InterBox Internet
Venkelbaan 58
2908 KE Capelle aan den IJssel
Telefoon 010-4422664
Email box@box.nl

KvK Rotterdam 24188264
BTW nummer NL800688910B01

IBAN NL32 ABNA 0412 4564 00
BIC ABNANL2A

Reason For Outage

Starttijd: **maandag 18 april 2022 10:00**
Eindtijd: **maandag 18 april 2022 20:53**

Wat vooraf ging:

Op zondag 10 april treedt er packetloss op in ons netwerk. Systemen blijven echter wel bereikbaar, zij het soms moeizaam. Na herstart van de core switch is de packetloss verdwenen. Na verder onderzoek vinden we alleen een aantal logregels over een klant switch die aangeeft dat MAC adressen ineens van VLAN wisselen. We voegen een controle toe zodat die meldingen voortaan als critical door ons notificatie systeem worden verzonden. Verder blijkt op een enkele switchport de spanning-tree optie nog aan te staan. Dat protocol zetten we ook op die poorten uit.

Reason For Outage

Beschrijving incident:

Op maandag 18 april 2022 treedt opnieuw packetloss op in ons netwerk, maar de impact is nu veel groter. Op deze dag gebeurde het volgende:

12:30 We krijgen meldingen er problemen zijn in het netwerk.

Opnieuw stellen we vast dat er packetloss optreedt, maar dat ook andere diensten soms slecht werken. Verbinding maken met onze VPN om toegang te krijgen het het management systeem lukt wel, maar toegang tot het systeem zelf lukt niet.

12:48 Omdat het Pasen is en dus niet op kantoor zijn, gaan we onderweg naar Capelle en Rotterdam om daar ter plekke te kunnen onderzoeken wat er mis gaat.

We verwachten dat het herstarten van de core switch het probleem oplost worden die opnieuw gestart. Helaas blijkt dat deze keer niet de oplossing.

De core switch bestaan uit twee losse apparaten (switch A en B) die onderling verbonden zijn waarbij gebruik wordt gemaakt van MLAG (multi chassis aggregation). Hiermee kunnen we systemen en klanten dubbel aansluiten, waarbij een van de switches kan uitvallen zonder gevolgen voor de verbinding.

We maken switch B stroomloos en daardoor verdwijnt de packetloss. Zodra switch B weer wordt gestart is de packetloss terug. Dan maken we switch A stroomloos en is de packetloss ook weg. Switch A weer starten zorgt voor terugkerende packetloss.

Omdat de oorzaak van packetloss daarmee van buiten de core switch lijkt te komen wordt in switch B een voor een een netwerkkabel los gehaald. Bij een van de kabels heeft dat resultaat. Deze netwerkkabel loopt naar een klant in onze colocatie. We laten deze kabel los.

Daarmee is de packetloss verdwenen, maar al onze diensten blijven moeizaam of zelfs helemaal niet functioneren.

Dan blijkt dat het storage platform geen data meer wil schrijven. Het lijkt er op dat door de netwerk problemen de hoeveelheid log informatie die geschreven wordt zo groot is geworden, dat de disk waarop die informatie staat razendsnel is volgelopen (vreemd genoeg trad dat op 10 april niet op). Logbestanden worden leeggemaakt, maar binnen een minuut zijn weer vele gigabytes logging geschreven. Op diezelfde disk staat ook vitale informatie voor het cluster zelf (journals), waardoor ook die data niet geschreven kan worden. Hierdoor vallen de zogenaamde monitor processen uit, waardoor het cluster niet meer werkt.

Het leegmaken van die logbestanden wordt snel automatisch gemaakt, zodat er ruimte blijft voor de monitor processen die weer gestart worden. Het storage systeem komt daardoor weer tot leven en diensten beginnen ook weer op gang te komen.

Het storage systeem is drievoudig uitgevoerd. De status geeft aan dat alle disken in het systeem actief zijn, maar toch besluit het om data te gaan recoveren.

We herstarten alle disk processen omdat we vermoeden dat sommige misschien toch niet meer actief zijn. Dat levert echter niet het gewenste resultaat. Daarna komen we tot ontdekking dat sommige disk processen helemaal niet meer gestart zijn, terwijl de status weergeeft dat ze dat wel zijn. We starten die processen op en daarna komt het systeem tot rust en komen alle diensten weer op.

We controleren alle diensten op hun werking en voeren daar waar nodig herstelwerkzaamheden uit.

Reason For Outage

Verdere analyse:

In de avond zoeken we nog verder naar oorzaken en oplossingen.

- Het eerste probleem waar we tegen aan liepen vandaag was dat door alle beveiligingen die we hebben ingebouwd we bijna ons eigen netwerk moesten kraken om toegang te krijgen. Zo konden we bijvoorbeeld niet met ons wifi netwerk verbinden omdat de autorisatie daarvan ook gebruik maakt van de systemen die niet werkten.

- De kabel die los is gehaald loopt naar een eveneens dubbel uitgevoerde switch van een colocatie klant. Deze switches zijn met twee netwerkkabels op onze core switches aangesloten waarvan er nu nog een los is gehaald. In de logfiles vinden we dat deze switches op die dag om 10:00:57 al een melding geven over geheugen gebruik voorbij een bepaalde threshold was. Die meldingen gaan ook in de avond gewoon door, alleen lijken ze nu door het loshalen van de netwerkkabel geen invloed te hebben op het netwerk.

- We loggen in op die klant switches en merken dat ze heel traag reageren. De memory utilization blijkt 98,31% te zijn. Veel opdrachten om te kijken wat er op de switch gebeurd werken door geheugen gebrek niet. We herstarten de switches dan maar om de beurt en de meldingen in de log verdwijnen en de switches reageren weer vlot.

- Omdat deze switches van precies hetzelfde merk en type zijn als we ook op andere plekken in ons netwerk en bij klanten gebruiken controleren we ook die switches allemaal op memory gebruik. We vinden nog een dubbel uitvoerde set die erg veel geheugen gebruikt. Daar kunnen we nog wel meer informatie over het gebruik uit halen en het geheugen lijkt vooral in gebruik door SNMP processen.

Reason For Outage

Vervolgstappen:

De volgende vervolgstappen zijn genomen of worden op korte termijn genomen:

- We hebben het ophalen van informatie uit switches met SNMP tijdelijk uitgeschakeld tot we weten waarom dat in de switch zoveel geheugen vraagt.

- We hebben een specifieke netwerk aansluiting gemaakt waarmee we zelfs zonder wifi en VPN verbinding toegang kunnen krijgen tot ons management netwerk. De beveiliging is in dit geval geregeld op basis van MAC adressen van onze laptops.

- Onze zogenaamde Jump servers, waarmee we toegang krijgen tot andere servers was al dubbel uitgevoerd, maar een van de twee draait nu vanaf een lokale storage, waardoor ze niet beiden afhankelijk zijn van hetzelfde storage systeem.

- De switches van de klant waarin we het hoge geheugen gebruik voor SNMP processen is gevonden zullen in overleg met de klant herstart worden.

- Er is een systeem controle toegevoegd die regelmatig op switches van dit merk/type controleert of het geheugen gebruik bepaalde grenzen overschrijd.

- De logs van het storage systeem worden niet meer op het storage systeem zelf vastgelegd, maar doorgestuurd naar onze centrale logging servers. Mocht het storage systeem ooit weer zo'n enorme hoeveelheid logs genereren, dan zal hooguit dat systeem vol lopen.

- In de hectiek zijn we laat met het plaatsen van een storingsmelding op boxstatus.nl, we prenten onszelf nogmaals in dat een volgende keer eerder te doen.