pfSense - Bug #10176

Multiple overlapping phase 2 SAs on VTI tunnels

01/10/2020 04:27 AM - Brian Candler

Status:	Feedback	**Start date:**	01/10/2020
Priority:	Normal	**Due date:**	
Assignee:	Jim Pingle	**% Done:**	100%
Category:	IPsec	**Estimated time:**	0.00 hour
Target version:	2.5.0		
Affected Version:		**Affected Architecture:**	

Description

This might be a configuration error, but if so, I can't see it. The problem occurs with VTI tunnels between:

- "A end": a HA pair of XG-1537 (2.4.4p3)

and two different "B ends" which are single (non-HA) pfSense boxes:

- (B1) a Dell R220 running 2.4.4p3 (this is "con1000" from the point of view of the A end, and "con4000" from the B1 end)
- (B2) a SG-1000 running 2.4.4p3 (this is "con2000" from the point of view of the A end, and also "con2000" from the B2 end)

What I see is that many overlapping phase2 connections are created. This doesn't actually stop the tunnels from working, but obviously something is wrong somewhere.

```
# on the A end
/usr/local/sbin/swanctl --list-sas | grep con1000 | wc -l
      12
/usr/local/sbin/swanctl --list-sas | grep con2000 | wc -l
      76

# on the B1 end
/usr/local/sbin/swanctl --list-sas | grep con4000 | wc -l
      12

# on the B2 end
/usr/local/sbin/swanctl --list-sas | grep con2000 | wc -l
      76
```

Actually the B1 and B2 ends also have a direct tunnel between them, and appear to have the same issue, so I don't think it's anything to do with the HA configuration.

```
# B1
/usr/local/sbin/swanctl --list-sas | grep con6000 | wc -l
       8

# B2
/usr/local/sbin/swanctl --list-sas | grep con5000 | wc -l
       8
```

Fuller --list-sas output from the A end, showing only the con1000 SAs to the B1 end:

```
con1000: #4360, ESTABLISHED, IKEv2, XXXXXXXX_i* XXXXXXXX_r
  local  'X.X.X.X' @ X.X.X.X[500]
  remote 'Y.Y.Y.Y' @ Y.Y.Y.Y[500]
  AES_CBC-128/HMAC_SHA2_256_128/PRF_HMAC_SHA2_256/MODP_2048
```

```
  established 24492s ago, reauth in 2826s
  con1000: #306966, reqid 1000, INSTALLED, TUNNEL, ESP:AES_CBC-128/HMAC_SHA2_256_128/MODP_2048
    installed 1187s ago, rekeying in 1530s, expires in 2413s
    in  c0d37246, 2897353 bytes, 24494 packets
    out c6da3dcd, 112669792 bytes, 87588 packets
    local  0.0.0.0/0|/0
    remote 0.0.0.0/0|/0
  con1000: #306967, reqid 1000, INSTALLED, TUNNEL, ESP:AES_CBC-128/HMAC_SHA2_256_128/MODP_2048
    installed 1146s ago, rekeying in 1408s, expires in 2454s
    in  c8c07783, 6915344 bytes, 69796 packets
    out c826b292, 310892936 bytes, 235526 packets
    local  0.0.0.0/0|/0
    remote 0.0.0.0/0|/0
  con1000: #306969, reqid 1000, INSTALLED, TUNNEL, ESP:AES_CBC-128/HMAC_SHA2_256_128/MODP_2048
    installed 1044s ago, rekeying in 1673s, expires in 2556s
    in  cdea40bc, 3814242 bytes, 41079 packets
    out c8c8d986, 181956192 bytes, 137160 packets
    local  0.0.0.0/0|/0
    remote 0.0.0.0/0|/0
  con1000: #306970, reqid 1000, INSTALLED, TUNNEL, ESP:AES_CBC-128/HMAC_SHA2_256_128/MODP_2048
    installed 993s ago, rekeying in 1949s, expires in 2607s
    in  ca2a3c90, 2457237 bytes, 32152 packets
    out ce583d9b, 140862888 bytes, 103906 packets
    local  0.0.0.0/0|/0
    remote 0.0.0.0/0|/0
  con1000: #306971, reqid 1000, INSTALLED, TUNNEL, ESP:AES_CBC-128/HMAC_SHA2_256_128/MODP_2048
    installed 963s ago, rekeying in 1684s, expires in 2637s
    in  cee1d7dc, 454128 bytes,  3866 packets
    out c7e834b9, 15782828 bytes, 12245 packets
    local  0.0.0.0/0|/0
    remote 0.0.0.0/0|/0
  con1000: #306972, reqid 1000, INSTALLED, TUNNEL, ESP:AES_CBC-128/HMAC_SHA2_256_128/MODP_2048
    installed 957s ago, rekeying in 1791s, expires in 2643s
    in  c312d39f, 2736480 bytes, 24750 packets
    out c25c5023, 110195864 bytes, 85358 packets
    local  0.0.0.0/0|/0
    remote 0.0.0.0/0|/0
  con1000: #306973, reqid 1000, INSTALLED, TUNNEL, ESP:AES_CBC-128/HMAC_SHA2_256_128/MODP_2048
    installed 912s ago, rekeying in 1620s, expires in 2688s
    in  c2265d92, 12641898 bytes, 119647 packets
    out c1e0e608, 518579896 bytes, 396490 packets
    local  0.0.0.0/0|/0
    remote 0.0.0.0/0|/0
  con1000: #306974, reqid 1000, INSTALLED, TUNNEL, ESP:AES_CBC-128/HMAC_SHA2_256_128/MODP_2048
    installed 732s ago, rekeying in 1812s, expires in 2868s
    in  c479aead, 5165104 bytes, 49388 packets
    out c12da9a9, 210883956 bytes, 161739 packets
    local  0.0.0.0/0|/0
    remote 0.0.0.0/0|/0
  con1000: #306975, reqid 1000, INSTALLED, TUNNEL, ESP:AES_CBC-128/HMAC_SHA2_256_128/MODP_2048
    installed 659s ago, rekeying in 1877s, expires in 2941s
    in  c634de90, 19159395 bytes, 249823 packets
    out cec9794d, 1069186884 bytes, 772087 packets
    local  0.0.0.0/0|/0
    remote 0.0.0.0/0|/0
  con1000: #306976, reqid 1000, INSTALLED, TUNNEL, ESP:AES_CBC-128/HMAC_SHA2_256_128/MODP_2048
    installed 465s ago, rekeying in 2231s, expires in 3135s
    in  c5bac3fc, 490906864 bytes, 441747 packets
    out cff7482e, 457592048 bytes, 402168 packets
    local  0.0.0.0/0|/0
    remote 0.0.0.0/0|/0
  con1000: #306979, reqid 1000, INSTALLED, TUNNEL, ESP:AES_CBC-128/HMAC_SHA2_256_128/MODP_2048
    installed 295s ago, rekeying in 2506s, expires in 3305s
    in  c0c1d451, 20607802 bytes, 195867 packets
    out cb1cded7, 847400800 bytes, 647600 packets
    local  0.0.0.0/0|/0
    remote 0.0.0.0/0|/0
```

Here is the con1000 tunnel configuration at the A end: (note I had to change the "pre-shared-key" XML tag to stop redmine mangling it)

```
        <phase1>
            <ikeid>1</ikeid>
            <iketype>ikev2</iketype>
            <interface>_vip5ce58f3a60ba7</interface>
            <remote-gateway>Y.Y.Y.Y</remote-gateway>
            <protocol>inet</protocol>
            <myid_type>myaddress</myid_type>
            <myid_data></myid_data>
            <peerid_type>peeraddress</peerid_type>
            <peerid_data></peerid_data>
            <encryption>
                <item>
                    <encryption-algorithm>
                        <name>aes</name>
                        <keylen>128</keylen>
                    </encryption-algorithm>
                    <hash-algorithm>sha256</hash-algorithm>
                    <dhgroup>14</dhgroup>
                </item>
            </encryption>
            <lifetime>28800</lifetime>
            <Xre-shared-key>XXXXXXXX</Xre-shared-key>
            <private-key></private-key>
            <certref></certref>
            <caref></caref>
            <authentication_method>pre_shared_key</authentication_method>
            <descr><![CDATA[lch-fw]]></descr>
            <nat_traversal>on</nat_traversal>
            <mobike>off</mobike>
            <margintime></margintime>
            <dpd_delay>10</dpd_delay>
            <dpd_maxfail>5</dpd_maxfail>
        </phase1>
...
        <phase2>
            <ikeid>1</ikeid>
            <uniqid>5ce644f67e37d</uniqid>
            <mode>vti</mode>
            <reqid>1</reqid>
            <localid>
                <type>network</type>
                <address>10.9.1.17</address>
                <netbits>29</netbits>
            </localid>
            <remoteid>
                <type>address</type>
                <address>10.9.1.18</address>
            </remoteid>
            <protocol>esp</protocol>
            <encryption-algorithm-option>
                <name>aes</name>
                <keylen>128</keylen>
            </encryption-algorithm-option>
            <encryption-algorithm-option>
                <name>aes128gcm</name>
                <keylen>128</keylen>
            </encryption-algorithm-option>
            <hash-algorithm-option>hmac_sha256</hash-algorithm-option>
            <pfsgroup>14</pfsgroup>
            <lifetime>3600</lifetime>
```

```
            <pinghost></pinghost>
            <descr></descr>
        </phase2>
```

And the corresponding con4000 tunnel configuration at the B1 end:

```
        <phase1>
            <ikeid>4</ikeid>
            <iketype>ikev2</iketype>
            <interface>wan</interface>
            <remote-gateway>X.X.X.X</remote-gateway>
            <protocol>inet</protocol>
            <myid_type>myaddress</myid_type>
            <myid_data></myid_data>
            <peerid_type>peeraddress</peerid_type>
            <peerid_data></peerid_data>
            <encryption>
                <item>
                    <encryption-algorithm>
                        <name>aes</name>
                        <keylen>128</keylen>
                    </encryption-algorithm>
                    <hash-algorithm>sha256</hash-algorithm>
                    <dhgroup>14</dhgroup>
                </item>
            </encryption>
            <lifetime>28800</lifetime>
            <Xre-shared-key>XXXXXXXX</Xre-shared-key>
            <private-key></private-key>
            <certref></certref>
            <caref></caref>
            <authentication_method>pre_shared_key</authentication_method>
            <descr><![CDATA[ldex-fw]]></descr>
            <nat_traversal>on</nat_traversal>
            <mobike>off</mobike>
            <margintime></margintime>
            <dpd_delay>10</dpd_delay>
            <dpd_maxfail>5</dpd_maxfail>
        </phase1>
...
        <phase2>
            <ikeid>4</ikeid>
            <uniqid>5ce644a266cf6</uniqid>
            <mode>vti</mode>
            <reqid>4</reqid>
            <localid>
                <type>network</type>
                <address>10.9.1.18</address>
                <netbits>29</netbits>
            </localid>
            <remoteid>
                <type>address</type>
                <address>10.9.1.17</address>
            </remoteid>
            <protocol>esp</protocol>
            <encryption-algorithm-option>
                <name>aes</name>
                <keylen>128</keylen>
            </encryption-algorithm-option>
            <encryption-algorithm-option>
                <name>aes128gcm</name>
                <keylen>128</keylen>
            </encryption-algorithm-option>
            <hash-algorithm-option>hmac_sha256</hash-algorithm-option>
```

```
            <pfsgroup>14</pfsgroup>
            <lifetime>3600</lifetime>
            <pinghost></pinghost>
            <descr></descr>
        </phase2>
```

Side note: there is OpenBGP routing on top of this, and there is some relaying of traffic via VTI interfaces. Specifically: A also has tunnels to AWS, and there is traffic which flows B1 -> A -> AWS, and B2 -> A -> AWS (i.e. in one VTI interface and out another VTI interface). I can't see how this has any relevance, given that VTI SAs match 0.0.0.0/0 and therefore should allow all traffic, but I thought it was worth mentioning.

Associated revisions

Revision 9a69dd4b - 06/04/2020 02:09 PM - Jim Pingle

Fix VTI responder only on 2.4.x. Fixes #10176

This only affects 2.4.x, the swanctl rewrite in 2.5.0 fixed this already

History

#1 - 01/10/2020 05:04 AM - Brian Candler

I should add: these overlapping SAs *don't* occur for VTI tunnels to AWS. I consistently get only a single phase2 SA for each AWS tunnel:

```
# At "A end", which also has 4 VTI tunnels to AWS
/usr/local/sbin/swanctl --list-sas | grep 'con[4567]000'
con7000: #4364, ESTABLISHED, IKEv1, XXXX_i* XXXX_r
  con7000: #307070, reqid 7000, INSTALLED, TUNNEL-in-UDP, ESP:AES_CBC-128/HMAC_SHA1_96/MODP_2048
con5000: #4362, ESTABLISHED, IKEv1, XXXX_i* XXXX_r
  con5000: #307072, reqid 5000, INSTALLED, TUNNEL-in-UDP, ESP:AES_CBC-128/HMAC_SHA1_96/MODP_1024
con4000: #4368, ESTABLISHED, IKEv1, XXXX_i* XXXX_r
  con4000: #307068, reqid 4000, INSTALLED, TUNNEL-in-UDP, ESP:AES_CBC-128/HMAC_SHA1_96/MODP_1024
con6000: #4367, ESTABLISHED, IKEv1, XXXX_i* XXXX_r
  con6000: #307059, reqid 6000, INSTALLED, TUNNEL-in-UDP, ESP:AES_CBC-128/HMAC_SHA1_96/MODP_2048
```

I do have a different problem with AWS tunnels (#10175) - it's the same HA pair for both these tickets.

#2 - 01/10/2020 06:25 AM - Jim Pingle

- Category set to IPsec

- Status changed from New to Duplicate

- Affected Version deleted (2.4.4-p3)

If there is anything actionable here it's almost certainly solved by #9603 and needs tested on 2.5.0 snapshots.

If it's not, then it's in strongSwan and not something we can control.

#3 - 02/15/2020 09:16 AM - Izaac Falken

I just watched this happen in 2.5.0-DEVELOPMENT (amd64) with a configuration straight out of:
https://docs.netgate.com/pfsense/en/latest/vpn/ipsec/ipsec-routed.html

So, no, #9603 does not solve it.

What now? A shrug "it's StrongSwan" is not an acceptable answer.

#4 - 02/15/2020 09:31 AM - Jim Pingle

Was it 2.5.0 on both ends? If either end is 2.4.x, it still could be that side triggering the problem.

#5 - 02/17/2020 04:50 PM - Jim Pingle

- Status changed from Duplicate to Feedback

- Assignee set to Jim Pingle

- Target version set to 2.5.0

I don't yet see a reason why it happened, but I caught one tunnel in my lab doing this, 2.5.0 to 2.5.0. An identical tunnel on another pair of lab boxes didn't do it.

I was able to stop both sides, restart, and a new copy appeared around when the tunnel rekeyed, but it isn't consistent.

I set "Child SA Close Action" on one side to "Restart/Reconnect" and I set the other side to "Close connection and clear SA", and so far it has not come back. That setting is available on 2.4.5 and 2.5.0. Leaving this on Feedback for a bit to see if it comes back or if I can get a better lead on what is happening otherwise.

#6 - 02/18/2020 08:04 AM - Jim Pingle

- Status changed from Feedback to In Progress

It took it longer to happen but it still happened when set that way. Still investigating.

#7 - 02/21/2020 09:29 AM - Izaac Falken

Jim Pingle wrote:

> Was it 2.5.0 on both ends? If either end is 2.4.x, it still could be that side triggering the problem.

Yes. (You've reproduced it yourself, so this comment is moot. But I didn't want to just leave the question hanging.)

#8 - 05/14/2020 01:19 PM - Jim Pingle

Looking at this again on 2.5.0, now that it's on strongSwan 5.8.4. I do not see any of my VMs with multiple overlapping child SA entries despite them having ~2 weeks of uptime. (Even some still on 12.0-RELEASE vs some on 12.1-STABLE). Looks like one of the changes in recent strongSwan release may have solved it.

I don't see any on 2.4.5 either (strongSwan 5.8.2) but the 2.4.5 systems I have up at the moment all have low uptimes so I can't be certain they just haven't shown the problem yet.

#9 - 06/03/2020 09:10 AM - Jim Pingle

Contrary to my last note, I am seeing this still, but it still appears to be unpredictable. A system that doesn't show it for long stretches of time will have three the next time I check it. Others have them after only a couple hours. Most only stay in the single digits, some end up super high (~70).

#10 - 06/04/2020 01:38 PM - Jim Pingle

Digging deeper in strongSwan most of the times this has happened in the past have been due to the use of IKEv2 with reauthentication and break-before-make, or a race condition where both sides attempt to initiate at nearly the same time.

Since I have been able to reproduce this on a few different pairs of lab systems, I've changed a couple in the following ways:

- Left both sides on reauthenticate, but enabled make-before-break (On the IPsec Advanced Settings tab) -- Note that this behavior must be supported by both peers. If both peers are pfSense, this should be OK.

- OR -

- Disabled reauthentication and enabled rekey instead (On 2.5.0, leave reauth blank and put a value in rekey. On 2.4.5, check Disable Reauth but leave Disable Rekey unchecked.

The race condition should be solved using advice from earlier on this issue. Set one side to responder only. On the initiator side, set the child SA close action to restart/reconnect. On the responder side, set the Child SA close action to Close/Clear.

After 24 hours all pairs I configured in both of these ways still only had one single child SA. But I'll leave them running and check again after a few more days of uptime/rekeys.

#11 - 06/04/2020 02:20 PM - Jim Pingle

There is a small bug on 2.4.x which prevents responder only from working on VTI, I've pushed a fix for that, but it's too late for 2.4.5-p1 unless circumstances dictate a rebuild. It's a one-line change if anyone wants to make it manually, it will appear here on the issue shortly.

#12 - 06/04/2020 02:20 PM - Jim Pingle

- Status changed from In Progress to Feedback

- % Done changed from 0 to 100

Applied in changeset 9a69dd4b8ff6eeeaf5779b7388a10743afae8e91.

#13 - 06/05/2020 05:14 AM - Marc L

I have a GNS3 lab setup with two pfSense VMs connected via IPSec (IKEv2, VTI). Multi-WAN with failover on one side. Whenever a gateway event/failover occurs there, more child SAs are created. Make-before-break is enabled on both sides.

It looks like after a fresh reboot it is capped/limited at two child SAs, can't push it beyond that. However when i manually disconnect the tunnel and click on connect again, i can easily produce as many Child SAs as i want. Basically everytime i suspend and resume the first WAN link (to trigger automatic failover) the number increases. I would assume that if i wait long enough after a reboot, it also works without disconnecting the tunnel first. In any case, make-before-break does not seem to be a ~~viable~~ reliable workaround on 2.4.x

I can't provoke duplicates on the latest 2.5.0 snapshot, seems to be fixed there...

#14 - 06/05/2020 08:18 AM - Jim Pingle

If it happens on disconnect/reconnect that is more likely the race condition case and not the reauth case. I wouldn't expect make-before-break to help there, but the initiator/responder/SA close action may help.

If it's better on 2.5.0 that could either be from the swanctl conversion, a newer strongSwan, or the initiator/responder setting working there when it doesn't work for VTI on 2.4.x. 2.4.5-p1 will have the newer version of strongSwan, you could apply the patch here on the issue to fix "responder only" for VTI, but if it's related to the swanctl changes then you'll have to run 2.5.0 to get that.

#15 - 06/05/2020 08:29 AM - Jim Pingle

All of my test pairs still only have a single SA this morning (2.4.5 and 2.5.0, multiple causes and changes mentioned here). I'll check them again after the weekend.

#16 - 08/03/2020 08:04 AM - Izaac Falken

So is there a final, required set of baseline versions and recommended configuration which can do into the docs? Or at least here?

#17 - 08/03/2020 10:26 AM - Jim Pingle

Side 1: IKEv2, Rekey configured, Reauth disabled, child SA close action set to restart/reconnect
Side 2: IKEv2, Rekey configured, Reauth disabled, responder only set, child SA close action left at default (clear)

That's the best setup I've hit so far, though I still do see multiple SAs on rare occasions it's nowhere near what it was previously.

#18 - 08/07/2020 07:38 AM - Izaac Falken

Did this. Within 48 hours I have six overlapping phase 2s and am in the #11000's in IPsec IDs.
I'm pretty sure it's time to conclude that VTI does not work and needs to be pulled as an option altogether.

#19 - 08/07/2020 07:54 AM - Jim Pingle

Except that it does work, and thousands of people are using is successfully, and pulling it would cause much more harm at this point.

There is clearly some environmental/situational problem here, not a general problem. It doesn't happen to every setup, even practically identical setups vary in behavior for as-yet-unknown reasons.

Something is triggering this in strongSwan, it would be more productive to focus on what is happening to cause the duplication. (For example, stop strongSwan on both ends, setup remote syslog for ipsec logs, start strongswan and monitor it closely and check the logs when the additional SAs start appearing).

#20 - 08/07/2020 10:05 AM - Izaac Falken

Jim Pingle wrote:

> Except that it does work, and thousands of people are using is successfully

Are they? Or are they just not noticing because they're not looking? Eventually, those IDs run out and one side or the other no longer receives packets. All the SADs and SPIs check out. They're just not there on the interface. This broken state can't even be repaired by restarting IPsec.

If you want to see it happen, string together eight sites with VTI links and iBGP. In about thirty days, the core falls over and can't get back up without a reboot. Seriously. I've tried. And I don't have time to throw the kernel on the debugger and figure it out. This has been a huge source of embarrassment for me as an advocate and for the product itself since the switch to strongSwan in 2.4.4.

It's pretty evident that VTIs are not behaving correctly. If you don't want to pull it, I'd recommend marking it experimental.

#21 - 08/07/2020 10:16 AM - Jim Pingle

If you want to discuss it, take it to the forum. As I said, there are many people using it with success. It doesn't affect every installation. Without figuring out *why*, complaining that it's broken is practically useless. I do have systems on which I can reproduce it, but it doesn't happen until they have been up for weeks, so it's hard to catch and difficult to notice exactly when the problem starts. If you can reproduce it faster, by all means, collect the necessary data. But if you are unwilling to assist in debug why it happens in your environment, then consider your complaint noted and stop posting unhelpful comments.

Running out of IDs when reconnecting should not be a problem on 2.4.5-p1 as it contains strongSwan 5.8.4, which includes the fix for https://wiki.strongswan.org/issues/2315 -- If that still affects your setup, then perhaps the bug hasn't really been fixed in strongSwan, but it's not directly related to this issue.

There are several similar bug reports for duplicate SAs in the strongSwan bug tracker, but last I looked, all of them imply it's a settings issue with IKEv2 and reauth, which the settings I recommended above should address. There may be some other problem in strongSwan, but again, it would be more helpful to figure out the specifics, and ultimately it's going to need to be fixed by strongSwan if it's a problem in strongSwan, which it appears to be at the moment (but we don't know for sure.)

#22 - 08/07/2020 04:18 PM - Izaac Falken

Ticket is marked for Feedback. Feedback is being provided.

#23 - 09/07/2020 02:22 AM - Marc L

I have successfully used your patch and suggested settings on a pair of SG-3100s. There i have two tunnels to AWS that work almost flawlessly. However on another cluster (two Dell R210 II servers) i'm consistently getting duplicates over relatively short periods of time. The patch is applied to all systems, and settings are identical.

Since it's so consistent, maybe we can grab interesting logs from there? Which ones do you need? What log levels would be interesting? This issue is a bummer and made us consider moving away from pfSense, which would be a pity because everything else has been awesome for us over the last few years. But site-to-site IPSec (VTI) tunnels will become more and more of a standard thing for us, they have to be reliable...

This thread may also be related https://forum.netgate.com/topic/153791/vpns-disconnecting-reported-memory-issue

#24 - 09/09/2020 04:25 PM - Kyle Mulligan

I was observing similar behavior on an SG-3100 to XG-1537 VTI tunnel (both 2.4.5-p1 w/ patch and recommended P1 settings). Something that stood out in the logs was the following repeating cycle of events...

Sep 8 17:43:54 rc.gateway_alarm 8970 >>> Gateway alarm: VTIINTERFACE_VTIV4 (Addr:172.31.0.17 Alarm:0 RTT:46.043ms RTTsd:6.019ms Loss:5%)
Sep 8 17:43:54 check_reload_status updating dyndns VTIINTERFACE_VTIV4
Sep 8 17:43:54 check_reload_status Restarting ipsec tunnels

```
Sep 8 17:43:54   check_reload_status       Restarting OpenVPN tunnels/interfaces
Sep 8 17:43:54   check_reload_status       Reloading filter
Sep 8 17:44:10   php-fpm   364   /rc.newipsecdns: IPSEC: One or more IPsec tunnel endpoints has changed its IP. Refreshing.
Sep 8 17:44:10   check_reload_status       Reloading filter
```

While this certainly seems related, it might be a bit different as it resulted in a fresh SA, but traffic was interrupted for a brief moment and OSPF had to spin back up. No explanation for the completely random monitoring alarms. They occurred from the SG-3100's PoV despite no other connectivity issues. In the end, I had to disable the gateway monitoring action, but it's been stable for over 24 hours and it's appropriately maintaining the SAs now.

I just chalked it up to an ARM quirk as I've never had these issues with our Qotom x86 boxes connected to the same XG-1537 with virtually identical configurations.

#25 - 09/22/2020 01:17 AM - Marc L

Could another difference-maker be NAT-T? As reported above, i'm consistently seeing duplicates on a cluster i'm operating. Interestingly, it only happens on Tunnels to AWS, but not on Tunnels to an EdgeRouter and to another pfSense (both under my control). The difference is that the AWS tunnels show NAT-T while the others don't.

Working Example (only one Child SA):

```
con1000: #5151, reqid 1000, INSTALLED, TUNNEL, ESP:AES_CBC-256/HMAC_SHA2_256_128/MODP_4096
  installed 94s ago, rekeying in 2677s, expires in 3507s
  in  ccebc395,  15388 bytes,   424 packets
  out cd9a3565,  35256 bytes,   378 packets
  local  0.0.0.0/0|/0
  remote 0.0.0.0/0|/0
```

Broken Example:

```
con4000: #5132, reqid 4000, INSTALLED, TUNNEL-in-UDP, ESP:AES_CBC-256/HMAC_SHA2_256_128/MODP_4096
  installed 2389s ago, rekeying in 172s, expires in 1211s
  in  cbaef5f6,  25106 bytes,   316 packets
  out 79ef4ff7,  62676 bytes,   435 packets
  local  0.0.0.0/0|/0
  remote 0.0.0.0/0|/0
con4000: #5133, reqid 4000, INSTALLED, TUNNEL-in-UDP, ESP:AES_CBC-256/HMAC_SHA2_256_128/MODP_4096
  installed 2319s ago, rekeying in 241s, expires in 1281s
  in  c4a7ad40,  20995 bytes,   285 packets
  out 3879fb50,  54144 bytes,   376 packets
  local  0.0.0.0/0|/0
  remote 0.0.0.0/0|/0
con4000: #5134, reqid 4000, INSTALLED, TUNNEL-in-UDP, ESP:AES_CBC-256/HMAC_SHA2_256_128/MODP_4096
  installed 2257s ago, rekeying in 317s, expires in 1343s
  in  cac3b309,   4844 bytes,    86 packets
  out f1fb3f7a,  18704 bytes,   132 packets
  local  0.0.0.0/0|/0
  remote 0.0.0.0/0|/0
con4000: #5135, reqid 4000, INSTALLED, TUNNEL-in-UDP, ESP:AES_CBC-256/HMAC_SHA2_256_128/MODP_4096
  installed 2234s ago, rekeying in 466s, expires in 1366s
  in  ce391b6d,  14330 bytes,   183 packets
  out c1248764,  36568 bytes,   250 packets
  local  0.0.0.0/0|/0
  remote 0.0.0.0/0|/0
con4000: #5136, reqid 4000, INSTALLED, TUNNEL-in-UDP, ESP:AES_CBC-256/HMAC_SHA2_256_128/MODP_4096
  installed 2195s ago, rekeying in 362s, expires in 1405s
  in  c9f21daf, 560345 bytes,  7857 packets
  out bf275006, 1507208 bytes, 10486 packets
  local  0.0.0.0/0|/0
```

```
    remote 0.0.0.0/0|/0
  con4000: #5146, reqid 4000, INSTALLED, TUNNEL-in-UDP, ESP:AES_CBC-256/HMAC_SHA2_256_128/MODP_4096
    installed 462s ago, rekeying in 2193s, expires in 3138s
    in  c08737f6,  54508 bytes,   710 packets
    out 70d7adba, 129676 bytes,   901 packets
    local  0.0.0.0/0|/0
    remote 0.0.0.0/0|/0
  con4000: #5147, reqid 4000, INSTALLED, TUNNEL-in-UDP, ESP:AES_CBC-256/HMAC_SHA2_256_128/MODP_4096
    installed 315s ago, rekeying in 2222s, expires in 3285s
    in  ca7144e0,  25618 bytes,   401 packets
    out be33b986,  78124 bytes,   545 packets
    local  0.0.0.0/0|/0
    remote 0.0.0.0/0|/0
  con4000: #5148, reqid 4000, INSTALLED, TUNNEL-in-UDP, ESP:AES_CBC-256/HMAC_SHA2_256_128/MODP_4096
    installed 225s ago, rekeying in 2365s, expires in 3375s
    in  c4e86d84,  37407 bytes,   513 packets
    out ead0f31a, 100964 bytes,   703 packets
    local  0.0.0.0/0|/0
    remote 0.0.0.0/0|/0
  con4000: #5150, reqid 4000, INSTALLED, TUNNEL-in-UDP, ESP:AES_CBC-256/HMAC_SHA2_256_128/MODP_4096
    installed 108s ago, rekeying in 2462s, expires in 3492s
    in  c5289559,  36070 bytes,   486 packets
    out c0bab313,  93256 bytes,   650 packets
    local  0.0.0.0/0|/0
    remote 0.0.0.0/0|/0
```

Then again, i had Tunnels to AWS with identical configuration working on my SG-3100s. But that setup has since been removed so i can't compare the output unfortunately...

#26 - 09/22/2020 01:43 AM - Brian Candler

I don't think NAT-T is the issue. All my firewalls have public IPs, and my tunnels don't have NAT-T (see status output in original post).

#27 - 10/07/2020 04:44 PM - Todd Blum

I'm having the same issue with duplicating VTI Phase2s with tunnels to AWS.

Did anyone find settings that fixed this with tunnels to AWS?

#28 - 10/27/2020 08:50 AM - Todd Blum

This is the response from Amazon. Since they weren't sure about 'Make before break' I will try the other settings they suggest:

Upon investigation, I also confirmed the multiple duplicated Phase2s of your VPN and randomly re-keying at unexpected time.
In regard to the proposed configuration changes, I assume Side 1 is AWS whereas Side 2 is your CGW device. Please see below comments:

> Side 1
> Disable Reauthentication
> Set to 'Responder Only'

"Reauthentication" is NOT supported by AWS VPN so you don't need to worry about it. It's always going to Rekey instead of Reauthentication.
AWS VPN is in 'Responder Only' by default. However, I see you have configured "Startup Action": Start. Please make sure both "Startup action" and "DPD timeout action" are configured back to below[1][2]:

DPD Timeout Action: Clear
Startup Action: Add

You can make changes to both via VPC Console -> Site-to-Site VPN Connections -> (Select your VPN)->Actions->Modify VPN Tunnel Options.

> Side 2
> Disable Reauthentication
> Set 'Child SA close action' to 'Restart/Reconnect'

"Reauthentication" is not supported by AWS VPN, so disabling it on the CGW won't cause any issue.
According to Netgate documentation[3], I understand 'Child SA close action' controls how the IPsec daemon behaves when a child SA (P2) is unexpectedly closed by the peer.
If "Restart/Reconnect" is configured on your side, it should eliminate the issue as DPD Timeout Action: Clear and Startup Action: Add are configured at AWS side.

Resource:
[1] Site-to-Site VPN tunnel initiation options - https://docs.aws.amazon.com/vpn/latest/s2svpn/initiate-vpn-tunnels.html
[2] Tunnel options for your Site-to-Site VPN connection - https://docs.aws.amazon.com/vpn/latest/s2svpn/VPNTunnels.html
[3]https://docs.netgate.com/pfsense/en/latest/vpn/ipsec/configure.html#advanced-options

Lastly, Make-before-break option is not available for configuration on AWS side. But I think that is the default behavior.
I hope the above information has addressed all your concerns. Please let me know if the issue persists after you make the above changes. Happy to further assist you on this issue.