぼ比例します。

ポリシング

ポリシングは、設定した最大レート（ビット/秒単位）を超えるトラフィックが発生しないようにして、1つのトラフィック クラスが全体のリソースを占有しないようにする方法です。トラフィックが最大レートを超えると、ASA は超過した分のトラフィックをドロップします。また、ポリシングでは、許可されるトラフィックの最大単一バーストも設定されます。

プライオリティ キューイング

LLQ プライオリティ キューイングを使用すると、特定のトラフィック フロー（音声やビデオのような遅延の影響を受けやすいトラフィックなど）をその他のトラフィックよりも優先でき

ます。プライオリティ キューイングでは、インターフェイスで LLQ プライオリティ キューが使用されます（インターフェイスのプライオリティ キューの設定 （6 ページ）を参照してください）。一方、他のトラフィックはすべて「ベストエフォート」キューに入ります。キューは無限大ではないため、いっぱいになってオーバーフローすることがあります。キューがいっぱいになると、以降のパケットはキューに入ることができず、すべてドロップされます。これはテール ドロップと呼ばれます。キューがいっぱいになることを避けるには、キューのバッファ サイズを大きくします。送信キューに入れることのできるパケットの最大数も微調整できます。これらのオプションを使用して、プライオリティ キューイングの遅延と強固さを制御できます。LLQ キュー内のパケットは、常に、ベストエフォート キュー内のパケットよりも前に送信されます。

QoS 機能の相互作用のしくみ

ASA で必要な場合は、個々の QoS 機能を単独で設定できます。ただし、普通は、たとえば一部のトラフィックを優先させて、他のトラフィックによって帯域幅の問題が発生しないようにするために、複数の QoS 機能を ASA に設定します。次のことを設定できます。

プライオリティ キューイング（特定のトラフィックについて）+ ポリシング（その他のトラフィックについて）

同じトラフィックのセットに対して、プライオリティ キューイングとポリシングを両方設定することはできません。

DSCP（DiffServ）の保存

DSCP（DiffServ）のマーキングは、ASA を通過するすべてのトラフィックで維持されます。ASA は、分類されたトラフィックをローカルにマーク/再マークすることはありません。たとえば、すべてのパケットの完全優先転送（EF）DSCP ビットを受け取り、「プライオリティ」処理が必要かどうかを判断し、ASA にそれらのパケットを LLQ に入れさせることができます。

QoS のガイドライン

コンテキスト モードのガイドライン

シングル コンテキスト モードでだけサポートされます。マルチ コンテキスト モードをサポートしません。

ファイアウォール モードのガイドライン

ルーテッド ファイアウォール モードでだけサポートされています。トランスペアレント ファイアウォール モードはサポートされません。

IPv6 のガイドライン

IPv6 はサポートされません。

その他のガイドラインと制限事項

- QoS は単方向に適用されます。ポリシー マップを適用するインターフェイスに出入りする（QoS 機能によって異なります）トラフィックだけが影響を受けます。
- プライオリティ トラフィックに対しては、**class-default** クラスマップは使用できません。
- プライオリティキューイングの場合、プライオリティキューは物理インターフェイス用に設定する必要があります。
- ポリシングでは、to-the-box トラフィックはサポートされません。
- ポリシングでは、VPN トンネルとの間で送受信されるトラフィックはインターフェイスのポリシングをバイパスします。
- ポリシングでは、トンネル グループ クラス マップを照合する場合、出力ポリシングのみがサポートされます。

QoS の設定

ASA に QoS を実装するには、次の手順を使用します。

手順

ステップ 1 プライオリティ キューのキューおよび TX リング制限の決定 （4 ページ）。
ステップ 2 インターフェイスのプライオリティ キューの設定 （6 ページ）。
ステップ 3 プライオリティ キューイングとポリシング用のサービス ルールの設定 （7 ページ）。

プライオリティ キューのキューおよび TX リング制限の決定

プライオリティ キューおよび TX リング制限を決定するには、次のワークシートを使用します。

キュー制限のワークシート

次のワークシートは、プライオリティ キューのサイズを計算する方法を示しています。キューは無限大ではないため、いっぱいになってオーバーフローすることがあります。キューがいっぱいになると、以降のパケットはキューに入ることができず、すべてドロップされます（テール ドロップと呼ばれます）。キューがいっぱいになることを避けるには、インターフェイスのプライオリティ キューの設定 （6 ページ）に従ってキューのバッファ サイズを調節します。

ワークシートに関するヒント:

- アウトバウンド帯域幅：たとえば、DSL のアップリンク速度は 768 Kbps などです。プロバイダーに確認してください。
- 平均パケットサイズ：この値は、コーデックまたはサンプリングサイズから決定します。たとえば、VoIP over VPN の場合は、160 バイトなどを使用します。使用するサイズがわからない場合は、256 バイトにすることをお勧めします。
- 遅延：遅延はアプリケーションによって決まります。たとえば、VoIP の場合の推奨される最大遅延は 200 ミリ秒です。使用する遅延がわからない場合は、500 ミリ秒にすることをお勧めします。

表 *1*：キュー制限のワークシート

1	_____ アウトバウンド帯域幅（*Mbps* または *Kbps*）	Mbps	×	**125**	=	_____ バイト数/ミリ秒		
		Kbps	×	**.125**	=	_____ バイト数/ミリ秒		
2	_____ ステップ *1* からのバイト数/ミリ秒		÷	_____ 平均パケットサイズ（バイト）	×	_____ 遅延（ミリ秒）	=	_____ キュー制限（パケット数）

TX リング制限のワークシート

次のワークシートは、TX リング制限の計算方法を示しています。この制限により、イーサネット送信ドライバが受け入れるパケットの最大数が決まります。この制限に達すると、ドライバはパケットをインターフェイスのキューに差し戻し、輻輳が解消されるまでパケットをバッファに格納できるようにします。この設定により、ハードウェアベースの送信リングがプライオリティの高いパケットに対して制限以上の余分な遅延を発生させないことが保証されます。

ワークシートに関するヒント:

- アウトバウンド帯域幅：たとえば、DSL のアップリンク速度は 768 Kbps などです。プロバイダーに確認してください。
- 最大パケットサイズ：通常、最大サイズは 1538 バイト、またはタグ付きイーサネットの場合は 1542 バイトです。ジャンボフレームを許可する場合（プラットフォームでサポートされている場合）、パケットサイズはさらに大きくなる場合があります。
- 遅延：遅延はアプリケーションによって決まります。たとえば、VoIP のジッタを制御するには、20 ミリ秒を使用します。

表 *2: TX* リング制限のワークシート

1	_____ アウトバウンド帯域幅（*Mbps* または *Kbps*）	Mbps	×	**125**	=	_____ バイト数/ミリ秒		
		Kbps	×	**0.125**	=	_____ バイト数/ミリ秒		
2	_____ ステップ *1* からのバイト数/ミリ秒		÷	_____ 最大パケット サイズ（バイト）	×	_____ 遅延（ミリ秒）	=	_____ *TX* リング制限（パケット数）

インターフェイスのプライオリティ キューの設定

物理インターフェイスでトラフィックに対するプライオリティ キューイングをイネーブルにする場合は、各インターフェイスでプライオリティ キューを作成する必要もあります。各物理インターフェイスは、プライオリティ トラフィック用と、他のすべてのトラフィック用に、2 つのキューを使用します。他のトラフィックについては、必要に応じてポリシングを設定できます。

手順

ステップ 1 **[Configuration]** > **[Device Management]** > **[Advanced]** > **[Priority Queue]** を選択して、[Add] をクリックします。

ステップ 2 次のオプションを設定します。

- [Interface]：プライオリティ キューをイネーブルにする物理インターフェイスの名前、または ASASM の場合は VLAN インターフェイス名。
- **キュー制限**：指定したインターフェイスが 500 ミリ秒の間隔で送信できる平均 256 バイトの パケットの数。指定できる値の範囲は 0 ～ 2048 で、2048 がデフォルトです。

  ネットワーク ノードに 500 ミリ秒よりも長く留まるパケットは、エンドツーエンド アプリケーションでタイムアウトをトリガーする可能性があります。そのようなパケットは、各ネットワーク ノードで破棄できます。

  キューは無限大ではないため、いっぱいになってオーバーフローすることがあります。キューがいっぱいになると、以降のパケットはキューに入ることができず、すべてドロップされます（テール ドロップと呼ばれます）。キューがいっぱいになることを避けるため、このオプションを使用してキューのバッファ サイズを大きくできます。

  このオプションの範囲の上限値は実行時に動的に決まります。主な決定要素は、キューのサポートに必要となるメモリと、デバイス上で使用可能なメモリの量です。

キューの制限を指定すると、優先順位の高い低遅延のキューとベストエフォート キューの両方に影響が及びます。

- **送信リング制限** ：プライオリティ キューの深さ。これは、指定したインターフェイスが 10 ミリ秒の間隔で送信できる最大 1550 バイトのパケットの数です。指定できる値の範囲は 3 ～ 511 で、511 がデフォルトです。

  この設定により、ハードウェアベースの伝送リングが優先順位の高いパケットに課す余分な遅延が 10 ミリ秒を超えないことが保証されます。

  このオプションは、Ethernet 伝送ドライバに送ることができる低遅延または通常プライオリティのパケットの最大数を設定します。この最大数を超えると、Ethernet 伝送ドライバがインターフェイスのキューにパケットを押し戻し、輻輳が解消されるまでパケットをキューにバッファします。

  値の範囲の上限は、実行時にダイナミックに決定されます。主な決定要素は、キューのサポートに必要となるメモリと、デバイス上で使用可能なメモリの量です。

  伝送リング制限の制限値を指定すると、優先順位の高い低遅延のキューとベストエフォート キューの両方に影響が及びます。

ステップ 3 [OK] をクリックし、さらに [Apply] をクリックします。

プライオリティ キューイングとポリシング用のサービス ルールの設定

同じポリシー マップ内の異なるクラス マップに対し、プライオリティ キューイングとポリシングを設定できます。有効な QoS 設定については、QoS 機能の相互作用のしくみ （3 ページ）を参照してください。

始める前に

- プライオリティ トラフィックに対しては、**class-default** クラス マップは使用できません。
- ポリシングでは、to-the-box トラフィックはサポートされません。
- ポリシングでは、VPN トンネルとの間で送受信されるトラフィックはインターフェイスのポリシングをバイパスします。
- ポリシングでは、トンネル グループ クラス マップを照合する場合、出力ポリシングのみがサポートされます。
- プライオリティ トラフィックの場合は、遅延が問題になるトラフィックだけを指定します。
- ポリシング トラフィックの場合は、他のすべてのトラフィックをポリシングすることも、トラフィックを特定のタイプに制限することもできます。

手順

ステップ 1 **[Configuration]** > **[Firewall]** > **[Service Policy]** を選択して、ルールを開きます。

新しいサービス ポリシー ルールの一部として QoS を設定できます。または、既存のサービス ポリシーを編集することもできます。

ステップ 2 ウィザードで [Rules] ページにウィザードに進み、途中でインターフェイス（またはグローバル）とトラフィック照合基準を選択します。

ポリシング トラフィックに関しては、優先していない全トラフィックをポリシングするように選択するか、トラフィックを一定の種類に制限できます。

ヒント トラフィック照合に ACL を使用する場合、ポリシングは ACL で指定された方向にのみ適用されます。つまり、送信元から宛先に向かうトラフィックがポリシングされ、宛先から送信元に向かうトラフィックはポリシングされません。

ステップ 3 [Rule Actions] ダイアログボックスで、[QoS] タブをクリックします。

ステップ 4 [Enable priority for this flow] を選択します。

このサービス ポリシー ルールが個別のインターフェイス用の場合、ASDM は自動的にこのインターフェイス用のプライオリティ キューを作成します（[Configuration] > [Device Management] > [Advanced] > [Priority Queue]。詳細については、インターフェイスのプライオリティ キューの設定 （6 ページ）を参照してください）。このルールがグローバル ポリシー用の場合は、サービス ポリシー ルールを設定する前に、1 つ以上のインターフェイスにプライオリティ キューを手動で追加する必要があります。

ステップ 5 指定したタイプのトラフィック ポリシングをイネーブルにするには、[Enable policing] を選択して、[Input policing] または [Output policing]（または両方の）チェックボックスをオンにします。トラフィック ポリシングのタイプごとに、次のオプションを設定します。

- **Committed Rate**：このトラフィック クラスのレート制限（8000 ～ 2000000000 ビット/秒の範囲）。たとえば、トラフィックを 5 Mbps に制限するには、5000000 と入力します。
- **Conform Action**：トラフィックがポリシング レートとバースト サイズを下回った場合に実行するアクション。トラフィックをドロップまたは送信できます。デフォルトでは、トラフィックは送信されます。
- **Exceed Action**：トラフィックがポリシング レートとバースト サイズを上回った場合に実行するアクション。ポリシング レートとバースト サイズを上回ったパケットをドロップまたは送信できます。デフォルトでは、超過パケットはドロップされます。
- **Burst Rate**：適合レート値にスロットリングするまでに、持続したバーストで許可された最大瞬間バイト数（1000 ～ 512000000 バイトの範囲）。バースト サイズはバイト単位の適

合レートの 1/32 として計算されます。たとえば、5 Mbps レートのバーストサイズは 156250 です。デフォルトは 1500 ですが、必要に応じて入力した値が再計算されます。

ステップ 6 [Finish]、[Apply] の順にクリックします。

QoS のモニター

ここでは、QoS をモニターする方法について説明します。

ASDM の QoS をモニターするには、コマンドライン インターフェイス ツールでコマンドを入力します。

QoS ポリシーの統計情報

トラフィック ポリシングの QoS 統計情報を表示するには、**show service-policy police** コマンドを使用します。

```
hostname# show service-policy police

Global policy:
 Service-policy: global_fw_policy

Interface outside:
 Service-policy: qos
  Class-map: browse
   police Interface outside:
    cir 56000 bps, bc 10500 bytes
    conformed 10065 packets, 12621510 bytes; actions: transmit
    exceeded 499 packets, 625146 bytes; actions: drop
    conformed 5600 bps, exceed 5016 bps
  Class-map: cmap2
   police Interface outside:
    cir 200000 bps, bc 37500 bytes
    conformed 17179 packets, 20614800 bytes; actions: transmit
    exceeded 617 packets, 770718 bytes; actions: drop
    conformed 198785 bps, exceed 2303 bps
```

QoS プライオリティの統計情報

priority コマンドを実装するサービス ポリシーの統計情報を表示するには、**show service-policy priority** コマンドを使用します。

```
hostname# show service-policy priority
Global policy:
 Service-policy: global_fw_policy
Interface outside:
 Service-policy: qos
  Class-map: TG1-voice
   Priority:
```

```
Interface outside: aggregate drop 0, aggregate transmit 9383
```

「Aggregate drop」は、このインターフェイスでの合計ドロップ数を示しています。「aggregate transmit」は、このインターフェイスで送信されたパケットの合計数を示しています。

QoS プライオリティ キューの統計情報

インターフェイスのプライオリティ キュー統計情報を表示するには、**show priority-queue statistics** コマンドを使用します。ベストエフォート（BE）キューと低遅延キュー（LLQ）の両方の統計情報が表示されます。次の例に、test という名前のインターフェイスに対する **show priority-queue statistics** コマンドの使用方法を示します。

```
hostname# show priority-queue statistics test

Priority-Queue Statistics interface test

Queue Type         = BE
Packets Dropped    = 0
Packets Transmit   = 0
Packets Enqueued   = 0
Current Q Length   = 0
Max Q Length       = 0

Queue Type         = LLQ
Packets Dropped    = 0
Packets Transmit   = 0
Packets Enqueued   = 0
Current Q Length   = 0
Max Q Length       = 0
hostname#
```

この統計情報レポートの内容は次のとおりです。

- 「Packets Dropped」は、このキューでドロップされたパケットの合計数を示します。
- 「Packets Transmit」は、このキューで送信されたパケットの合計数を示します。
- 「Packets Enqueued」は、このキューでキューイングされたパケットの合計数を示します。
- 「Current Q Length」は、このキューの現在の深さを示します。
- 「Max Q Length」は、このキューで発生した最大の深さを示します。

QoS の履歴

機能名	プラットフォームリリース	説明
プライオリティ キューイングとポリシング	7.0(1)	QoS プライオリティ キューイングとポリシングが導入されました。 次の画面が導入されました。 [Configuration] > [Device Management] > [Advanced] > [Priority Queue] > [Configuration] > [Firewall] > [Service Policy Rules]
シェーピングおよび階層型プライオリティ キューイング	7.2(4)/8.0(4)	QoS シェーピングおよび階層型プライオリティ キューイングが導入されました。 次の画面が変更されました。[Configuration] > [Firewall] > [Service Policy Rules]。
ASA 5585-X での 10 ギガビット イーサネットによる標準プライオリティ キューのサポート	8.2(3)/8.4(1)	ASA 5585-X の 10 ギガビット イーサネット インターフェイスでの標準プライオリティ キューのサポートが追加されました。

翻訳について
このドキュメントは、米国シスコ発行ドキュメントの参考和訳です。リンク情報につきましては、日本語版掲載時点で、英語版にアップデートがあり、リンク先のページが移動/変更されている場合がありますことをご了承ください。あくまでも参考和訳となりますので、正式な内容については米国サイトのドキュメントを参照ください。