Whitepaper AWS

Creazione di un'infrastruttura di rete AWS multi-VPC sicura e scalabile

Creazione di un'infrastruttura di rete AWS multi-VPC sicura e scalabile: Whitepaper AWS

Table of Contents

Creazione di un'infrastruttura di rete AWS multi-VPC sicura e scalabile

Data di pubblicazione: 10 giugno 2020 (Cronologia del documento)

Riassunto

I clienti AWS spesso si affidano a centinaia di account e VPC per segmentare i carichi di lavoro ed espandere il footprint. Questo livello di scalabilità comporta in molti casi la nascita di problematiche relative alla condivisione delle risorse, alla connettività tra VPC e alla connettività da ambiente On-Premise a VPC.

Questo whitepaper descrive le best practice per la creazione di architetture di rete scalabili e sicure in una rete di grandi dimensioni tramite servizi AWS come Amazon VPC, AWS Transit Gateway, AWS PrivateLink e AWS Direct Connect Gateway. Vengono illustrate soluzioni per la gestione dell'infrastruttura in crescita, in modo da offrire scalabilità, disponibilità elevata e sicurezza mantenendo bassi i costi generali.

Introduzione

I clienti AWS iniziano creando risorse in un singolo account AWS che rappresenta un confine di gestione che segmenta autorizzazioni, costi e servizi. Tuttavia, man mano che l'organizzazione del cliente cresce, diventa necessaria una maggiore segmentazione dei servizi per monitorare i costi, controllare l'accesso e fornire una gestione dell'ambiente più semplice. Una soluzione con più account consente di risolvere questi problemi fornendo account specifici per i servizi IT e gli utenti all'interno di un'organizzazione. AWS fornisce diversi strumenti per gestire e configurare questa infrastruttura, tra cui la zona di destinazione AWS e AWS Control Tower.

Figura 1. Struttura degli account delle zone di destinazione

La zona di destinazione AWS e AWS Control Tower automatizzano la configurazione e l'integrazione di più servizi AWS per fornire un ambiente di base, altamente controllato e con più account che offre funzionalità IAM (Identity and Access Management), governance, sicurezza dei dati, progettazione della rete e registrazione.

La soluzione di zona di destinazione AWS nella Figura 1 include quattro account: l'account AWS Organizations (utilizzato per gestire la configurazione e l'accesso agli account gestiti dalla zone di destinazione AWS), l'account dei servizi condivisi (utilizzato per la creazione dei servizi condivisi dell'infrastruttura come i servizi di directory), l'account dell'archivio dei registri (registrazione centralizzata nei bucket S3) e l'account di sicurezza (che deve essere utilizzato da un team aziendale responsabile della sicurezza e della conformità a scopo di verifica o per eseguire operazioni di sicurezza di emergenza in caso di incidente negli account spoke).

Questo whitepaper introduce un account dei servizi di rete di proprietà del team di rete che gestisce l'infrastruttura AWS. I servizi di rete e l'infrastruttura di rete per l'account sono condivisi da tutti gli account e i VPC in modo centralizzato (analogamente a una progettazione di tipo hub-spoke). Questa progettazione migliora la gestione della zona di destinazione e aiuta a ridurre i costi, eliminando la necessità di duplicare i servizi di rete in ogni VPC e account spoke.

> **Note**
>
> In questo whitepaper, il termine "zona di destinazione" viene utilizzato in modo generico per indicare una configurazione con più account/VPC scalabile, sicura e di alto livello in cui

vengono distribuiti i carichi di lavoro. Questa configurazione può essere creata utilizzando qualsiasi strumento.

La maggior parte dei clienti inizia distribuendo l'infrastruttura con pochi VPC. Il numero di VPC di proprietà di un cliente è in genere correlato al numero di account, utenti e ambienti in uso (produzione, sviluppo, test e così via). Man mano che l'utilizzo del cloud aumenta, il numero di utenti, business unit, applicazioni e regioni con cui un cliente interagisce si moltiplica, portando alla creazione di nuovi VPC.

Con l'aumento del numero di VPC, diventa essenziale disporre di funzionalità di gestione di più VPC per la rete cloud del cliente. Questo whitepaper illustra le best practice per tre aree specifiche della connettività ibrida e tra VPC:

- Connettività di rete: interconnessione di VPC e reti On-Premise su larga scala.
- Sicurezza di rete: creazione di punti di uscita centralizzati per l'accesso a Internet ed endpoint come il gateway NAT, endpoint VPC e AWS PrivateLink.
- Gestione DNS: risoluzione dei nomi DNS nella zona di destinazione e in ambienti ibridi.

Connettività da VPC a VPC

I clienti possono utilizzare due diversi modelli di flusso VPC per configurare ambienti multi-VPC: molti-a-molti o hub-and-spoke. Nell'approccio molti-a-molti, il traffico tra ogni VPC viene gestito individualmente. Nel modello hub-and-spoke, tutto il traffico tra VPC passa attraverso una risorsa centrale, che lo instrada in base a regole stabilite.

Argomenti

Peering VPC

Il modo più semplice per connettere due VPC è tramite il peering VPC. In questa configurazione, una connessione consente connettività bidirezionale completa tra i VPC. Questa connessione peering viene utilizzata per instradare il traffico tra i VPC. È inoltre possibile eseguire il peering dei VPC tra account e regioni AWS. Il peering VPC comporta solo costi per il traffico che attraversa la connessione (non è prevista una tariffa oraria per l'infrastruttura).

Il peering VPC è una connettività point-to-point e non supporta il routing transitivo. Se ad esempio è presente una connessione peering VPC tra VPC A e VPC B e tra VPC A e VPC C, un'istanza in VPC B non può transitare attraverso VPC A per raggiungere VPC C. Per instradare i pacchetti tra VPC B e VPC C, è necessario creare una connessione peering VPC diretta.

Su larga scala, quando sono presenti decine o centinaia di VPC, l'interconnessione tramite peering comporta la creazione di una mesh di centinaia o migliaia di connessioni peering, che sono difficili da gestire e dimensionare. È previsto un limite massimo di 125 connessioni peering per VPC.

Figura 2. Configurazione della rete tramite il peering VPC

Se si utilizza il peering VPC, è necessaria la connettività On-Premise (VPN e/o Direct Connect) a ciascun VPC. Le risorse in un VPC non possono raggiungere l'ambiente On-Premise tramite la connettività ibrida di un VPC in peering (Figura 2).

Il peering VPC è ottimale quando le risorse in un VPC devono comunicare con le risorse in un altro VPC, l'ambiente di entrambi i VPC è controllato e protetto e il numero di VPC da connettere è inferiore a 10 (per permettere la gestione individuale di ciascuna connessione). Il peering VPC offre il costo complessivo più basso rispetto ad altre opzioni per la connettività tra VPC.

Soluzione VPC di transito

I VPC di transito risolvono alcune carenze del peering VPC introducendo una progettazione di tipo hub-and-spoke per la connettività tra VPC. In una rete VPC di transito, un VPC centrale (VPC hub) si connette con ogni altro VPC (VPC spoke) attraverso una connessione VPN che in genere utilizza il protocollo BGP su IPsec. Il VPC centrale contiene istanze EC2 che eseguono appliance software che instradano il traffico in entrata verso le destinazioni previste utilizzando l'overlay VPN (Figura 3). Il peering VPC di transito offre i vantaggi seguenti:

- Il routing transitivo è abilitato utilizzando la rete VPN overlay, con una progettazione di tipo hub-and-spoke più semplice.
- Quando si utilizza software di terze parti nell'istanza EC2 nel VPC di transito hub, è possibile usufruire delle funzionalità di sicurezza avanzata (firewall/IPS/IDS di livello 7) del fornitore. Se i clienti utilizzano lo stesso software nell'ambiente On-Premise, ottengono il vantaggio di un'esperienza operativa e di monitoraggio unificata.

Figura 3. VPC di transito con CSR Cisco

Il VPC di transito presenta alcune problematiche, tra cui costi più elevati per l'esecuzione di appliance virtuali, velocità effettiva per VPC limitata (fino a 1,25 Gbps per tunnel VPN) e costi aggiuntivi di configurazione e gestione (i clienti devono gestire la disponibilità e la ridondanza delle istanze EC2).

Transit Gateway

AWS Transit Gateway offre una progettazione di tipo hub-and-spoke per la connessione di VPC e reti On-Premise come servizio completamente gestito, senza richiedere il provisioning di appliance

virtuali come i router CSR Cisco. Non è richiesto alcun overlay VPN e AWS gestisce la disponibilità elevata e la scalabilità.

Transit Gateway consente ai clienti di connettere migliaia di VPC. È possibile collegare tutte le soluzioni di connettività ibrida (connessioni VPN e Direct Connect) a una singola istanza di Transit Gateway, consolidando e controllando l'intera configurazione di routing AWS dell'organizzazione in un'unica posizione (Figura 4). Transit Gateway controlla la modalità di instradamento del traffico tra tutte le reti spoke connesse utilizzando tabelle di routing. Questo modello hub-and-spoke semplifica la gestione e riduce i costi operativi perché i VPC si connettono solo a Transit Gateway per accedere alle reti connesse.

Figura 4. Progettazione hub-and-spoke con AWS Transit Gateway

Transit Gateway è una risorsa regionale e consente di connettere migliaia di VPC all'interno della stessa regione AWS. È possibile creare più istanze di Transit Gateway per regione, ma non è possibile eseguire il peering delle istanze all'interno di una regione AWS ed è possibile stabilire la connessione a un massimo di tre istanze di Transit Gateway tramite una singola connessione Direct Connect per la connettività ibrida. Per questi motivi, è consigliabile limitare l'architettura a una sola istanza di Transit Gateway che connette tutti i VPC in una determinata regione e usare le tabelle di routing di Transit Gateway per isolare le istanze quando necessario. Un motivo valido per creare più Transit Gateway è limitare il raggio di diffusione di configurazioni errate.

L'istanza di Transit Gateway dell'organizzazione deve essere inserita nel relativo account dei servizi di rete. Ciò consente la gestione centralizzata da parte degli ingegneri di rete che gestiscono tale account. AWS Resource Access Manager (RAM) consente condividere un'istanza di Transit Gateway per connettere i VPC tra più account nell'organizzazione AWS all'interno della stessa regione. AWS RAM consente di condividere in modo semplice e sicuro le risorse AWS con qualsiasi account AWS o all'interno dell'organizzazione AWS. Per ulteriori informazioni, consulta il post di blog Automating AWS Transit Gateway attachments to a transit gateway in a central account.

Argomenti

Confronto tra Transit Gateway e VPC di transito

Transit Gateway offre diversi vantaggi rispetto a un VPC di transito:

- Transit Gateway elimina la complessità di gestione delle connessioni VPN con centinaia di VPC.
- Transit Gateway elimina la necessità di gestire e dimensionare le appliance software basate su EC2. AWS è responsabile della gestione di tutte le risorse necessarie per instradare il traffico.
- Transit Gateway elimina la necessità di gestire la disponibilità elevata, fornendo un'infrastruttura Multi-AZ altamente disponibile e ridondante.
- Transit Gateway migliora la larghezza di banda per la comunicazione tra VPC fino a una velocità massima di 50 Gbps per zona di disponibilità.
- Transit Gateway semplifica i costi degli utenti grazie a un semplice modello di calcolo dei costi per ora per GB trasferito.
- Transit Gateway riduce la latenza rimuovendo i proxy EC2 e la necessità di incapsulamento VPN.

Confronto tra Transit Gateway e peering VPC

Transit Gateway risolve la complessità legata alla creazione e alla gestione di più connessioni di peering VPC su larga scala. Sebbene ciò renda TGW una buona opzione di default per la maggior parte delle architetture di rete, il peering VPC è comunque una scelta valida grazie ai seguenti vantaggi che offre rispetto a TGW:

- Costi inferiori: con il peering VPC si paga solo il costo di trasferimento dei dati. Transit Gateway prevede una tariffa oraria per collegamento oltre alle tariffe per il trasferimento dei dati.
- Nessun limite di larghezza di banda: con Transit Gateway, la larghezza di banda massima (burst) per connessione VPC è di 50 Gbps. Il peering VPC non prevede una larghezza di banda aggregata. I limiti delle prestazioni di rete delle singole istanze e i limiti di flusso (10 Gbps all'interno di un gruppo di collocazione e 5 Gbps negli altri casi) si applicano a entrambe le opzioni. Solo il peering VPC supporta i gruppi di collocazione.
- Latenza: a differenza del peering VPC, Transit Gateway costituisce un hop aggiuntivo tra VPC.
- Compatibilità dei gruppi di sicurezza: il riferimento ai gruppi di sicurezza funziona con il peering VPC all'interno della regione. Attualmente non funziona con Transit Gateway.

All'interno della configurazione della zona di destinazione è possibile utilizzare il peering VPC in combinazione con il modello hub-and-spoke abilitato da Transit Gateway.

AWS PrivateLink

I clienti possono voler esporre privatamente un'applicazione o un servizio che si trova in un VPC (fornitore di servizi) ad altri VPC consumer all'interno di una regione AWS in modo che solo i VPC consumer avviino connessioni al VPC fornitore di servizi. Un esempio è la possibilità per le applicazioni private di accedere alle API dei fornitore di servizi.

Per utilizzare AWS PrivateLink è necessario creare un sistema Network Load Balancer per l'applicazione nel VPC, nonché creare una configurazione del servizio endpoint VPC che punti a tale bilanciatore del carico. Un consumer di servizi crea quindi un endpoint di interfaccia per il servizio. In questo modo, nella sottorete viene creata un'interfaccia di rete elastica con un indirizzo IP privato che funge da punto di ingresso per il traffico destinato al servizio. Non è necessario che il consumer e il servizio si trovino nello stesso VPC. Se il VPC è diverso, i VPC consumer e fornitore di servizi possono avere intervalli di indirizzi IP sovrapposti. Oltre a creare l'endpoint VPC di interfaccia per accedere ai servizi in altri VPC, è possibile creare endpoint VPC di interfaccia per accedere privatamente ai servizi AWS supportati tramite AWS PrivateLink (Figura 5).

Figura 5. AWS PrivateLink

La scelta tra Transit Gateway, peering VPC e AWS PrivateLink dipende dalla connettività.

AWS PrivateLink: è consigliabile utilizzare AWS PrivateLink quando si dispone di una configurazione client/server in cui si desidera permettere a uno o più VPC consumer l'accesso unidirezionale a uno specifico servizio o set di istanze nel VPC fornitore di servizi. Solo i client nel VPC consumer possono avviare una connessione al servizio nel VPC fornitore di servizi. Questa è una buona opzione anche quando il client e i server nei due VPC hanno indirizzi IP sovrapposti in quanto AWS PrivateLink utilizza le interfacce di rete elastica nel VPC client in modo tale che non ci siano conflitti di IP con il fornitore di servizi. È possibile accedere agli endpoint AWS PrivateLink tramite peering VPC, VPN e AWS Direct Connect.

Peering VPC e Transit Gateway: è consigliabile utilizzare il peering VPC e Transit Gateway quando si vuole abilitare la connettività IP di livello 3 tra i VPC.

L'architettura conterrà una combinazione di queste tecnologie al fine di soddisfare i diversi casi d'uso. Tutti questi servizi possono essere combinati e gestiti tra loro. Ad esempio, AWS PrivateLink per la gestione della connettività client-server in stile API, il peering VPC per la gestione dei requisiti di connettività diretta in cui possono essere necessari gruppi di collocazione nella regione o connettività

tra regioni e Transit Gateway per semplificare la connettività dei VPC su larga scala, nonché il consolidamento edge per la connettività ibrida.

Condivisione Amazon VPC

La condivisione dei VPC è utile quando l'isolamento di rete tra i team non richiede una gestione rigorosa da parte del proprietario del VPC, mentre gli utenti e le autorizzazioni a livello di account la richiedono. Con un VPC condiviso, più account AWS creano le risorse delle applicazioni (ad esempio le istanze Amazon EC2) in istanze di Amazon VPC condivise gestite centralmente. In questo modello l'account che possiede il VPC (proprietario) condivide una o più sottoreti con altri account (partecipanti). Una volta condivisa una sottorete, i partecipanti possono visualizzare, creare, modificare ed eliminare le proprie risorse delle applicazioni nelle sottoreti condivise. Non possono invece visualizzare, modificare o eliminare le risorse che appartengono ad altri partecipanti o al proprietario del VPC. La sicurezza tra le risorse nei VPC condivisi viene gestita utilizzando gruppi di sicurezza e liste di controllo degli accessi di rete per la sottorete.

Vantaggi della condivisione dei VPC:

- Progettazione semplificata: nessuna complessità per la connettività tra VPC
- Meno VPC gestiti
- Segregazione dei compiti tra i team di rete e i proprietari delle applicazioni
- Migliore utilizzo degli indirizzi IPv4
- Costi inferiori: nessun costo di trasferimento dei dati tra istanze appartenenti ad account diversi all'interno della stessa zona di disponibilità

Nota: quando si condivide una sottorete con più account, i partecipanti devono avere un certo livello di collaborazione poiché condividono lo spazio IP e le risorse di rete. Se necessario, è possibile scegliere di condividere una sottorete diversa per ogni account partecipante. Una sottorete per partecipante consente alla lista di controllo degli accessi di rete di fornire l'isolamento di rete oltre ai gruppi di sicurezza.

La maggior parte delle architetture dei clienti contiene più VPC, molti dei quali sono condivisi con due o più account. È possibile utilizzare Transit Gateway e il peering VPC per connettere i VPC condivisi. Supponiamo ad esempio di avere 10 applicazioni. Ogni applicazione richiede il proprio account AWS. Le app possono essere classificate in due portfolio di applicazioni (le app all'interno dello stesso portfolio hanno requisiti di rete simili, App 1-5 nella categoria relativa al marketing e App 6-10 nella categoria relativa alle vendite).

È possibile avere un VPC per portfolio di applicazioni (due VPC in totale) e il VPC viene condiviso con i diversi account dei proprietari delle applicazioni all'interno di tale portfolio. I proprietari delle app distribuiscono le app nel rispettivo VPC condiviso (in questo caso, nelle diverse sottoreti per la segmentazione e l'isolamento delle route di rete tramite liste di controllo degli accessi di rete). I due VPC condivisi sono connessi tramite Transit Gateway. Con questa configurazione, è possibile passare dal dover connettere 10 VPC a soli 2 (Figura 6).

Figura 6. Configurazione di esempio – VPC condiviso

> **Note**
>
> I partecipanti alla condivisione dei VPC non possono creare tutte le risorse AWS in una sottorete condivisa. Per ulteriori informazioni, consulta la pagina relativa alle limitazioni di Amazon VPC.

Connettività ibrida

Questa sezione è incentrata sulla connessione sicura delle risorse cloud con i data center On-Premise. Sono disponibili due approcci per abilitare la connettività ibrida:

1. Connettività uno-a-uno: con questa configurazione viene creata una connessione VPN e/o un'interfaccia virtuale privata Direct Connect per ogni VPC. A tale scopo è possibile utilizzare il gateway virtuale privato (VGW, Virtual Private Gateway). Questa opzione è ideale per un numero limitato di VPC, mentre quando un cliente esegue il dimensionamento dei VPC, la gestione della connettività ibrida per ogni singolo VPC può diventare complessa.
2. Consolidamento dell'edge: con questa configurazione i clienti consolidano la connettività IT ibrida per più VPC in un unico endpoint. Tutti i VPC condividono le connessioni ibride. A tale scopo è possibile utilizzare AWS Transit Gateway e il gateway Direct Connect.

Argomenti

VPN

Figura 7. Opzioni di terminazione di AWS VPN

Ci sono tre modi per configurare una rete VPN in AWS:

1. Consolidamento della connettività VPN in Transit Gateway: questa opzione utilizza il collegamento del gateway di transito alla VPN in Transit Gateway. Transit Gateway supporta la terminazione IPsec per la connessione Site-to-Site VPN. I clienti possono creare tunnel VPN per Transit Gateway e accedere ai VPC collegati. Transit Gateway supporta connessioni VPN sia statiche che dinamiche basate su BGP. Transit Gateway supporta anche la strategia ECMP (Equal-Cost Multi-Path) per i collegamenti alla VPN. Ogni connessione VPN ha una velocità effettiva massima di 1,25 Gbps e l'abilitazione di ECMP permette di aggregare la velocità effettiva tra le connessioni VPN. Con questa opzione si pagano sia i costi per Transit Gateway che quelli per AWS VPN. Questa opzione è consigliata per la connettività VPN. Per ulteriori informazioni, consulta la panoramica di AWS VPN.

2. Terminazione della rete VPN nell'istanza EC2: questa opzione viene utilizzata dai clienti in ambienti edge quando desiderano un set di caratteristiche software di un fornitore specifico (come Cisco DMVPN o GRE) o la coerenza operativa tra diverse distribuzioni VPN. È possibile utilizzare la progettazione del VPC di transito per il consolidamento edge, ma è importante ricordare che tutte le considerazioni chiave della sezione relativa alla connettività da VPC a VPC per il VPC di transito sono applicabili alla connettività VPN ibrida. L'utente è responsabile della gestione della disponibilità elevata e paga i costi per le istanze EC2, nonché quelli delle eventuali licenze software dei fornitori.
3. Terminazione della rete VPN in un gateway virtuale privato (VGW): questa opzione consente una progettazione della connettività uno-a-uno in cui si crea una connessione VPN (composta da una coppia di tunnel VPN ridondanti) per VPC. Si tratta di un ottimo modo per iniziare a lavorare con la connettività VPN in AWS, ma man mano che si dimensiona il numero di VPC, la progettazione del consolidamento edge con Transit Gateway diventa un'opzione migliore. La velocità effettiva della rete VPN verso un VPC è limitata a 1,25 Gbps e il bilanciamento del carico ECMP non è supportato. Dal punto di vista dei prezzi, si pagano solo i costi per AWS VPN, senza alcun costo per l'esecuzione di un VGW. Per ulteriori informazioni, consulta Prezzi di AWS VPN e AWS VPN nel gateway virtuale privato.

Direct Connect

Sebbene la rete VPN in Internet sia un'ottima opzione per iniziare, la connettività Internet può non essere affidabile per il traffico di produzione. A causa di questa inaffidabilità, molti clienti scelgono AWS Direct Connect, che consente una connettività tra i data center dei clienti e AWS con fibra dedicata, costante, a bassa latenza e con larghezza di banda elevata. È possibile utilizzare AWS Direct Connect per la connessione ai VPC in quattro modi:

Figura 8. Quattro modalità di connessione dei data center On-Premise alla zona di destinazione

- Creazione di un'interfaccia virtuale (VIF) privata per un VGW collegato a un VPC: è possibile creare 50 VIF per connessione Direct Connect, così da poter stabilire la connessione a un massimo di 50 VPC (un'interfaccia VIF fornisce la connettività a un VPC). Viene creato un peering BGP per VPC. La connettività con questa configurazione è limitata alla regione AWS in cui si trova la posizione Direct Connect. La mappatura uno-a-uno tra VIF e VPC (e la mancanza di accesso globale) rende questo metodo il meno adatto per l'accesso ai VPC nella zona di destinazione.
- Creazione di un'interfaccia virtuale privata per un gateway Direct Connect associato a più VGW (ogni VGW è collegato a un VPC): un gateway Direct Connect può connettersi a un massimo di 10

VGW in tutto il mondo (esclusa la Cina) in qualsiasi account AWS. Si tratta di un'ottima opzione se una zona di destinazione è costituita da un numero limitato di VPC (dieci o meno VPC) e/o è necessario disporre di accesso globale. Viene creato un peering BGP per gateway Direct Connect per connessione Direct Connect. Il gateway Direct Connect serve solo per il flusso di traffico nord/sud e non consente la connettività da VPC a VPC.

- Creazione di un'interfaccia VIF di transito per un gateway Direct Connect associato a Transit Gateway: è possibile associare un'istanza di Transit Gateway a un gateway Direct Connect tramite una connessione Direct Connect dedicata o ospitata in esecuzione a 1 Gbps o più. Questa opzione permette di connettere il data center On-Premise a un massimo di tre istanze di Transit Gateway (che possono connettersi a migliaia di VPC) in diverse regioni AWS e diversi account AWS tramite un'interfaccia VIF e un peering BGP. Si tratta della configurazione più semplice tra le quattro opzioni per la connessione di più VPC su larga scala, ma è necessario tenere presenti le limitazioni di Transit Gateway. Un limite fondamentale è la possibilità di annunciare solo 20 intervalli CIDR da un'istanza di Transit Gateway a un router On-Premise attraverso l'interfaccia VIF di transito. Le opzioni 1 e 2 richiedono il pagamento dei costi di Direct Connect. L'opzione 3 richiede anche il pagamento dei costi di collegamento del gateway di transito e di trasferimento dei dati. Per ulteriori informazioni, consulta la documentazione sulle associazioni di Transit Gateway in Direct Connect.
- Creazione di una connessione VPN a Transit Gateway tramite interfaccia virtuale pubblica Direct Connect: un'interfaccia virtuale pubblica permette di accedere a tutti gli endpoint e i servizi pubblici di AWS utilizzando gli indirizzi IP pubblici. Quando si crea un collegamento alla VPN in Transit Gateway, si ottengono due indirizzi IP pubblici per la terminazione VPN nell'estremità AWS. Questi IP pubblici sono raggiungibili tramite l'interfaccia virtuale pubblica. È possibile creare il numero desiderato di connessioni VPN a diverse istanze di Transit Gateway tramite interfaccia virtuale pubblica. Quando si crea un peering BGP nell'interfaccia virtuale pubblica, AWS annuncia l'intero intervallo IP pubblico di AWS al router. Per permettere solo un determinato tipo di traffico (ad esempio per permettere solo il traffico verso gli endpoint di terminazione VPN), è consigliabile utilizzare un firewall On-Premise. Questa opzione può essere utilizzata per crittografare Direct Connect a livello di rete.

Sebbene la terza opzione (interfaccia virtuale di transito per il gateway Direct Connect) possa sembrare l'opzione migliore perché consente di consolidare tutta la connettività On-Premise per una determinata regione AWS in un unico punto (Transit Gateway) utilizzando una singola sessione BGP per connessione Direct Connect, considerando alcuni limiti e caratteristiche dell'opzione 3, è probabile che i clienti scelgano entrambe le opzioni 2 e 3 per i loro requisiti di connettività della zona di destinazione. La Figura 9 illustra una configurazione di esempio in cui vengono utilizzate un'interfaccia virtuale di transito come metodo di default per la connessione ai VPC e un'interfaccia

virtuale privata per un caso d'uso edge in cui è necessario trasferire quantità molto grandi di dati da un controller di dominio On-Premise al VPC multimediale. L'interfaccia virtuale privata viene utilizzata per evitare i costi di trasferimento dei dati di Transit Gateway. Come best practice, è consigliabile disporre di almeno due connessioni in due diverse posizioni Direct Connect per la massima ridondanza, per un totale di quattro connessioni. Si crea un'interfaccia VIF per connessione per un totale di quattro VIF private e quattro VIF di transito. Si crea anche una VPN come connettività di backup alle connessioni AWS Direct Connect.

Figura 9. Esempio di architettura di riferimento per la connettività ibrida

Utilizzare l'account dei servizi di rete per creare risorse Direct Connect che consentono la delimitazione dei confini amministrativi di rete. La connessione Direct Connect, il gateway Direct Connect e Transit Gateway possono trovarsi tutti in un account dei servizi di rete. Per condividere la connettività di AWS Direct Connect con la zona di destinazione, è sufficiente condividere Transit Gateway tramite RAM con altri account.

Uscita centralizzata per l'accesso a Internet

Quando si distribuiscono le applicazioni nella zona di destinazione, molte app richiedono solo l'accesso a Internet in uscita (ad esempio per il download di librerie, patch o aggiornamenti del sistema operativo). È possibile ottenere questo risultato utilizzando di preferenza un gateway NAT (Network Address Translation), o in alternativa, un'istanza EC2 configurata con SNAT (Source NAT) come hop successivo per tutti gli accessi Internet in uscita. Le applicazioni interne si trovano in sottoreti private, mentre le istanze NAT EC2 e del gateway NAT si trovano in una sottorete pubblica.

Uso del gateway NAT

L'implementazione di un gateway NAT in ogni VPC spoke può diventare costosa perché si paga una tariffa oraria per ogni gateway NAT implementato (consulta la pagina Prezzi di Amazon VPC), quindi la centralizzazione può essere un'opzione valida. A tale scopo, è necessario creare un VPC in uscita nell'account dei servizi di rete e instradare tutto il traffico in uscita dai VPC spoke tramite un gateway NAT che si trova in questo VPC utilizzando Transit Gateway, come illustrato nella Figura 10.

Nota: quando si centralizza il gateway NAT utilizzando Transit Gateway, si paga un costo aggiuntivo per l'elaborazione dei dati di Transit Gateway, rispetto all'approccio decentralizzato di esecuzione di un gateway NAT in ogni VPC. In alcuni ambienti edge, quando si inviano quantità molto grandi di dati tramite il gateway NAT da un VPC, può essere più conveniente mantenere il gateway NAT in locale nel VPC per evitare il costo di elaborazione dei dati di Transit Gateway.

Figura 10. Gateway NAT centralizzato con l'utilizzo di Transit Gateway (panoramica)

Figura 11. Gateway NAT centralizzato con l'utilizzo di Transit Gateway (progettazione della tabella di routing)

Con questa configurazione, i collegamenti VPC spoke sono associati alla tabella di routing 1 (RT1) e vengono propagati alla tabella di routing 2 (RT2). Abbiamo aggiunto esplicitamente una route blackhole per impedire ai due VPC di comunicare tra loro. Se si desidera permettere la comunicazione tra VPC, è possibile rimuovere la voce di routing "10.0.0.0/8 -> Blackhole" da RT1. Si permette così la comunicazione tramite il gateway NAT. È inoltre possibile propagare i collegamenti VPC spoke a RT1 (o in alternativa, è possibile utilizzare una tabella di routing e associare/propagare tutto in tale tabella), consentendo il flusso di traffico diretto tra i VPC tramite Transit Gateway.

Aggiungiamo una route statica in RT1 che instrada tutto il traffico al VPC in uscita. A causa di questa route statica, Transit Gateway invia tutto il traffico Internet attraverso le interfacce di rete elastica nel VPC in uscita. Una volta nel VPC in uscita, il traffico segue le regole definite nella tabella di routing della sottorete in cui sono presenti le interfacce di rete elastica di Transit Gateway. Aggiungiamo una route nella tabella di routing della sottorete che instrada tutto il traffico verso il gateway NAT. La tabella di routing della sottorete del gateway NAT ha il Gateway Internet (IGW) come hop successivo. Per consentire il flusso di ritorno del traffico, è necessario aggiungere una voce statica nella tabella di routing della sottorete del gateway NAT che instrada tutto il traffico verso i VPC spoke a Transit Gateway come hop successivo.

Disponibilità elevata

Per la disponibilità elevata, è consigliabile utilizzare due gateway NAT (uno in ciascuna zona di disponibilità). All'interno di una zona di disponibilità, il gateway NAT ha un contratto di servizio (SLA, Service Level Agreement) che prevede una disponibilità del 99,9%. La ridondanza in caso di errore dei componenti all'interno di una zona di disponibilità è gestita da AWS in base al contratto SLA.
Il traffico non viene trasmesso durante lo 0,1% del tempo in cui il gateway NAT può non essere disponibile in una zona di disponibilità. In caso di problema di un'intera zona di disponibilità, l'endpoint Transit Gateway e il gateway NAT in tale zona di disponibilità non funzionano e tutto il traffico viene trasmesso attraverso gli endpoint Transit Gateway e del gateway NAT nell'altra zona di disponibilità.

Sicurezza

La sicurezza si ottiene grazie a gruppi di sicurezza nelle istanze di origine, route blackhole nelle tabelle di routing di Transit Gateway e lista di controllo degli accessi di rete della sottorete in cui si trova il gateway NAT.

Scalabilità

Un gateway NAT può supportare fino a 55.000 connessioni simultanee a ciascuna destinazione univoca. Dal punto di vista della velocità effettiva, i limiti dipendono dalle prestazioni del gateway NAT. Transit Gateway non è un bilanciatore del carico e non distribuisce il traffico in modo uniforme nel gateway NAT in più zone di disponibilità. Il traffico attraverso Transit Gateway rimane all'interno di una zona di disponibilità, se possibile. Se l'istanza EC2 da cui ha origine il traffico si trova nella zona di disponibilità 1, il traffico esce dall'interfaccia di rete elastica di Transit Gateway nella stessa zona di disponibilità 1 nel VPC in uscita e passa all'hop successivo in base alla tabella di routing della sottorete in cui si trova l'interfaccia di rete elastica. Per un elenco completo di regole, consulta la pagina relativa a regole e limiti del gateway NAT.

Per ulteriori informazioni, consulta il post di blog Creating a single internet exit point from multiple VPCs Using AWS Transit Gateway.

Utilizzo di un'istanza EC2 per l'uscita centralizzata

L'utilizzo di un'appliance firewall basata su software (in EC2) disponibile in Marketplace AWS come punto di uscita è simile alla configurazione del gateway NAT. È possibile scegliere questa opzione se si desidera utilizzare le funzionalità IPS/IDS (Intrusion Prevention/Detection System) e firewall di livello 7 delle varie offerte dei fornitori.

Nella Figura 12 il gateway NAT viene sostituito con un'istanza EC2 (con la funzionalità SNAT abilitata nell'istanza EC2). Ci sono alcune considerazioni chiave in merito a questa opzione:

Disponibilità elevata

In questa configurazione, l'utente è responsabile del monitoraggio dell'istanza EC2, del rilevamento degli errori e della sostituzione dell'istanza EC2 con un'istanza di backup/standby. La maggior parte dei fornitori AWS offre un'automazione preconfigurata per il proprio software distribuito in questa configurazione. Tale automazione consente quanto segue:

- Rilevamento degli errori dell'istanza EC2-1 primaria
- Modifica della tabella di routing "Route Table Egx 1" in modo da instradare tutto il traffico all'istanza EC2-2 di backup in caso di errore dell'istanza primaria. Ciò deve avvenire anche per le sottoreti nella zona di disponibilità 2.

Figura 12. Soluzione NAT centralizzata con l'utilizzo di istanze EC2 e Transit Gateway

Scalabilità

Transit Gateway non è un bilanciatore del carico e non distribuisce il traffico in modo uniforme tra le istanze nelle due zone di disponibilità. Il traffico attraverso Transit Gateway rimane all'interno di una zona di disponibilità, se possibile. I limiti dipendono dalle capacità di larghezza di banda di una singola istanza EC2. È possibile il dimensionamento verticale dell'istanza EC2 quando l'utilizzo aumenta.

Se il fornitore scelto per l'ispezione del traffico in uscita non supporta l'automazione per il rilevamento degli errori o se è necessario il dimensionamento orizzontale, è possibile utilizzare una progettazione alternativa. Con questo tipo di progettazione (Figura 13), non viene creato un collegamento VPC in

Transit Gateway per il VPC in uscita, ma vengono creati un collegamento alla VPN IPsec e una rete VPN IPsec da Transit Gateway verso le istanze EC2 utilizzando BGP per lo scambio di route.

Vantaggi

- Il rilevamento degli errori e il reinstradamento del traffico sono gestiti da BGP. Non è richiesta l'automazione della tabella di routing della sottorete VPC.
- È possibile utilizzare la strategia ECMP BGP per il bilanciamento del carico del traffico tra più istanze EC2 ed è possibile il dimensionamento orizzontale.

Figura 13. Soluzione NAT centralizzata con l'utilizzo di istanze EC2 e VPN di Transit Gateway

Considerazioni chiave

- Costi di gestione della rete VPN nelle istanze EC2.
- La larghezza di banda di Transit Gateway è limitata a 1,25 Gbps per tunnel VPN. Con la strategia ECMP Transit Gateway può supportare fino a 50 Gbps di larghezza di banda VPN totale. Le funzionalità VPN e di elaborazione dei pacchetti dell'appliance del fornitore possono costituire un fattore limitante.
- Questo tipo di progettazione presuppone che l'istanza EC2 firewall funzioni con la stessa interfaccia di rete elastica per il traffico in entrata e in uscita.
- Se si abilita il bilanciamento del carico ECMP del traffico tra più istanze EC2, è necessario applicare la funzionalità SNAT al traffico nell'istanza EC2 per consentire la simmetria del flusso di ritorno, il che significa che la destinazione non conoscerà la vera origine.

Sicurezza di rete centralizzata per il traffico da VPC a VPC e da ambiente On-Premise a VPC

AWS fornisce gruppi di sicurezza e liste di controllo degli accessi di rete per le sottoreti per implementare la sicurezza di rete all'interno della zona di destinazione. Si tratta di firewall di livello 4. Ci possono essere scenari in cui un cliente desidera implementare una soluzione firewall/IPS/IDS di livello 7 all'interno della propria zona di destinazione per ispezionare il flusso del traffico tra VPC o tra un data center On-Premise e un VPC. A tale scopo è possibile utilizzare Transit Gateway e appliance software di terze parti in esecuzione nelle istanze EC2. Utilizzando l'architettura nella Figura 14 è possibile abilitare il flusso del traffico da VPC a VPC e da ambiente On-Premise a VPC tramite le istanze EC2. La configurazione è simile a quella già discussa nella Figura 12, ma viene inoltre rimossa la route blackhole nella tabella di routing 1 per permettere il flusso di traffico VPC interno e viene abilitato il collegamento della VPN e/o del gateway virtuale Direct Connect alla tabella di routing 1 per permettere il flusso di traffico ibrido. In questo modo si consente il flusso di tutto il traffico proveniente dagli spoke verso il VPC in uscita prima dell'invio alla destinazione. È necessario disporre di route statiche nella tabella di routing della sottorete VPC in uscita (dove si trovano le appliance EC2 firewall) per l'invio del traffico destinato ai VPC spoke e di routing CIDR On-Premise attraverso Transit Gateway dopo l'ispezione del traffico.

> **Note**
>
> Le informazioni di routing non vengono propagate dinamicamente da Transit Gateway alla tabella di routing della sottorete e devono essere inserite staticamente. È previsto un limite flessibile di 50 route statiche in una tabella di routing di una sottorete.

Figura 14. Controllo del traffico da VPC a VPC e da VPC ad ambiente On-Premise

Considerazioni chiave per l'invio di traffico alle istanze EC2 per l'ispezione in linea:

- Costi aggiuntivi per l'elaborazione dei dati di Transit Gateway
- Necessità di passaggio del traffico attraverso due hop aggiuntivi (istanza EC2 e Transit Gateway)
- Potenziali colli di bottiglia in termini di larghezza di banda e prestazioni
- Complessità aggiuntiva legata a manutenzione, gestione e dimensionamento delle istanze EC2:

- Rilevamento di errori e failover all'istanza di standby
- Monitoraggio dell'utilizzo e dimensionamento orizzontale/verticale
- Configurazione firewall, gestione delle patch
- Conversione SNAT (Source Network Address Translation) del traffico durante il bilanciamento del carico per consentire un flusso simmetrico

È consigliabile essere selettivi in merito al traffico che passa attraverso le istanze EC2. Un modo per procedere consiste nel definire le zone di sicurezza e ispezionare il traffico tra zone non attendibili. Una zona non attendibile può essere un sito remoto gestito da una terza parte, un VPC di un fornitore non controllato personalmente o di cui non si ha fiducia oppure un VPC sandbox o di sviluppo che prevede un framework di sicurezza meno rigoroso rispetto al resto dell'ambiente. La Figura 15 illustra una situazione in cui è consentito il flusso di traffico diretto tra reti attendibili durante l'ispezione del flusso di traffico da/verso reti non attendibili utilizzando istanze EC2 in linea. Nell'esempio sono state create tre zone:

- Zona non attendibile: riguarda tutto il traffico proveniente dalla rete VPN verso il sito remoto non attendibile o il VPC del fornitore di terze parti.
- Zona di produzione: contiene il traffico proveniente dal VPC di produzione e dal controller di dominio del cliente On-Premise.
- Zona di sviluppo: contiene il traffico proveniente dai due VPC di sviluppo.

Di seguito sono riportate le regole di esempio definite per la comunicazione tra zone:

1. Zona non attendibile – zona di produzione: comunicazione non permessa
2. Zona di produzione - zona di sviluppo: comunicazione permessa tramite appliance firewall EC2 nel VPC in uscita
3. Zona non attendibile – zona di sviluppo: comunicazione permessa tramite appliance firewall EC2 nel VPC in uscita
4. Zona di produzione – zona di produzione e zona di sviluppo – zona di sviluppo: comunicazione diretta tramite Transit Gateway

In questa configurazione sono presenti tre zone di sicurezza, ma possono essercene di più. È possibile utilizzare più tabelle di routing e route blackhole per ottenere un isolamento della sicurezza e un flusso di traffico ottimale. La scelta delle zone appropriate dipende dalla strategia generale di

progettazione della zona di destinazione (struttura degli account, progettazione VPC). È possibile creare zone per abilitare l'isolamento tra business unit, applicazioni, ambienti e così via.

In questo esempio la rete VPN remota non attendibile viene terminata in Transit Gateway e tutto il traffico viene inviato alle appliance firewall software in EC2 per l'ispezione. In alternativa, è possibile terminare le reti VPN direttamente nelle istanze EC2 invece che in Transit Gateway. Con questo approccio, il traffico VPN non attendibile non interagisce mai direttamente con Transit Gateway. Il numero di hop nel flusso di traffico si riduce di uno ed è possibile risparmiare sui costi di AWS VPN. Per consentire gli scambi di route dinamici (per consentire a Transit Gateway di conoscere il routing CIDR della rete VPN remota tramite BGP), le istanze del firewall devono essere connesse a Transit Gateway tramite VPN. Nel modello di collegamento TGW nativo, è necessario aggiungere route statiche nella tabella di routing TGW per il routing CIDR VPN con l'hop successivo come VPC in uscita/di sicurezza. Nella configurazione esaminata (Figura 15) è presente una route di default verso il VPC in uscita per tutto il traffico, quindi non è necessario aggiungere esplicitamente alcuna route statica specifica. Con questo approccio si passa da un endpoint di terminazione VPN di Transit Gateway completamente gestito a un'istanza EC2 autogestita, aggiungendo costi di gestione VPN e un carico aggiuntivo nell'istanza EC2 in termini di calcolo e memoria.

Figura 15. Isolamento del traffico utilizzando Transit Gateway e definendo le zone di sicurezza

DNS

Quando si avvia un'istanza in un VPC non di default, AWS fornisce all'istanza un nome host DNS privato (e potenzialmente un nome host DNS pubblico), in base agli attributi DNS specificati per il VPC e al fatto che l'istanza disponga di un indirizzo IPv4 pubblico. Quando l'attributo "enableDnsSupport" è impostato su true, si ottiene una risoluzione DNS all'interno del VPC da Route 53 Resolver (con offset +2 per l'indirizzo IP nel routing CIDR del VPC). Di default, Route 53 Resolver risponde alle query DNS per i nomi di dominio VPC, ad esempio i nomi di dominio per le istanze EC2 o i bilanciatori del carico di Elastic Load Balancing. Con il peering VPC, gli host in un VPC possono risolvere i nomi host DNS pubblici in indirizzi IP privati per le istanze nei VPC in peering, a condizione che sia abilitata l'opzione per farlo. Lo stesso vale per i VPC connessi tramite AWS Transit Gateway. Per ulteriori informazioni, consulta Abilitazione del supporto per la risoluzione DNS per una connessione peering VPC.

Per mappare le istanze a un nome di dominio personalizzato, è possibile utilizzare Amazon Route 53 per creare un record di mappatura da DNS a IP personalizzato. Una zona ospitata di Amazon Route 53 è un container che contiene informazioni su come si desidera che Amazon Route 53 risponda alle query DNS per un dominio e i relativi sottodomini. Le zone ospitate pubbliche contengono informazioni DNS risolvibili nella rete Internet pubblica, mentre le zone ospitate private sono un'implementazione specifica che presenta informazioni solo ai VPC che sono stati collegati alla zona ospitata privata specifica. In una configurazione di zona di destinazione in cui sono presenti più VPC/account, è possibile associare una singola zona ospitata privata a più VPC in diversi account AWS e regioni. Gli host finali nei VPC utilizzano il rispettivo IP di Route 53 Resolver (con offset +2 per l'indirizzo IP nel routing CIDR del VPC) come server dei nomi per le query DNS. Route 53 Resolver nel VPC accetta le query DNS solo dalle risorse all'interno di un VPC.

DNS ibrido

Il coordinamento della risoluzione DNS tra la configurazione della zona di destinazione AWS e le risorse On-Premise è uno degli elementi più critici in una rete ibrida. I clienti che implementano ambienti ibridi di solito dispongono già di un sistema di risoluzione DNS e desiderano una soluzione DNS che funzioni in combinazione con il loro sistema corrente. Quando si integra il sistema DNS per i VPC in una regione AWS con il DNS per la rete, sono in genere necessari un endpoint in entrata Route 53 Resolver (per le query DNS che vengono inoltrate al VPC) e un endpoint in uscita Route 53 Resolver (per le query che vengono inoltrate dai VPC alla rete). Come illustrato nella Figura 16, è possibile configurare gli endpoint Resolver in uscita per inoltrare le query ricevute dalle istanze EC2

nei VPC ai server DNS nella rete. Per inoltrare le query selezionate, da un VPC all'ambiente On-Premise, è necessario creare regole di Route 53 Resolver che specificano i nomi di dominio per le query DNS da inoltrare (ad esempio example.com) e gli indirizzi IP dei resolver DNS nella rete in cui si desidera inoltrare le query. Per le query in entrata dall'ambiente On-Premise alle zone ospitate di Route 53, i server DNS nella rete possono inoltrare le query agli endpoint Resolver in entrata in un VPC specificato.

Figura 16. Risoluzione DNS ibrida tramite Route 53 Resolver

In questo modo i resolver DNS On-Premise possono risolvere facilmente i nomi di dominio per le risorse AWS, come le istanze EC2 o i record in una zona ospitata privata di Route 53 associata al VPC.

Non è consigliabile creare endpoint Resolver Route 53 in ogni VPC della zona di destinazione. È preferibile centralizzarli in un VPC in uscita centrale (nell'account dei servizi di rete). Questo approccio permette una migliore gestione mantenendo bassi i costi (viene addebitata una tariffa oraria per ogni endpoint in entrata/uscita creato). L'endpoint in entrata e in uscita centralizzato viene condiviso con il resto della zona di destinazione.

Risoluzione in uscita: utilizza l'account dei servizi di rete per scrivere le regole del resolver (in base alle quali le query DNS verranno inoltrate ai server DNS On-Premise). Utilizzando Resource Access Manager (RAM), condividi queste regole di Route 53 Resolver con più account (associandole ai VPC negli account). Le istanze EC2 nei VPC spoke possono inviare query DNS a Route 53 Resolver e il servizio Route 53 Resolver inoltrerà queste query al server DNS On-Premise tramite gli endpoint Route 53 Resolver in uscita nel VPC in uscita. Non è necessario eseguire il peering dei VPC spoke al VPC in uscita o connetterli tramite Transit Gateway. Non utilizzare l'IP dell'endpoint resolver in uscita come DNS primario nei VPC spoke. I VPC spoke devono utilizzare Route 53 Resolver (per l'offset del routing CIDR del VPC) all'interno del VPC.

Figura 17. Centralizzazione degli endpoint Route 53 Resolver nel VPC in uscita

Risoluzione DNS in entrata: crea endpoint in entrata Route 53 Resolver in un VPC centralizzato e associa tutte le zone ospitate private nella zona di destinazione al VPC centralizzato. Per ulteriori informazioni, consulta la pagina relativa all'associazione di più VPC a una zona ospitata privata. Più zone ospitate private (PHZ, Private Hosted Zone) associate a un VPC non possono sovrapporsi. Come illustrato nella Figura 17, l'associazione della zona ospitata privata con il VPC centralizzato consente ai server On-Premise di risolvere il nome DNS per qualsiasi voce in qualsiasi zona ospitata

privata (associata a VPC centrale) utilizzando l'endpoint in entrata nel VPC centralizzato. Per ulteriori informazioni sulle configurazioni DNS ibride, consulta Centralized DNS management of hybrid cloud with Amazon Route 53 and AWS Transit Gateway e Opzioni DNS cloud ibride per Amazon VPC.

Accesso centralizzato agli endpoint privati VPC

Un endpoint VPC permette di stabilire una connessione privata tra il VPC e i servizi AWS supportati senza richiedere un gateway Internet o un dispositivo NAT. Con questo endpoint di interfaccia, le istanze nel VPC non richiedono indirizzi IP pubblici per comunicare con gli endpoint del servizio AWS. Il traffico tra il VPC e altri servizi non esce dalla dorsale di rete (backbone) della rete AWS. Attualmente è possibile effettuare il provisioning di due tipi di endpoint: endpoint di interfaccia (basati su AWS PrivateLink) ed endpoint gateway. Il provisioning degli endpoint gateway è gratuito e non c'è un caso d'uso valido per la centralizzazione.

Endpoint VPC di interfaccia

Un endpoint di interfaccia è costituito da una o più interfacce di rete elastiche con un indirizzo IP privato che funge da punto di ingresso per il traffico destinato a un servizio AWS supportato. Quando si effettua il provisioning di un endpoint di interfaccia, gli utenti sostengono un costo per ogni ora di esecuzione dell'endpoint. Di default viene creato un endpoint di interfaccia in ogni VPC da cui si desidera accedere al servizio AWS. Ciò può risultare costoso e complesso da gestire nella configurazione della zona di destinazione in cui un cliente desidera interagire con un servizio AWS specifico in più VPC. Per evitare ciò, è possibile ospitare gli endpoint di interfaccia in un unico VPC centralizzato. Tutti i VPC spoke utilizzeranno questi endpoint centralizzati.

Quando si crea un endpoint VPC per un servizio AWS, è possibile abilitare il DNS privato. Questa impostazione, se abilitata, crea una zona ospitata privata di Route 53 gestita da AWS che consente la risoluzione dell'endpoint del servizio AWS pubblico nell'IP privato dell'endpoint di interfaccia. La zona ospitata privata gestita funziona solo all'interno del VPC con l'endpoint di interfaccia. Nella configurazione esaminata, se si vuole che i VPC spoke siano in grado di risolvere il nome DNS degli endpoint VPC ospitati in un VPC centralizzato, la zona ospitata privata gestita non funziona. Per ovviare a questo problema, è possibile disabilitare l'opzione di creazione automatica del nome DNS privato quando viene creato un endpoint di interfaccia. In alternativa, è possibile creare una zona ospitata privata di Route 53 manualmente e aggiungere un record Alias con il nome completo dell'endpoint del servizio AWS che punta all'endpoint di interfaccia, come illustrato nella Figura 18.

Figura 18. Zona ospitata privata creata manualmente

Associamo questa zona ospitata privata agli altri VPC nella zona di destinazione. Questa configurazione permette ai VPC spoke di risolvere i nomi completi degli endpoint del servizio negli endpoint di interfaccia nel VPC centralizzato.

> **Note**
>
> Per accedere alla zona ospitata privata condivisa, gli host nei VPC spoke devono utilizzare l'IP di Route 53 Resolver del loro VPC. Gli endpoint di interfaccia sono accessibili anche dalle reti On-Premise tramite VPN e Direct Connect. È possibile utilizzare le regole di inoltro condizionale per inviare tutto il traffico DNS per i nomi completi degli endpoint del servizio agli endpoint in entrata di Route 53 Resolver, che risolverà le richieste DNS in base alla zona ospitata privata.

Nella Figura 19 Transit Gateway consente il flusso di traffico dai VPC spoke agli endpoint di interfaccia centralizzati. Crea endpoint VPC e la relativa zona ospitata privata nell'account dei servizi di rete e condividili con i VPC spoke negli account spoke. Per ulteriori dettagli sulla condivisione delle informazioni degli endpoint con altri VPC, consulta il post di blog Integrating AWS Transit Gateway with AWS PrivateLink and Amazon Route 53 Resolver.

Nota: un approccio di questo tipo con endpoint VPC distribuito, ovvero un endpoint per VPC, permette di applicare policy di privilegio minimo agli endpoint VPC. Con un approccio centralizzato, si applicano e si gestiscono le policy per l'accesso a tutti i VPC spoke in un singolo endpoint. Con l'aumentare dei VPC, diventa più complesso mantenere il privilegio minimo con un unico documento di policy. Un unico documento di policy comporta anche un raggio di azione più ampio. Sono inoltre previste limitazioni per la dimensione del documento di policy (20.480 caratteri).

Figura 19. Centralizzazione degli endpoint VPC di interfaccia

Conclusione

Man mano che si dimensiona l'utilizzo di AWS e si distribuiscono applicazioni nella zona di destinazione AWS, il numero di VPC e componenti di rete aumenta. In questo whitepaper è stato illustrato come gestire l'infrastruttura in crescita in modo da offrire scalabilità, disponibilità elevata e sicurezza mantenendo bassi i costi. Prendere le giuste decisioni di progettazione quando si utilizzano servizi come Transit Gateway, VPC condiviso, AWS Direct Connect, endpoint VPC e appliance software di terze parti diventa fondamentale. È importante comprendere le considerazioni chiave di ciascun approccio e partire dai requisiti specifici per determinare quale opzione o combinazione di opzioni soddisfa al meglio le esigenze.

Collaboratori

Hanno contribuito alla stesura di questo documento:

- Sidhartha Chauhan, Solutions Architect, Amazon Web Services
- Amir Abu-akeel, Cloud Infrastructure Architect, Amazon Web Services
- Sohaib Tahir, Solutions Architect, Amazon Web Services

Cronologia del documento

Per ricevere una notifica sugli aggiornamenti di questo whitepaper, iscriviti al feed RSS.

update-history-change	update-history-description	update-history-date
Aggiornamento di minore entità	Sezione Confronto tra Transit Gateway e peering VPC aggiornata.	2 aprile 2021
Whitepaper aggiornato	Testo corretto per assicurare la corrispondenza con le opzioni illustrate nella Figura 7.	10 giugno 2020
Aggiornamento di minore entità	Testo corretto per assicurare la corrispondenza con le opzioni illustrate nella Figura 7.	10 giugno 2020
Pubblicazione iniziale	Whitepaper pubblicato.	15 novembre 2019

Avvisi

I clienti sono responsabili della propria valutazione autonoma delle informazioni contenute in questo documento. Questo documento: (a) è solo a scopo informativo, (b) mostra le offerte e le pratiche attuali dei prodotti AWS soggette a modifiche senza preavviso e (c) non crea alcun impegno o garanzia da parte di AWS e dei suoi fornitori, licenziatari o affiliate. I prodotti o servizi AWS sono forniti "così come sono" senza garanzie, dichiarazioni o condizioni di alcun tipo, sia esplicite che implicite. Le responsabilità e gli obblighi di AWS verso i propri clienti sono disciplinati dagli accordi AWS e il presente documento non fa parte né modifica alcun accordo tra AWS e i suoi clienti.