Why vEdges Unable To Establish IPSec Tunnels If NAT is being Used?

Contents

Introduction

This document describes the problem that may arise when vEdge routers are using IPSec encapsulation for data plane tunnels and one device is behind Network Address Translation (NAT) device doing Symmetric NAT (RFC3489) or Address Dependent Mapping (RFC4787), while another has Direct Internet Access (DIA) or some other type of NAT configured on the transport side interface.

Background information

> **Note**: This article is applicable for vEdge routers only and was written based on behavior seen in vEdge software 18.4.1 and 19.1.0. In newer releases behavior may be different. Please consult with documentation or contact the Cisco Technical Assistance Center (TAC) in case of doubts.

For the purpose of the demonstration, the problem was reproduced in the SD-WAN TAC lab. Devices settings are summarised in the table here:

hostname	site-id	system-ip	private-ip	public-ip
vedge1	232	10.10.10.232	192.168.10.232	198.51.100.232
vedge2	233	10.10.10.233	192.168.9.233	192.168.9.233
vsmart	1	10.10.10.228	192.168.0.228	192.168.0.228
vbond	1	10.10.10.231	192.168.0.231	192.168.0.231

Transport side configuration is quite generic on both devices. This is the configuration of vEdge1:

```
vpn 0
 interface ge0/0
  ip address 192.168.10.232/24
  !
  tunnel-interface
   encapsulation ipsec
   color biz-internet
   no allow-service bgp
   no allow-service dhcp
   allow-service dns
   allow-service icmp
   no allow-service sshd
   no allow-service netconf
   no allow-service ntp
   no allow-service ospf
   no allow-service stun
   allow-service https
  !
  no shutdown
 !
 ip route 0.0.0.0/0 192.168.10.11
!
```

vEdge2:

```
interface ge0/1
  ip address 192.168.9.233/24
  !
  tunnel-interface
   encapsulation ipsec
   color biz-internet
   no allow-service bgp
   no allow-service dhcp
   allow-service dns
   allow-service icmp
   no allow-service sshd
   no allow-service netconf
   no allow-service ntp
   no allow-service ospf
   no allow-service stun
   allow-service https
  !
  no shutdown
 !
 ip route 0.0.0.0/0 192.168.9.1
```

In order to demonstrate the problem in this document, Virtual Adaptive Security Appliance (ASAv) firewall resides between two vEdge routers. ASAv is doing address translations according to these rules:

- If traffic from vEdge1 is intended for controllers, source ports 12346-12426 are translated to 52346-52426
- If traffic from vEdge1 is intended for data plane connections to other sites, source ports 12346-12426 are translated to 42346-42426
- All other traffic from vEdge1 is also mapped to the same public address (198.51.100.232)

This is ASAv NAT configuration for reference:

```
object network VE1
```

```
 host 192.168.10.232
object network CONTROLLERS
 subnet 192.168.0.0 255.255.255.0
object network VE1_NAT
 host 198.51.100.232
object service CONTROL
 service udp source range 12346 12445 destination range 12346 12445
object service CC_NAT_CONTROLLERS
 service udp source range 52346 52445 destination range 12346 12445
object service CC_NAT_OTHER
 service udp source range 42346 42445 destination range 12346 12445
object network ALL
 subnet 0.0.0.0 0.0.0.0
nat (ve1-iface,ve2-iface) source static VE1 VE1_NAT destination static CONTROLLERS CONTROLLERS
service CONTROL CC_NAT_CONTROLLERS
nat (ve1-iface,ve2-iface) source static VE1 VE1_NAT destination static ALL ALL service CONTROL
CC_NAT_OTHER
nat (ve1-iface,ve2-iface) source dynamic VE1 VE1_NAT
```

Problem

Working Scenario

In the normal state, we can observe that data plane tunnels are established, Bidirectional Forwarding Detection (BFD) is in **up** state.

Please notice which public port used on vEdge1 device (52366) to establish control connections with controllers:

```
vEdge1# show control local-properties wan-interface-list

 NAT TYPE: E -- indicates End-point independent mapping
           A -- indicates Address-port dependent mapping
           N -- indicates Not learned
           Note: Requires minimum two vbonds to learn the NAT type

                          PUBLIC           PUBLIC PRIVATE          PRIVATE
PRIVATE                                 MAX   RESTRICT/           LAST          SPI TIME    NAT  VM
INTERFACE                 IPv4             PORT   IPv4             IPv6
PORT    VS/VM COLOR             STATE CNTRL CONTROL/     LR/LB  CONNECTION   REMAINING   TYPE CON

STUN                                              PRF
-------------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------------------
-----------
ge0/0                     198.51.100.232  52366  192.168.10.232  ::
12366    2/1  biz-internet     up     2      no/yes/no   No/No  0:00:00:28   0:11:59:17  N    5
```

On vEdge2 no NAT is being used, hence private address and ports are the same:

```
vEdge2# show control local-properties wan-interface-list

 NAT TYPE: E -- indicates End-point independent mapping
           A -- indicates Address-port dependent mapping
           N -- indicates Not learned
           Note: Requires minimum two vbonds to learn the NAT type

                          PUBLIC           PUBLIC PRIVATE          PRIVATE
```

```
PRIVATE                              MAX   RESTRICT/           LAST         SPI TIME    NAT  VM
INTERFACE               IPv4              PORT   IPv4          IPv6
PORT    VS/VM COLOR            STATE CNTRL CONTROL/     LR/LB  CONNECTION   REMAINING   TYPE CON

STUN                                                PRF
-------------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------------------
-----------
ge0/1                    192.168.9.233   12366  192.168.9.233   ::
12366    2/1  biz-internet     up     2      no/yes/no   No/No  0:00:00:48   0:11:58:53  N    5
```

In the **show tunnel statistics**from vEdge1 we can see tx/rx counters are incrementing:

```
vEdge1# show tunnel statistics dest-ip 192.168.9.233


TCP
TUNNEL                                      SOURCE  DEST
TUNNEL                                              MSS
PROTOCOL  SOURCE IP       DEST IP         PORT    PORT   SYSTEM IP     LOCAL COLOR   REMOTE COLOR
MTU     tx-pkts  tx-octets  rx-pkts  rx-octets  ADJUST
-------------------------------------------------------------------------------------------------
-----------------------------------------------------------
ipsec     192.168.10.232  192.168.9.233  12366   12366  10.10.10.233  biz-internet  biz-internet
1441    223      81163      179      40201      1202
```

From the same output from vEdge2 you can see as well rx/rx packets counters are incrementing. Please notice destination port (42366) is different from port used to establish control connections (52366):

```
vEdge2# show tunnel statistics dest-ip 198.51.100.232


TCP
TUNNEL                                      SOURCE  DEST
TUNNEL                                              MSS
PROTOCOL  SOURCE IP      DEST IP         PORT    PORT   SYSTEM IP     LOCAL COLOR   REMOTE COLOR
MTU     tx-pkts  tx-octets  rx-pkts  rx-octets  ADJUST
-------------------------------------------------------------------------------------------------
-----------------------------------------------------------
ipsec     192.168.9.233  198.51.100.232  12366   42366  10.10.10.232  biz-internet  biz-internet
1441    296      88669      261      44638      1201
```

But BFD sessions are still up on both devices:

```
vEdge1# show bfd sessions site-id 233 | tab

                                     SRC    DST                  SITE
DETECT      TX
SRC IP          DST IP         PROTO  PORT   PORT   SYSTEM IP     ID    LOCAL COLOR   COLOR
STATE  MULTIPLIER  INTERVAL  UPTIME      TRANSITIONS
-------------------------------------------------------------------------------------------------
-----------------------------------------------------------
192.168.10.232  192.168.9.233  ipsec  12366  12366  10.10.10.233  233   biz-internet  biz-
```

```
internet  up     7           1000      0:00:02:42  0



vEdge2# show bfd sessions site-id 232 | tab

                                     SRC    DST                   SITE
DETECT      TX
SRC IP          DST IP          PROTO  PORT   PORT   SYSTEM IP     ID    LOCAL COLOR   COLOR
STATE  MULTIPLIER  INTERVAL  UPTIME      TRANSITIONS
-------------------------------------------------------------------------------------------
-----------------------------------------------------------
192.168.9.233  198.51.100.232  ipsec  12366  52366  10.10.10.232  232   biz-internet  biz-
internet  up     7           1000      0:00:03:00  0
```

Different ports used for control and data plane connections does not cause any issues, connectivity is in place.

Failure Scenario

The user wants to enable Direct Internet Access (DIA) on vEdge2 router. In order to do so, this configuration was applied to vEdge2:

```
vpn 0
 interface ge0/1
  nat
   respond-to-ping
  !
 !
!
vpn 1
 ip route 0.0.0.0/0 vpn 0
!
```

And BFD session went down unexpectedly and moreover stays in the downstate. After clearing tunnel statistics you can see that RX counter does not increase in the **show tunnel statistics** output:

```
vEdge2# show tunnel statistics dest-ip 198.51.100.232


TCP
TUNNEL                                  SOURCE  DEST
TUNNEL                                          MSS
PROTOCOL  SOURCE IP      DEST IP         PORT    PORT   SYSTEM IP     LOCAL COLOR   REMOTE COLOR
MTU     tx-pkts  tx-octets  rx-pkts  rx-octets  ADJUST
-------------------------------------------------------------------------------------------
-----------------------------------------------------------
ipsec     192.168.9.233  198.51.100.232  12346   52366  10.10.10.232  biz-internet  biz-internet
1442    282      48222      0        0          1368

vEdge2# show bfd sessions site-id 232
                                     SOURCE TLOC      REMOTE TLOC
DST PUBLIC                       DST PUBLIC         DETECT      TX
SYSTEM IP        SITE ID  STATE       COLOR            COLOR           SOURCE IP
IP                               PORT        ENCAP  MULTIPLIER  INTERVAL(msec) UPTIME
TRANSITIONS
```

```
------------------------------------------------------------------------------------------------
------------------------------------------------------------------------------------------------
-------------
10.10.10.232     232      down        biz-internet     biz-internet     192.168.9.233
198.51.100.232                  52366       ipsec  7           1000           NA              0

vEdge2# show tunnel statistics dest-ip 198.51.100.232


TCP
TUNNEL                                    SOURCE  DEST
TUNNEL                                            MSS
PROTOCOL  SOURCE IP      DEST IP          PORT    PORT   SYSTEM IP     LOCAL COLOR   REMOTE COLOR
MTU     tx-pkts  tx-octets  rx-pkts  rx-octets  ADJUST
------------------------------------------------------------------------------------------------
-----------------------------------------------------------
ipsec     192.168.9.233  198.51.100.232  12346   52366  10.10.10.232  biz-internet  biz-internet
1442    285      48735      0        0          1368
```

Initially, customer suspected that problem related to Tunnel MTU. If you compare outputs above with outputs from "Working Scenario" section, you can notice that in working scenario Tunnel MTU is 1441 versus 1442 in the failed scenario. Based on the documentation, Tunnel MTU should be 1442 (1500 default interface MTU - 58 bytes for tunnel overhead), but once BFD is up, Tunnel MTU is lowered by 1 byte. For your reference, outputs from **show tunnel statistics** together with **show tunnel statistics bfd** provided below for case when BFD is in **down** state:

```
vEdge1# show tunnel statistics dest-ip 192.168.9.233 ; show tunnel statistics bfd dest-ip
192.168.9.233


TCP
TUNNEL                                    SOURCE  DEST
TUNNEL                                            MSS
PROTOCOL  SOURCE IP       DEST IP         PORT    PORT   SYSTEM IP     LOCAL COLOR   REMOTE COLOR
MTU     tx-pkts  tx-octets  rx-pkts  rx-octets  ADJUST
------------------------------------------------------------------------------------------------
-----------------------------------------------------------
ipsec     192.168.10.232  192.168.9.233  12346   12346  10.10.10.233  biz-internet  biz-internet
1442    133      22743      0        0          1362


                                                          BFD   BFD   BFD     BFD     BFD   BFD
BFD     BFD
                                                          ECHO  ECHO  ECHO    ECHO    PMTU  PMTU
PMTU    PMTU
TUNNEL                                    SOURCE  DEST    TX    RX    TX      RX      TX    RX
TX      RX
PROTOCOL  SOURCE IP       DEST IP         PORT    PORT    PKTS  PKTS  OCTETS  OCTETS  PKTS  PKTS
OCTETS  OCTETS
------------------------------------------------------------------------------------------------
----------------
ipsec     192.168.10.232  192.168.9.233  12346   12346  133   0     22743   0       0     0
0       0

vEdge1# show tunnel statistics dest-ip 192.168.9.233 ; show tunnel statistics bfd dest-ip
192.168.9.233


TCP
TUNNEL                                    SOURCE  DEST
```

```
TUNNEL                                                MSS
PROTOCOL  SOURCE IP       DEST IP         PORT    PORT   SYSTEM IP     LOCAL COLOR   REMOTE COLOR
MTU     tx-pkts  tx-octets  rx-pkts  rx-octets  ADJUST
-------------------------------------------------------------------------------------------------
-------------------------------------------------------
ipsec     192.168.10.232  192.168.9.233  12346   12346  10.10.10.233  biz-internet  biz-internet
1442    134      22914      0        0          1362


                                                      BFD   BFD   BFD     BFD     BFD   BFD
BFD     BFD
                                                      ECHO  ECHO  ECHO    ECHO    PMTU  PMTU
PMTU    PMTU
TUNNEL                                  SOURCE  DEST  TX    RX    TX      RX      TX    RX
TX      RX
PROTOCOL  SOURCE IP       DEST IP         PORT    PORT   PKTS  PKTS  OCTETS  OCTETS  PKTS  PKTS
OCTETS  OCTETS
-------------------------------------------------------------------------------------------------
----------------
ipsec     192.168.10.232  192.168.9.233  12346   12346  134   0     22914   0       0     0
0       0
```

And if BFD is in up state:

```
vEdge1# show tunnel statistics dest-ip 192.168.9.233 ; show tunnel statistics bfd dest-ip
192.168.9.233 ;


TCP
TUNNEL                                  SOURCE  DEST
TUNNEL                                          MSS
PROTOCOL  SOURCE IP       DEST IP         PORT    PORT   SYSTEM IP     LOCAL COLOR   REMOTE COLOR
MTU     tx-pkts  tx-octets  rx-pkts  rx-octets  ADJUST
-------------------------------------------------------------------------------------------------
-------------------------------------------------------
ipsec     192.168.10.232  192.168.9.233  12346   12346  10.10.10.233  biz-internet  biz-internet
1441    3541     610133     3504     592907     1361


                                                      BFD   BFD   BFD     BFD     BFD   BFD
BFD     BFD
                                                      ECHO  ECHO  ECHO    ECHO    PMTU  PMTU
PMTU    PMTU
TUNNEL                                  SOURCE  DEST  TX    RX    TX      RX      TX    RX
TX      RX
PROTOCOL  SOURCE IP       DEST IP         PORT    PORT   PKTS  PKTS  OCTETS  OCTETS  PKTS  PKTS
OCTETS  OCTETS
-------------------------------------------------------------------------------------------------
----------------
ipsec     192.168.10.232  192.168.9.233  12346   12346  3522  3491  589970  584816  19    13
20163   8091

vEdge1# show tunnel statistics dest-ip 192.168.9.233 ; show tunnel statistics bfd dest-ip
192.168.9.233 ;


TCP
TUNNEL                                  SOURCE  DEST
TUNNEL                                          MSS
PROTOCOL  SOURCE IP       DEST IP         PORT    PORT   SYSTEM IP     LOCAL COLOR   REMOTE COLOR
MTU     tx-pkts  tx-octets  rx-pkts  rx-octets  ADJUST
```

```
------------------------------------------------------------------------------------------------------------
-------------------------------------------------------
ipsec     192.168.10.232  192.168.9.233  12346   12346  10.10.10.233  biz-internet  biz-internet
1441    3542     610297     3505     593078     1361


                                                          BFD   BFD   BFD     BFD     BFD   BFD
BFD     BFD
                                                          ECHO  ECHO  ECHO    ECHO    PMTU  PMTU
PMTU    PMTU
TUNNEL                                    SOURCE  DEST    TX    RX    TX      RX      TX    RX
TX      RX
PROTOCOL  SOURCE IP       DEST IP         PORT    PORT    PKTS  PKTS  OCTETS  OCTETS  PKTS  PKTS
OCTETS  OCTETS
------------------------------------------------------------------------------------------------------------
----------------
ipsec     192.168.10.232  192.168.9.233  12346   12346  3523  3492  590134  584987  19    13
20163   8091
```

Note: By the way, we can determine BFD packet size together with encapsulation by looking to outputs above. Note that only one BFD packet was received between two outputs, hence substracting BFD Echo RX Octets value 584987 - 584816 will give us 171-byte result. It can be useful to precisely calculate bandwidth used by BFD itself.

The reason for BFD stuck in **down** state is not MTU, but NAT configuration obviously. This is the only thing changed between **Working scenario** and **Failed scenario**. You can see here that as a result of DIA configuration, NAT static mapping was automatically created by vEdge2 in the translation table to allow data plane IPSec traffic bypass:

```
vEdge2# show ip nat filter nat-vpn 0 nat-ifname ge0/1 vpn 0 protocol udp 192.168.9.233
198.51.100.232

                              PRIVATE                          PRIVATE  PRIVATE
PUBLIC  PUBLIC
NAT  NAT                      SOURCE          PRIVATE DEST     SOURCE   DEST     PUBLIC SOURCE
PUBLIC DEST     SOURCE  DEST    FILTER        IDLE          OUTBOUND  OUTBOUND  INBOUND  INBOUND
VPN  IFNAME  VPN  PROTOCOL  ADDRESS        ADDRESS          PORT     PORT     ADDRESS
ADDRESS         PORT    PORT    STATE         TIMEOUT      PACKETS   OCTETS    PACKETS  OCTETS
DIRECTION
------------------------------------------------------------------------------------------------------------
------------------------------------------------------------------------------------------------------------
------
0    ge0/1   0    udp       192.168.9.233  198.51.100.232  12346    52366    192.168.9.233
198.51.100.232  12346   52366   established  0:00:00:59  53        8321      0        0        -
```

As you can see, port 52366 is being used instead of 42366. This is because vEdge2 expects 52366 port and learned it from OMP TLOCs advertised by vSmart:

```
vEdge2# show omp tlocs ip 10.10.10.232 | b PUBLIC

PUBLIC           PRIVATE
ADDRESS                                                                    PSEUDO
PUBLIC                    PRIVATE  PUBLIC  IPV6    PRIVATE  IPV6     BFD
FAMILY   TLOC IP          COLOR            ENCAP  FROM PEER        STATUS    KEY     PUBLIC IP
PORT    PRIVATE IP       PORT     IPV6    PORT    IPV6     PORT     STATUS
```

```
--------------------------------------------------------------------------------------------------
--------------------------------------------------------------------------------
ipv4     10.10.10.232     biz-internet     ipsec  10.10.10.228     C,I,R     1
198.51.100.232   52366   192.168.10.232   12346    ::      0       ::        0        down
```

Solution

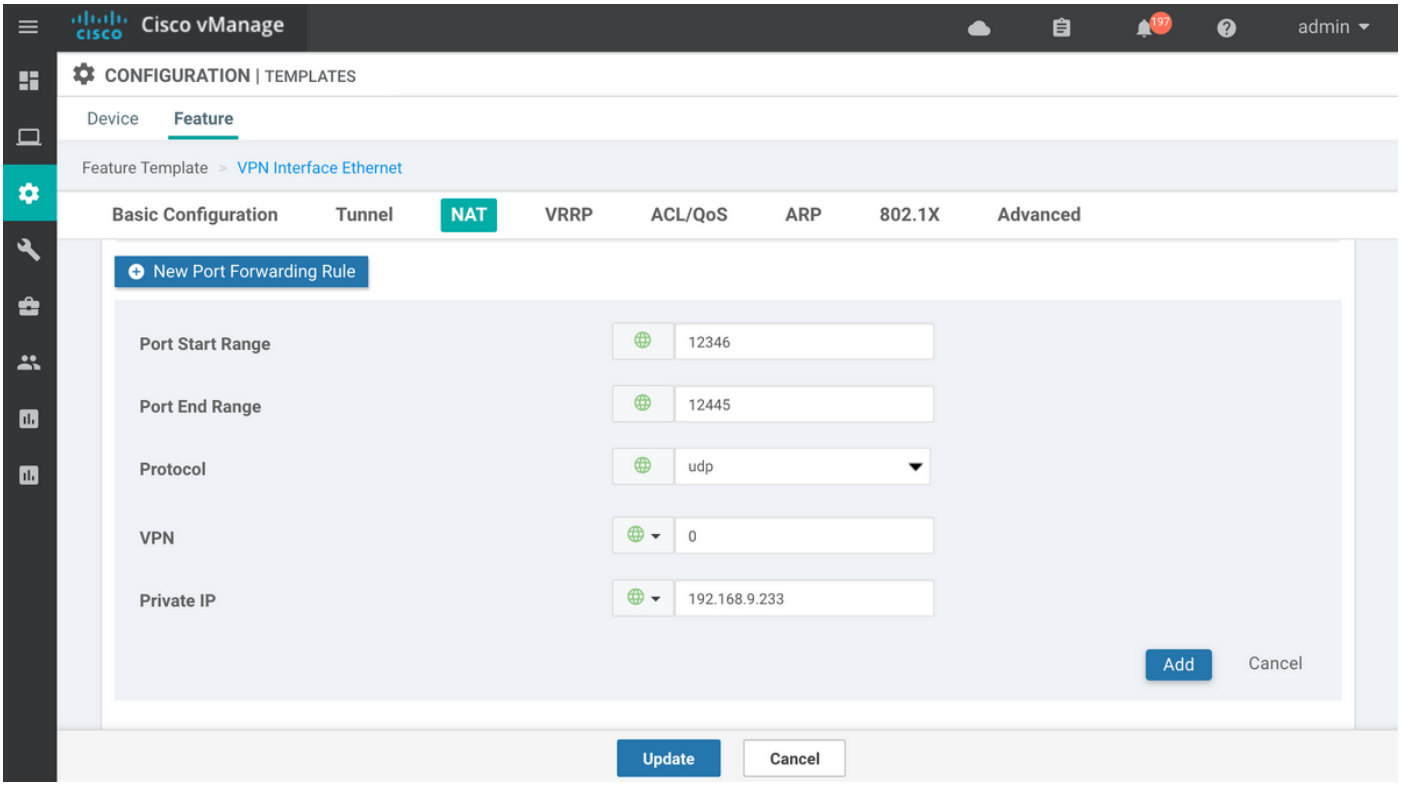
NAT Port-Forward

From first glance, workaround for such type of problems is simple. You can configure static NAT exemption port forwarding on vEdge2 transport interface to bypass filtering for data plane connections from any sources forcefully:

```
vpn 0
 interface ge0/1
  nat
   respond-to-ping
   port-forward port-start 12346 port-end 12445 proto udp
    private-vpn         0
    private-ip-address 192.168.9.233
   !
  !
 !
!
```

Here range 12346 to 12446 accommodate all possible initial ports (12346, 12366, 12386, 12406, and 12426 plus port-offset). For more information on this refer to "Firewall Ports for Viptela Deployments".

If Device Feature Templates are being used instead of CLI template, then to achieve the same, we need to update or add new VPN Ethernet Feature Template for corresponding transport (vpn 0) interface with **New Port Forwarding Rule**, as shown in the image:



Explicit ACL

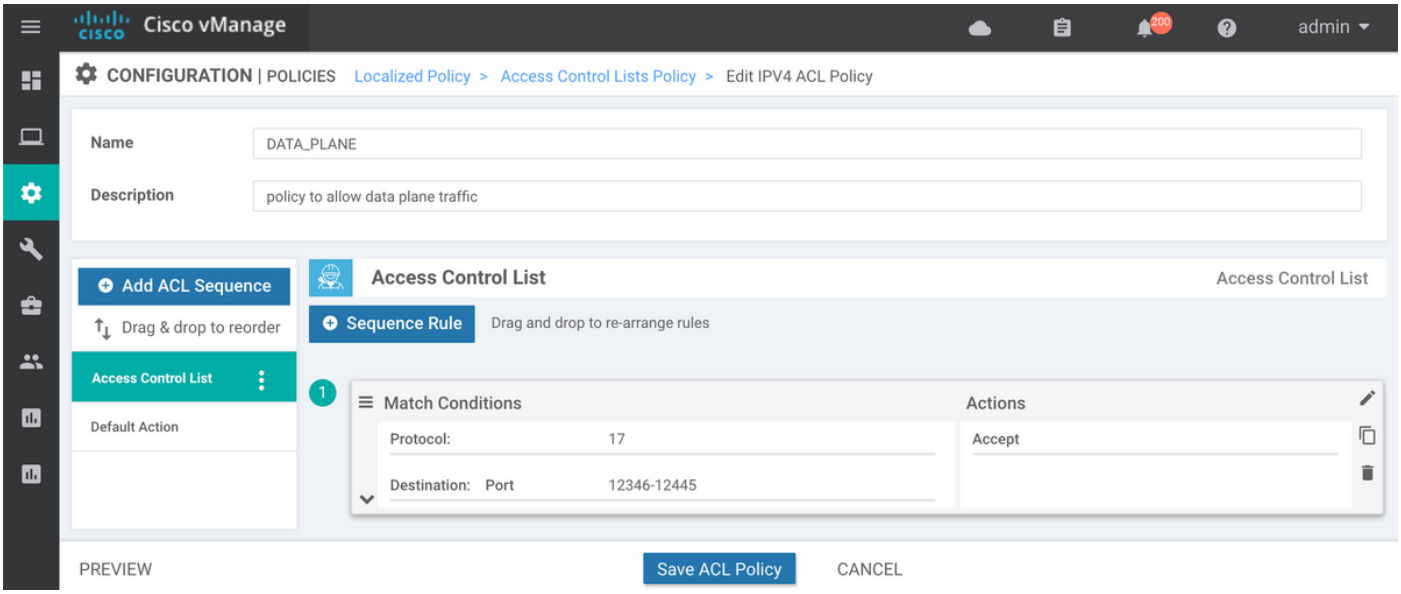
Also, another solution with an explicit ACL is possible. If **implicit-acl-logging** is configured under **policy** section, you may notice the following message in the **/var/log/tmplog/vdebug** file:

```
local7.notice: Jun  8 17:53:29 vEdge2 FTMD[980]: %Viptela-vEdge2-FTMD-5-NTCE-1000026: FLOW LOG
vpn-0 198.51.100.232/42346 192.168.9.233/12346 udp: tos: 192  inbound-acl, Implicit-ACL, Result:
denyPkt count 2: Byte count 342 Ingress-Intf ge0/1 Egress-intf cpu
```

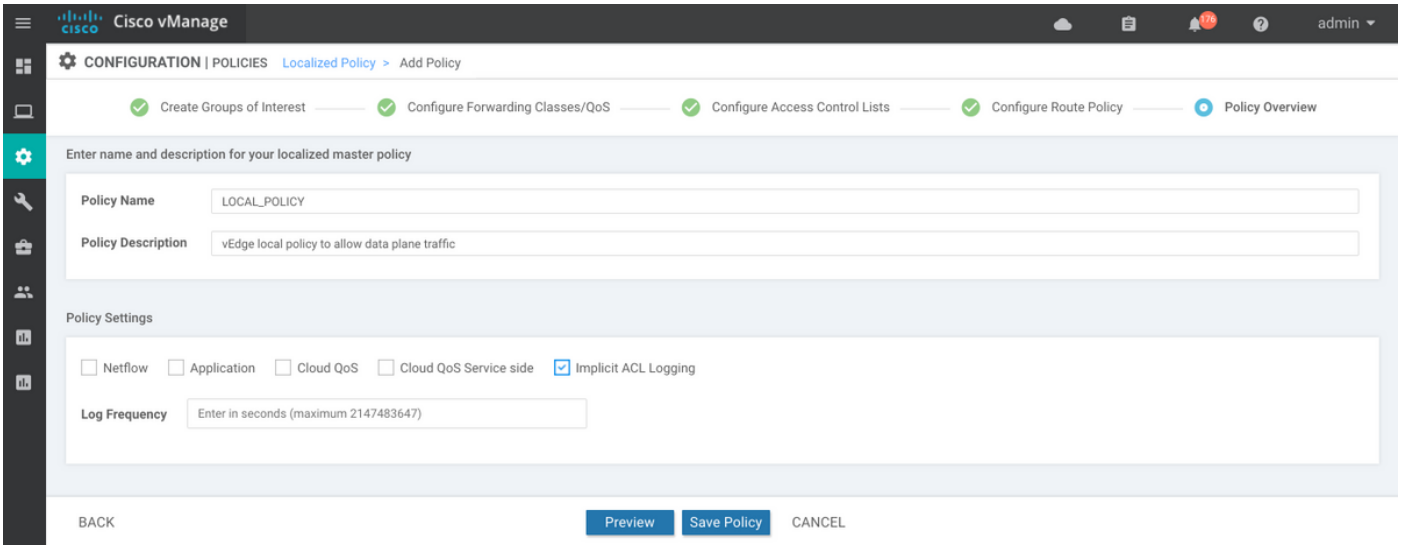
It explains the root cause and hence you need to explicitly allow incoming data plane packets in the Access Control List (ACL) on vEdge2 like this:

```
vpn 0
 interface ge0/1
  ip address 192.168.9.233/24
  nat
   respond-to-ping
  !
  tunnel-interface
   encapsulation ipsec
   color biz-internet
   no allow-service bgp
   no allow-service dhcp
   allow-service dns
   allow-service icmp
   no allow-service sshd
   no allow-service netconf
   no allow-service ntp
   no allow-service ospf
   no allow-service stun
   allow-service https
  !
  mtu      1506
  no shutdown
  access-list DATA_PLANE in
 !
!
policy
 implicit-acl-logging
 access-list DATA_PLANE
  sequence 10
   match
destination-port 12346 12445 protocol 17 ! action accept ! ! default-action drop ! !
```

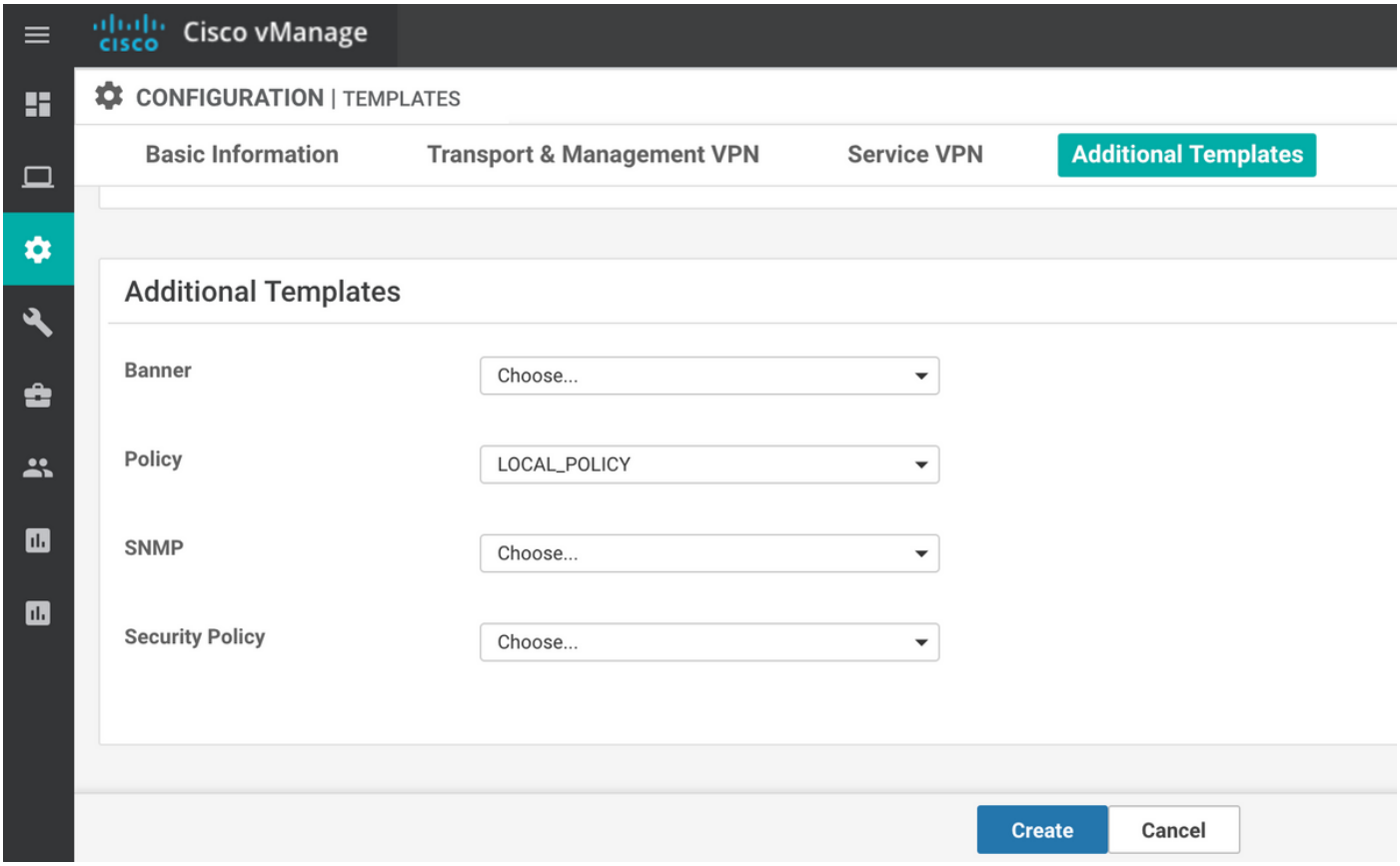
If Device Feature Templates are being used, then you need to create Localized Policy and configure ACL on **Configure Access Control Lists** wizard step:



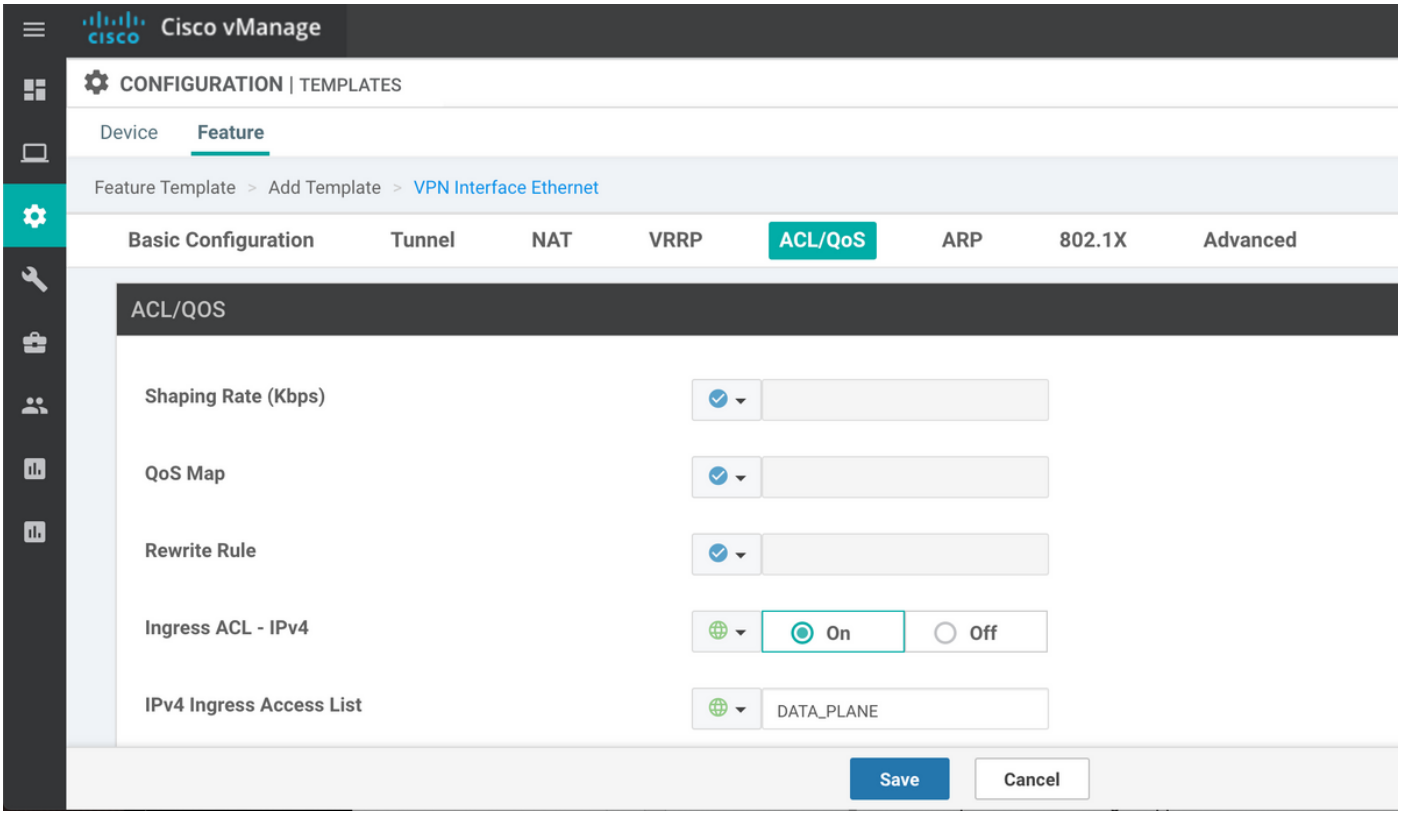
If **implicit-acl-logging** is not yet enabled, it might be a good idea to enable it on the final step before click on **Save Policy** button:



Localized policy (named **LOCAL_POLICY** in our case) should be referenced in the Device Template:



And then ACL (named **DATA_PLANE** in our case) should be applied under VPN Interface Ethernet Feature Template in the ingress (in) direction:



Once ACL is configured and applied to the interface to bypass data plane traffic, BFD session is more to the **up** state again:

```
vEdge2# show tunnel statistics dest-ip 198.51.100.232 ; show bfd sessions site-id 232
```

```
TCP
TUNNEL                                    SOURCE  DEST
TUNNEL                                            MSS
PROTOCOL  SOURCE IP      DEST IP          PORT    PORT   SYSTEM IP     LOCAL COLOR   REMOTE COLOR
MTU     tx-pkts  tx-octets  rx-pkts  rx-octets  ADJUST
-------------------------------------------------------------------------------------------------
-------------------------------------------------------------
ipsec     192.168.9.233  198.51.100.232  12346   42346  10.10.10.232  biz-internet  biz-internet
1441    1768     304503     1768     304433     1361

                                       SOURCE TLOC      REMOTE TLOC
DST PUBLIC                        DST PUBLIC          DETECT      TX
SYSTEM IP        SITE ID  STATE       COLOR            COLOR            SOURCE IP
IP                                PORT        ENCAP  MULTIPLIER  INTERVAL(msec) UPTIME
TRANSITIONS
-------------------------------------------------------------------------------------------------
-------------------------------------------------------------------------------------------------
-------------
10.10.10.232     232      up          biz-internet     biz-internet     192.168.9.233
198.51.100.232                   52346       ipsec  7           1000           0:00:14:36      0
```

Other Considerations

Please note that workaround with ACL is much more practical than NAT port-forwarding because you may also match based on source addresses of the remote site for greater security and to protect from DDoS attacks to your device, e.g:

```
 access-list DATA_PLANE
  sequence 10
   match
    source-ip        198.51.100.232/32
    destination-port 12346 12445
    protocol         17
   !
   action accept
   !
  !
```

Also please note that for any other incoming traffic (not specified with **allowed-services**) e.g. for default **iperf** port 5001 explicit ACL **seq 20** like in this example this won't make any effect as opposed to data plane traffic:

```
policy
 access-list DATA_PLANE
  sequence 10
   match
    source-ip        198.51.100.232/32
    destination-port 12346 12445
    protocol         17
   !
   action accept
   !
  !
  sequence 20
   match
    destination-port 5001
    protocol         6
```

```
   !
   action accept
   !
  !
```

And you still need NAT port-forward exemption rule for **iperf** to work:

```
vEdgeCloud2# show running-config vpn 0 interface ge0/1 nat
vpn 0
 interface ge0/1
  nat
   respond-to-ping
   port-forward port-start 5001 port-end 5001 proto tcp
    private-vpn        0
    private-ip-address 192.168.9.233
   !
  !
 !
!
```

Conclusion

This is expected behavior on vEdge routers caused by NAT software design specifics and can't be avoided.