ACI でのパケット ドロップ障害の説明

目次

概要

本書では、各障害タイプと、その障害が発生したときの手順について説明します。 シスコ アプリケーション セントリック インフラストラクチャ（ACI）ファブリックの通常操作中に、管理者は特定のタイプのパケット ドロップの障害に遭遇する場合があります。

著者：Cisco TAC エンジニア、Joseph Ristaino、Takuya Kishida

管理対象オブジェクト

Cisco ACI では、すべてのエラーは管理対象オブジェクト（MO）で挙げられます。 たとえば、

障害「*F11245 - 入力のドロップ パケット レート (l2IngrPktsAg15min:dropRate)*」は MO *l2IngrPktsAg15min* の *dropRate* パラメータに関してです。

ここでは**管理対象オブジェクト (MO) 関連のドロップ パケット エラー**について説明します。

	例	説明	サンプ
l2IngrPkts	l2IngrPkts5min l2IngrPkts15min l2IngrPkts1h など...	これは VLAN ごとの各時間内の入力パケットの統計を表します	dropP floodP multic unica
l2IngrPktsAg	l2IngrPktsAg15min l2IngrPktsAg1h l2IngrPktsAg1d など...	これは EPG、BD、VRF 等ごとの入力パケットの統計を表します 例：EPG の統計は EPG に属する VLAN の統計の集約を表します	dropP floodP multic unica
eqptIngrDropPkts	eqptIngrDropPkts15min eqptIngrDropPkts1h eqptIngrDropPkts1d など...	これはインターフェイスごとの各時間内の入力ドロップ パケット統計を表します	*1 for *1 err *1 buf

*1: SUP_REDIRECT での転送ドロップの -EX プラットフォーム制限のため、eqptIngrDropPkts のこれらのカウンタは 1.3(2) のリリース以降使用されていません。

この実装は将来再度変更となる場合もございますので注意してください。

ハードウェア ドロップ カウンタのタイプ

ACI モードで動作する Nexus 9000 スイッチには、ASIC での入力インターフェイスのドロップの原因に関する主要なハードウェア カウンタが 3 つあります。

l2IngrPkts、l2IngrPktsAg の dropRate には、これらのカウンタが含まれます。 eqptIngrDropPkts の上記の表の 3 つのパラメータ (forwardingRate、errorRate、bufferRate) は、それぞれ 3 つのインターフェイス カウンタを表します。

転送

転送ドロップは ASIC の LookUp (LU) ブロックでドロップされるパケットです。 LU ブロックでは、パケット転送の判断は、パケット ヘッダー情報に基づいて行われます。 パケットをドロップする判断の場合、転送ドロップがカウントされます。 これが起きる原因はさまざまですが、主要なものに焦点をあてて説明します。

SECURITY_GROUP_DENY

通信を許可す契約の欠如が原因のドロップ。

パケットがファブリックに入ると、スイッチは送信元と宛先 EPG を参照してこの通信を可能にする契約があるかどうかを確認します。 送信元および宛先が異なる EPG にあり、その間でこのパケット タイプを許可する契約がない場合、スイッチはパケットをドロップし、SECURITY_GROUP_DENY であると分類します。 この場合転送ドロップ カウンタが増えます。

VLAN_XLATE_MISS

不適切な VLAN によるドロップ。

パケットがファブリックに入ると、スイッチはパケットを参照して、このポートの設定でこのパケットの受け入れが可能か判断します。 たとえば、10 の 802.1Q タグ付きファブリックにフレームが入るとします。 スイッチのポートに VLAN 10 があれば、その内容を確認し、宛先 MAC に基づいて転送の判断を行います。 ただし、VLAN 10 がポートにない場合は、ドロップされ VLAN_XLATE_MISS に分類されます。 この場合転送ドロップ カウンタが増えます。

「XLATE」、つまり「変換」となる原因は、ACI ではリーフ スイッチがカプセル化された 802.1Q のフレームを受け入れ、VXLAN もしくはファブリック内でその他の正規化に使用する新しい VLAN に変更するためです。 導入されていない VLAN のフレームが入ると 「変換」は失敗します。

ACL_DROP

sup-tcam によるドロップ。

ACI のスイッチの sup-tcam には、通常の L2/L3 転送の判断に加えて適用する特殊なルールが含まれます。 sup-tcam ルールは組み込み型でユーザ設定はできません。 sup-tcam ルールの目的は主に一部の例外やコントロール プレーン トラフィックを処理することにあり、ユーザがチェックしたりモニタしたりするようには意図されていません。 パケットが sup-tcam ルールに抵触していて、パケットをドロップするルールである場合、ドロップされたパケットは ACL_DROP としてカウントされ、転送ドロップ カウンタでカウントされます。 これが発生すると、通常はパケットが基本的な ACI の転送の原則に反する転送をされようとしていることを意味します。

ドロップの名前が ACL_DROP であっても、この「ACL」はスタンドアロン NX-OS デバイスや他のルーティング/スイッチング デバイスに設定できる通常のアクセス コントロール リストと同じではないことに注意してください。

SUP_REDIRECT

これはドロップではありません。

SUP がリダイレクトされたパケット (つまり CDP/LLDP/UDLD/BFD など) は、パケットが正しく処理されて CPU に転送されていたとしても、転送ドロップとしてカウントされることがあります。

これは N9K-C93180YC-EX など -EX プラットフォームでのみ発生します。 しかしこれは -EX プラットフォームの ASIC の制限によるものなので、「ドロップ」に数えるべきではありません。

エラー

スイッチが無効なフレームを受信すると、エラーとしてドロップされます。 この例として、FCS や CRC エラーのフレームなどがあります。

バッファ

スイッチがフレームを受信し、入出力のいずれかで使用できるバッファ クレジットがない場

合は、フレームは「バッファ」でドロップされます。 これは、ネットワークのどこかで輻輳が発生していることを示唆しています。 エラーを示すリンクが満杯だったり、宛先を含むリンクが輻輳しているかもしれません。

CLI でのドロップ状態表示

管理対象オブジェクト

APIC の 1 つにセキュア シェル（SSH）接続し、次のコマンドを実行します。

```
apic1# moquery - c l2IngrPktsAg15min
```

このコマンドは、このクラス l2IngrPktsAg15min 用にすべてのオブジェクト インスタンスを提供します。

特定のオブジェクトを照会するフィルタの例を示します。 この例では、フィルタは、「tn-TENANT1/ap APP1/epg EPG1」を含む **dn** の属性のオブジェクトのみを示します。

また、この例では必要な属性だけを表示するために egrep を使用しています。

出力例 1： TENANT1、アプリケーション プロファイル APP1、epg EPG1 の EPG カウンタ オブジェクト（l2IngrPktsAg15min）。

```
apic1# moquery -c l2IngrPktsAg15min -f 'l2.IngrPktsAg15min.dn*"tn-TENANT1/ap-APP1/epg-EPG1"' |
egrep 'dn|drop[P,R]|rep'
dn : uni/tn-TENANT1/ap-APP1/epg-EPG1/CDl2IngrPktsAg15min dropPer : 30 <--- number of drop packet
in the current periodic interval (600sec) dropRate : 0.050000 <--- drop packet rate =
dropPer(30) / periodic interval(600s) repIntvEnd : 2017-03-03T15:39:59.181-08:00 <--- periodic
interval = repIntvEnd - repIntvStart repIntvStart : 2017-03-03T15:29:58.016-08:00 = 15:39 -
15:29
= 10 min = 600 sec
```

または、オブジェクト dn がわかっている場合、**-c** の代わりにオプション **-d** を使用して特定のオブジェクトを取得することができます。

出力例 2： TENANT1、アプリケーション プロファイル APP1、epg EPG 2 の EPG カウンタ オブジェクト（l2IngrPktsAg15min）。

```
apic1# moquery -d uni/tn-TENANT1/ap-APP1/epg-EPG2/CDl2IngrPktsAg15min | egrep 'dn|drop[P,R]|rep'
dn               : uni/tn-jw1/BD-jw1/CDl2IngrPktsAg15min
dropPer          : 30
dropRate         : 0.050000
repIntvEnd       : 2017-03-03T15:54:58.021-08:00
repIntvStart     : 2017-03-03T15:44:58.020-08:00
```

ハードウェアカウンタ

エラーが表示される場合や、CLI を使用してスイッチポートのパケット ドロップを確認する場合は、一番の方法はハードウェアのプラットフォーム カウンタを表示することです。 一部の例外を除き、ほとんどすべてのカウンタは **show interface** を使用して表示できます。 3 つの主要なドロップの原因はプラットフォーム カウンタを使用してのみ表示できます。 これらを表示するには、次の手順を実行します。

リーフ

リーフに SSH 接続し、次のコマンドを実行します。

ACI-LEAF#**vsh_lc**
module-1# **show platform internal counters port <X>**
** X はポート番号を表します

イーサネット 1/31 の出力例：

```
ACI-LEAF# vsh_lc
vsh_lc
module-1#
module-1# show platform internal counters port 31
Stats for port 31
(note: forward drops includes sup redirected packets too)
IF          LPort                    Input                        Output
                                Packets       Bytes          Packets    Bytes
eth-1/31    31  Total          400719   286628225      2302918    463380330
                Unicast        306610   269471065       453831     40294786
                Multicast           0           0      1849091    423087288
                Flood           56783     8427482            0            0
              Total Drops       37327                        0
                Buffer              0                        0
                Error               0                        0
                Forward         37327
                LB                  0
                AFD RED                                      0
                    ----- snip -----
```

スパイン

ボックス型スパイン（N9K-C9336PQ）は、リーフとまったく同じです。

モジュラ スパイン用（N9K-C9504 など）,の場合、プラットフォーム カウンタを表示する前に特定のラインカードを接続する必要があります。 スパインに SSH 接続して、次のコマンドを実行します。

ACI-SPINE#**vsh**

ACI-SPINE# **attach module <X>**

module-2# **show platform internal counters port <Y>**.

** X は、表示したいラインカードのモジュール番号を表します

Y はポート番号を表します

イーサネット 2/1 の出力例：

```
ACI-SPINE# vsh
Cisco iNX-OS Debug Shell
This shell should only be used for internal commands and exists
for legacy reasons. User should use ibash infrastructure as this
will be deprecated.
```

```
ACI-SPINE#
ACI-SPINE# attach module 2
Attaching to module 2 ...
To exit type 'exit', to abort type '$.'
Last login: Mon Feb 27 18:47:13 UTC 2017 from sup01-ins on pts/1
No directory, logging in with HOME=/
Bad terminal type: "xterm-256color". Will assume vt100.
module-2#
module-2# show platform internal counters port 1
Stats for port 1
(note: forward drops includes sup redirected packets too)
IF          LPort                    Input                        Output
                                Packets       Bytes          Packets    Bytes
eth-2/1      1  Total        85632884  32811563575     126611414   25868913406
                Unicast      81449096  32273734109     104024872   23037696345
                Multicast     3759719    487617769      22586542   2831217061
                Flood               0            0             0            0
              Total Drops           0                          0
                Buffer              0                          0
                Error               0                          0
                Forward             0
                LB                  0
                AFD RED                                        0
                     ----- snip -----
```

障害

F11245 - 入力のドロップ パケット レート (l2IngrPktsAg15min:dropRate)

説明：

このエラーはレイヤ 2 パケットが「転送ドロップ」を理由にドロップされたときに増加します。 さまざまな理由があるため、

最も一般的な原因について説明します。

N9K-C93180YC-EX などの -EX プラットフォームでは CPU にリダイレクトする必要がある L2 パケットの制限があり (つまり CDP/LLDP/UDLD/BFD など)、「転送ドロップ」として記録され、同時に CPU にコピーされます。 これは Nexus 9000 の EX モデルで使用される ASIC の制限が原因です。

このため、コントロール プレーン プロトコルがインターフェイスで多数有効にされていると、これらのエラーが発生する可能性があります。

解決策：

サービスへの影響がないため、推奨されるベスト プラクティスは「**統計情報のしきい値**」のセクションの説明のとおり、エラーのしきい値を大きくすることです。 方法については「統計情報のしきい値」の手順を参照してください。

F100264 - 入力バッファのドロップ パケット レート (eqptIngrDropPkts5min:bufferRate)

説明：

このエラーはバッファの理由でパケットがポートにドロップされているときに増加します。前述のように、通常これはインターフェイスの輻輳が出入力どちらかの方向で起こっているときに発生します。

解決策：

このエラーは、輻輳による環境で実際にドロップされたパケットを表します。 ドロップされたパケットは ACI ファブリックで稼働するアプリケーションの問題を引き起こしている可能性があります。 ネットワーク管理者はパケット フローを分離し、輻輳が予想外のトラフィック フロー、非効率なロード バランシングなど、 あるいはこれらのポートの想定内の使用によるものであるかどうかを判断する必要があります。

F100696 - 入力転送のパケット ドロップ（ eqptIngrDropPkts5min:forwardingRate ）

注: バージョン 1.3(2) 以降、転送ドロップは eqptIngrDropPkts5min オブジェクトから削除されているため、この問題についてはこのエラーは発生しないはずです。

このエラーの発生シナリオはいくつかあります。 最も一般的なものは以下です。

説明 1) スパイン ドロップ

ARP または IP パケットがプロキシ ルックアップのスパインに転送され、エンドポイントがファブリックで不明の場合、特別な収集パケットが生成され、適切な BD の マルチキャスト グループ アドレスのすべてのリーフに送信されます。 これは、ブリッジ ドメイン (BD) の各リーフからエンドポイントを検出するための ARP 要求をトリガーします。 制約があるため、リーフで受信された収集パケットもファブリックに反映され、スパイン リンクでの転送ドロップをトリガーします。 転送ドロップは、Generation 1 スパイン ハードウェアでのみ増加します。

解決策 1)

問題は ACI ファブリックに未知のユニキャスト トラフィックを送信するデバイスに起因することはわかっているので、どのデバイスがこれを引き起こしているか分析してそれを防止できるかどうか検討する必要があります。 通常はモニタリングの目的でサブネットの IP アドレスをスキャンまたはプローブするデバイスによって引き起こされています。 どの IP がこのトラフィックを送信しているのかを調べるには、エラーを示すスパイン インターフェイスに接続されたリーフに SSH 接続します。

次に、以下のコマンドを実行して収集パケットをトリガーしている送信元 IP アドレス（SIP）を確認します。

```
ACI-LEAF# show ip arp internal event-history event | grep glean | grep sip | more
    [116] TID 11304:arp_handle_inband_glean:3035: log_collect_arp_glean;sip = 192.168.21.150;dip
= 192.168.20.100;info = Rece
ived glean packet is an IP packet
    [116] TID 11304:arp_handle_inband_glean:3035: log_collect_arp_glean;sip = 192.168.21.150;dip
```

```
= 192.168.20.100;info = Rece
ived glean packet is an IP packet
    [116] TID 11304:arp_handle_inband_glean:3035: log_collect_arp_glean;sip = 192.168.21.150;dip
= 192.168.20.100;info = Rece
ived glean packet is an IP packet
    [116] TID 11304:arp_handle_inband_glean:3035: log_collect_arp_glean;sip = 192.168.21.150;dip
= 192.168.20.100;info = Rece
ived glean packet is an IP packet
```

そこから、「192.168.21.150 」がファブリックにこのトラフィックを送信している理由を調査し、それを軽減できるかどうか検討することができます。

説明 2) リーフ ドロップ

このエラーがリーフ インターフェイスで発生する場合、考えられる原因は前述の SECURITY_GROUP_DENY のドロップです。

解決策 2)

リーフには、契約違反が原因で却下されるパケットのログが記録されます。CPU リソースを保護するため、このログはすべてを記録するわけではありませんが、それでも大量のログを提供します。

必要なログを取得するには、エラーが起こっているインターフェイス エラーがポートチャネルの一部である場合は、ポートチャネルに以下のコマンドと grep を使用する必要があります。 そうでない場合は物理インターフェイスを使用できます。

契約のドロップ量に応じてこのログはすぐにロールオーバーできます。

```
ACI-LEAF# show logging ip access-list internal packet-log deny | grep port-channel2 | more
[ Sun Feb 19 14:16:12 2017 503637 usecs]: CName: jr:sb(VXLAN: 2129921), VlanType: FD_VLAN, Vlan-
Id: 59, SMac: 0x8c604
f0288fc, DMac:0x0022bdf819ff, SIP: 192.168.21.150, DIP: 192.168.20.3, SPort: 0, DPort: 0, Src
Intf: port-channel2, Pr
oto: 1, PktLen: 98
[ Sun Feb 19 14:16:12 2017 502547 usecs]: CName: jr:sb(VXLAN: 2129921), VlanType: FD_VLAN, Vlan-
Id: 59, SMac: 0x8c604
f0288fc, DMac:0x0022bdf819ff, SIP: 192.168.21.150, DIP: 192.168.20.3, SPort: 0, DPort: 0, Src
Intf: port-channel2, Pr
oto: 1, PktLen: 98
[ Sun Feb 19 14:16:12 2017 500387 usecs]: CName: jr:sb(VXLAN: 2129921), VlanType: FD_VLAN, Vlan-
Id: 59, SMac: 0x8c604
f0288fc, DMac:0x0022bdf819ff, SIP: 192.168.21.150, DIP: 192.168.20.3, SPort: 0, DPort: 0, Src
Intf: port-channel2, Pr
oto: 1, PktLen: 98
[ Sun Feb 19 14:16:12 2017 499779 usecs]: CName: jr:sb(VXLAN: 2129921), VlanType: FD_VLAN, Vlan-
Id: 59, SMac: 0x8c604
f0288fc, DMac:0x0022bdf819ff, SIP: 192.168.21.150, DIP: 192.168.20.3, SPort: 0, DPort: 0, Src
Intf: port-channel2, Pr
oto: 1, PktLen: 98
[ Sun Feb 19 14:16:12 2017 499624 usecs]: CName: jr:sb(VXLAN: 2129921), VlanType: FD_VLAN, Vlan-
Id: 59, SMac: 0x8c604
f0288fc, DMac:0x0022bdf819ff, SIP: 192.168.21.150, DIP: 192.168.20.3, SPort: 0, DPort: 0, Src
Intf: port-channel2, Pr
oto: 1, PktLen: 98
```

この場合では、「192.168.21.150」は ICMP メッセージ (IP protocol number 1) を「192.168.20.3」に送信しようとしています。 ただし、2 つの EPG 間に ICMP を可能に

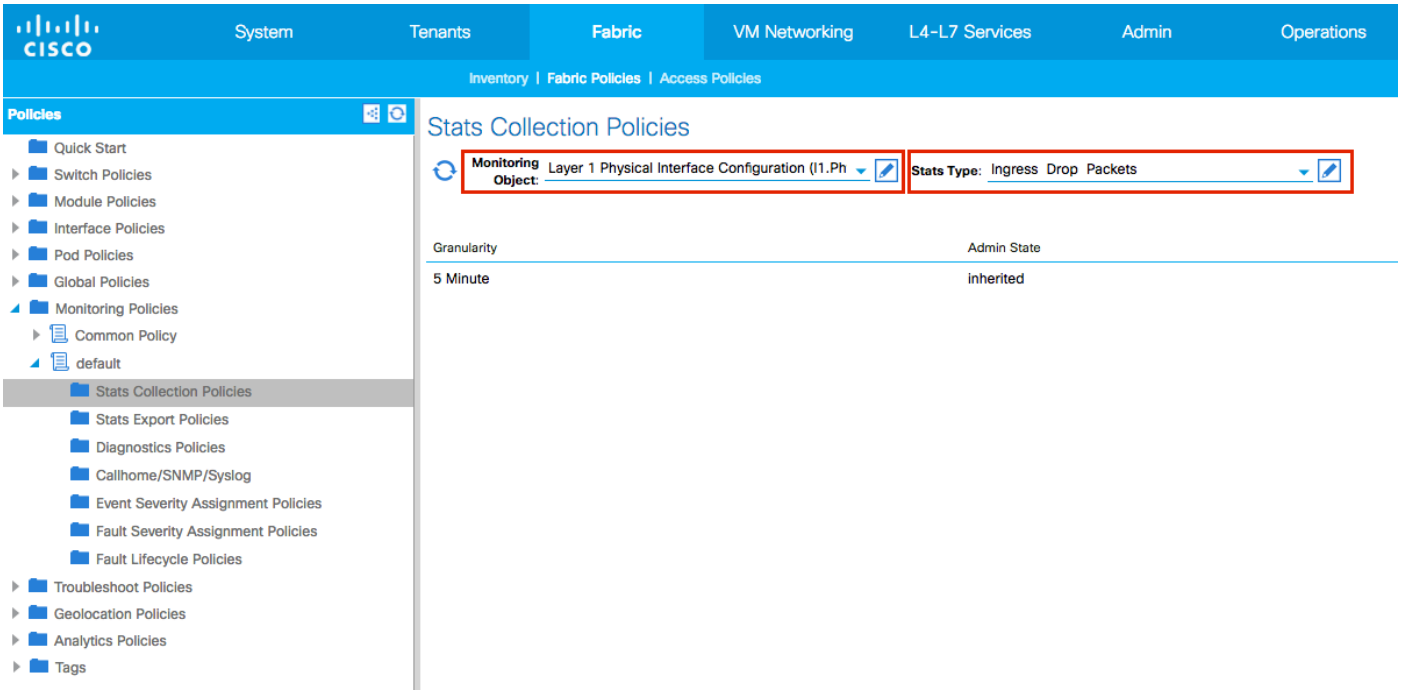
する契約がないため、パケットはドロップされます。 ICMP が許可されるべきである場合、2 つの EPG 間に契約を追加できます。

統計情報のしきい値

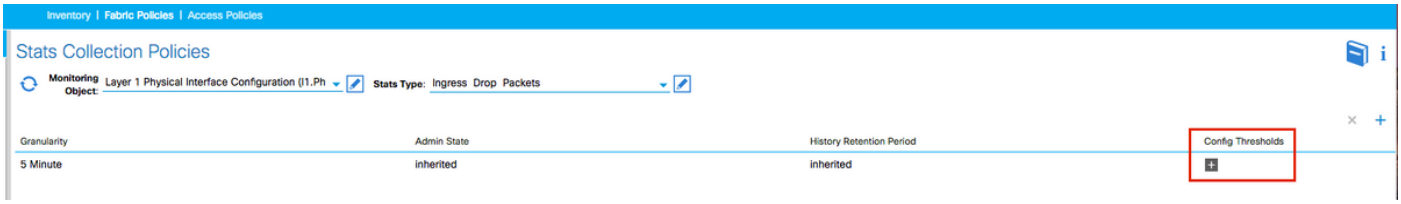
このセクションは、潜在的にドロップ カウンタのエラーを引き起こす可能性のある統計情報オブジェクトのしきい値を変更する方法を説明します。

次の例は *eqptIngrDropPkts* での転送 *ドロップ*のしきい値を変更します。

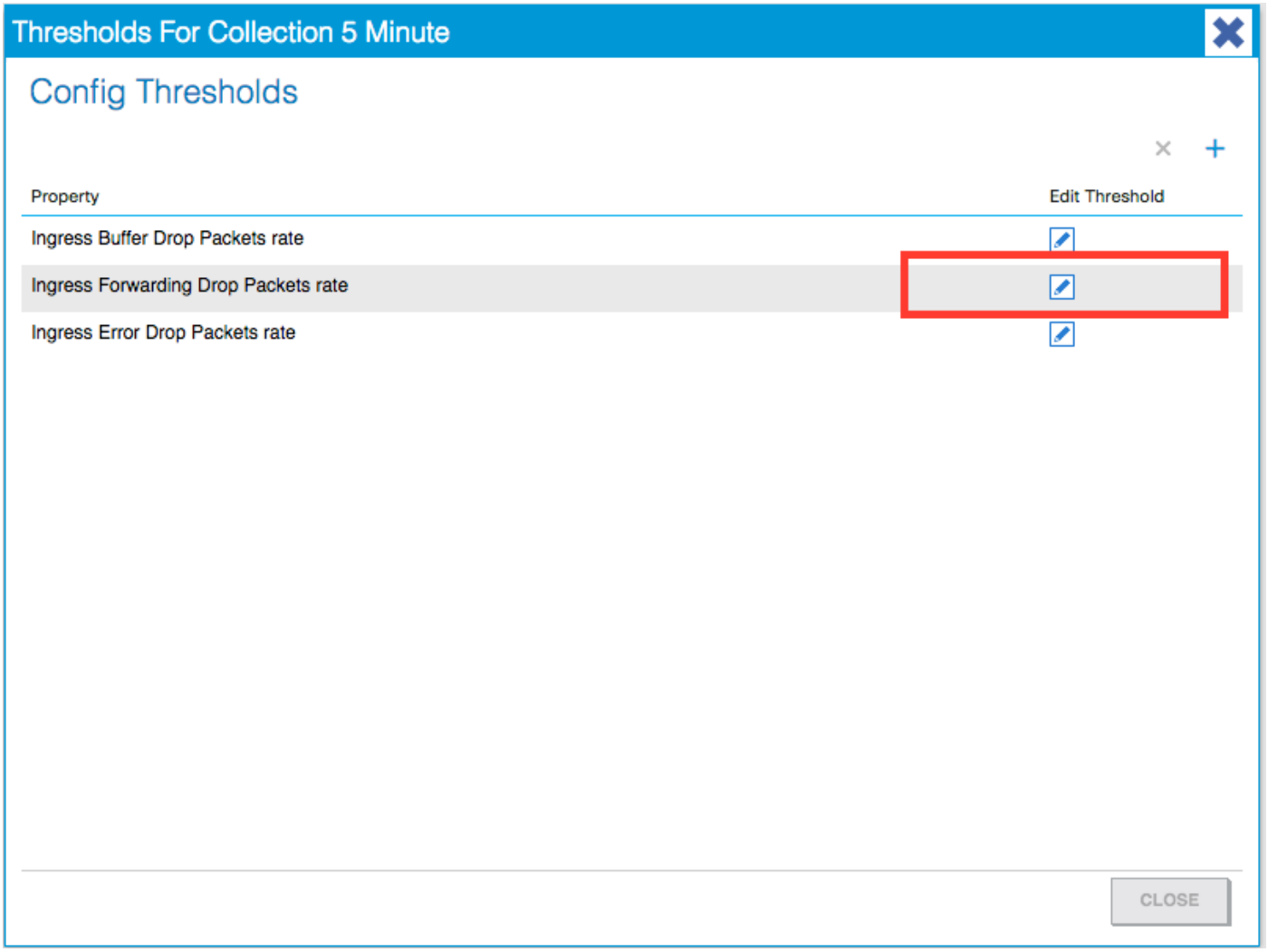
1. [Fabric] > [Fabric Policies] > [Monitoring Policies] > [Default] > [Stats Collection Policies] に移動します。
2. [Monitoring Object] のドロップダウンで、[Layer 1 Physical Interface Configuration (l2.PhysIf)] を選択します。そして、[Stats Type] には、[Ingress Drop Packets] を選択します。



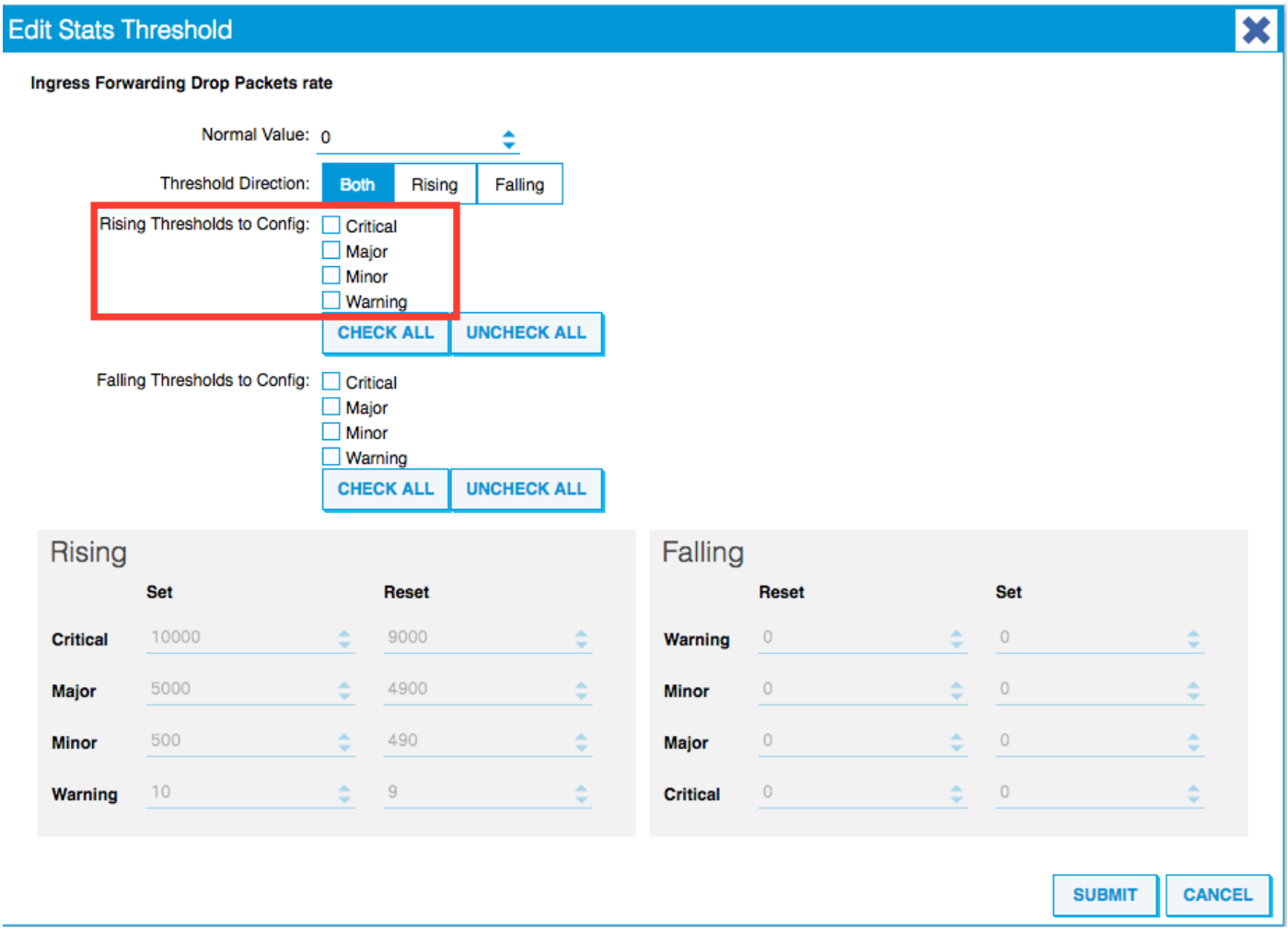
3. [Config Thresholds] の横の [+] をクリックします。



4.バッファドロップのしきい値を編集します。



5. 転送ドロップ レートのメジャー、マイナー、警告の設定の上昇しきい値を無効にすることが推奨されます。


Edit Stats Threshold
Ingress Forwarding Drop Packets rate
Normal Value: 0
Threshold Direction: Both Rising Falling
Rising Thresholds to Config: Critical Major Minor Warning
CHECK ALL UNCHECK ALL
Falling Thresholds to Config: Critical Major Minor Warning
CHECK ALL UNCHECK ALL
Rising
Set Reset
Critical 10000 9000
Major 5000 4900
Minor 500 490
Warning 10 9
Falling
Reset Set
Warning 0 0
Minor 0 0
Major 0 0
Critical 0 0
SUBMIT CANCEL