SCTE·ISBE CABLE-TEC EXPO 2017
DENVER, CO
OCTOBER 17-20

2017 Fall Technical Forum
SCTE·ISBE • NCTA • CABLELABS

Cloud Overlay (CLOVER)

Extending the Cloud Virtually

A Technical Paper prepared for SCTE/ISBE by

John Jason Brzozowski
Fellow
Comcast
Philadelphia, PA 19103
484-962-0060
john_brzozowski@comcast.com

Mark Brittingham, Comcast
Chris Luke, Comcast
Zheng Yin, Comcast

Table of Contents

Introduction

The term “cloud” is practically a household word today. References to “cloud” have matured, in a relatively short period of time, from what was an abstract concept, to infrastructure and resources used widely by consumers. Adoption of cloud infrastructure has obviously surpassed the specialized usage by large enterprise or service provider adopters. While the use of the cloud has evolved, how adopters access and utilize it has remained largely unchanged. In fact, one could argue that classic, aging networking techniques remain pervasively used today to gain access to third party cloud (TPC) resources and infrastructure. The aggressive adoption of cloud technologies seems to be pushing the limits of traditional techniques, not to mention the associated business and cost models.

The approach described in this paper is one that has been developed to modernize how cloud adopters connect to and utilize modern TPC infrastructures. The objective of this approach, which we call the Cloud Overlay (and hereafter referenced as “CLOVER”), is to marry automation, modern networking techniques, and existing, well-known protocols to help maximize how applications and services are securely deployed to third party clouds. Further, many of the techniques outlined in this paper can be extended and utilized within an enterprise or service provider network to enhance how internal users leverage their own private clouds.

CLOVER sets out to leverage more deliberately the concepts of overlay and underlay networking to provide seamless connectivity to cloud resources that are both on and off network. Today, the line is blurred, perhaps even non-existent, between the concepts of overlay and underlay networking, mainly because they often follow the same layer 3 path. For clarification, an underlay is analogous to how a typical Virtual Private Network (VPN) functions, where the VPN connection is the underlay and the overlay includes corporate email or Intranet communications over the VPN. Further, for CLOVER it is essential to clearly and distinctly differentiate between a service or application interface, and a control interface for a host or collection of hosts that have been deployed into TPC provider infrastructure.

Background

Not all applications or services that run in the cloud are equal. Today’s technology landscape goes far beyond hosting a simple web server in the cloud for an e-commerce offering, or even a personal blog. Most applications and services are quite complex, requiring advanced functionality for storage, performance, security, and network connectivity. In fact, many cloud adopters use TPCs as an extension of their own internal infrastructures -- which means that in many cases, what is hosted “in the cloud” may not even be reachable or usable over the Internet, only internally. Conversely, there are many applications and services that utilize a hybrid model, where they, in fact, are intended to be reachable and usable over the Internet. However, doing so requires access to a complex maze of backend services that are rightfully locked down in secure, on network data centers that are not reachable over the Internet. It is this hybrid model that has introduced complexity, and, in some cases, costs into the TPC adoption equation. This model has simultaneously stressed traditional networking and security technology, while fueling innovation that greatly enhances how TPC infrastructures could be used by the bulk of adopters.

Traditional Cloud Models

The rapid proliferation of TPC offerings, coupled with the desire to minimize the burdens on network data center maintenance, provided much of the early motivation for adopters to migrate toward the use of third party clouds. The attraction of third-party clouds offered the following:

- The promise of a cost-optimized, financially attractive cloud infrastructure, at least in the early stages
- Improved deployment agility, efficiency, and velocity enabled by TPC automation
- The potential to minimize capital investments in new and/or aging on-network, specialized data centers

Organizations that own and operate their own networks, data centers, and cloud infrastructures, and have also embraced TPCs, likely have come to the realization that it's quite a complex juggling act to keep up with demand for the cloud and virtualization, while managing capacity, quenching the thirst for new technology, and minimizing impact and downtime. The same adopters have also likely come to the conclusion that the grass is not green(er) on the other side.

Reference Architecture

There are a multitude of connectivity models available today to enable connectivity to TPCs. Two primary examples will be referenced within this paper: the Full Service TPC, and the Dedicated Infrastructure TPC. Reference to these models will help punctuate some of the challenges associated with their expansion today and in the future, in particular as TPC usage continues to balloon. Specifically, the reference models are:

1.1. Full Service TPC

The Full service TPC is a common model that is most analogous to traditional hosting models, where all resources required by an adopter are deployed on a TPC provider's infrastructure. This includes everything from compute, storage, and possibly application- or service-specific data. The adopter, in this case, effectively outsources most, if not all, infrastructure related activities to the TPC provider.

1.2. Dedicated Infrastructure TPC

Dedicated infrastructure TPC is a model that has grown in popularity. In it, the adopter and the TPC provider establish and maintain dedicated connectivity to the TPC infrastructure. Effectively, the adopter treats the TPC infrastructure as an extension of its own network, data center, or cloud infrastructure. Most TPC providers have an offering of this type.

Observations

Both traditional models carry distinct and immediate benefits. In the Full Service model, the potential to maximize the economies of scale and cost reductions afforded to the TPC provider are by far the most attractive attributes -- providing that the cost savings are at least in part passed on to the adopter or customer of the TPC. Delegating day-to-day operational responsibilities related to managing downtime, releasing upgrades, and augmenting capacity specific to the underlying TPC hardware all yield direct benefits to adopters of this model. Infrastructure delegation does not include responsibilities associated

with the applications or services that are running within the TPC environments, which is deliberate and often viewed as beneficial. All of these responsibilities remain with the application or service owner which is typically the customer of the TPC provider. With these models, adopters are able to focus primarily on application or service excellence and agility. Delegating responsibility related to the underlying infrastructure often allows for the reallocation of time, energy, and budget to application or service innovation and development.

While there are many bona fide benefits, there are also a number of considerations that must still be considered when considering one or more of the TPC models that are available today. The most obvious is relate to "shared fate" and the consequences of the inability to migrate between TPCs, or to use multiple TPC providers simultaneously. Particularly for the Full Service model adopter, if its TPC provider experience issues with their infrastructure (which have been rare to date), those issues typically impact large populations of the TPC customer base. A single TPC adopter is rarely impacted in an isolated manner.

Additionally, many TPC providers have invested heavily in technologies specific to their respective platforms. In most cases, these investments were driven by customer or industry demand, which has truly benefited those adopters. However, those same investments and innovations typically make it difficult (if not impossible) to migrate from one TPC provider to another, or to use multiple TPCs simultaneously.

Further, there are cases -- with the Dedicated Infrastructure model, specifically -- where localization, performance, and redundancy are significantly affected by the resulting network topology. The investment and resources required by adopters of the dedicated TPC model often also requires dedicated access to capacity for a subset of TPC data centers. While offering numerous benefits, this model does have a potentially adverse impact on redundancy and localization -- which, in turn, impacts performance, and ultimately the consumer experience.

Finally, and not insignificantly, the capacity, reliability, and performance of the underlying network between the TPC provider network and the adopter network is critical, specifically for the dedicated infrastructure model. Unpredictability and fluctuations on either side can introduce significant instability and customer impact. Resource requirements in the form of capacity planning and management are essential to the successful use of the Dedicated TPC deployment model. In practice, not all adopters of TPCs have the required resources -- human and financial -- to consider the Dedicated Infrastructure deployment model.

Next-Generation Cloud

The "cloud" has clearly helped to fuel innovation and the deployment of new applications and services. As previously mentioned, this phenomenon helped to push technology to new limits -- and, in some cases, is or soon will be pushing some adjacent technologies past their breaking point. Cloud-based networking and connectivity are near the top of the list of areas where we, as a community of adopters, continue to use traditional techniques that need to be revisited and/or redesigned. CLOVER, our moniker for "Cloud Overlay," is an outgrowth of those realizations, and aims to offers the following:

- Independence and flexibility across multiple third-party cloud infrastructures
- Automation
- Scale
- Distribution and localization

- Improved redundancy

CLOVER more clearly delineates between the “underlay” and “overlay” aspects of next generation, cloud-oriented applications and services. Today, developers and engineers blur the lines between the underlay and overlay, treating them both equally. While this has worked for years, this approach forces “baggage” to be carried from one chapter to the next, from an infrastructure engineering perspective. Applications and services that have been deployed into the cloud have distinct communication properties -- namely, the application or service interface, and the control interface. To date, all of these communications properties are generally managed as single flows, to and from resources in the cloud. By separating application and services flows from control flows, traffic and connectivity can be separately managed. This allows TPC adopters to maximize and more effectively utilize cloud resources, and the robust infrastructures being built and deployed by TPC providers.

Reference Architecture

CLOVER builds heavily on concepts pertaining to underlay and overlay connectivity and communications. To clarify, underlay connectivity is a term describing the encapsulation or transmission of application, service, or end-user communications using an alternate transport. Overlay communications are typically what is being carried or encapsulated, while the underlay, as expected, is the carrier. CLOVER expands on this by differentiating application and service communications properties from the control communications. The classic example is the common “storefront” web application. Such an application typically consists of the following:

1. One or more web server front-ends
2. One or more backend systems, for example, databases

In the web storefront example, a cluster of web servers lacking customer and product databases does not make for much of an experience or much of a storefront for that matter. Conversely, exposing a customer or product catalog database directly over the Internet is inconceivable without proper security and scale considerations. The two challenges jointly make the case for coupling an HTTP-based application or service interface -- in this case, care of the web servers -- with a control interface to the customer and product databases, via a structured query language (SQL).

The core of the CLOVER target architecture allows for a separation of control communications from application or service communications, with each being managed independently. Control properties often require that communications be secure, and often not available over the Internet-at-Large, leveraging well-known underlay networking concepts. Underlay communications indicate the use of an underlying transport, often Internet Protocol (IP)-based, that can encapsulate or carry others forms of overlay communications -- which are, incidentally, Internet Protocol-based communications as well. With the pervasiveness of Internet Protocol version 6 (IPv6) today, there is often a wide range of underlay and overlay combinations. Overlay communications may include one or more of the following:

- IPv6 only
- IPv4 only
- Dual Stack (IPv4 and IPv6 are both enabled and used simultaneously)

In conjunction with the overlay combinations above, underlay communications can also be deployed and operated in various versions of the Internet Protocol. Unlike overlay communications, underlay

communications need not operate in dual-stack mode, where both IPv4 and IPv6 are enabled and used simultaneously. Underlay communications may, for example, prefer the use of IPv6 with only a fallback to IPv4, in the event that IPv6 is inoperable or is not supported. Notably, a fall back mode of operation is not analogous to dual-stack mode. Beyond the examples listed below, there are additional options for underlay communications, however, addressing these is out of scope for this paper. Further, there are several transport modes or protocols that can be used in conjunction with Internet Protocol to carry overlay Internet Protocol communications. Each of the below offer varying levels of security, including both authorization and encryption. Typically, those that are the most secure often have the greatest overhead, which, in turn, introduces the possibility of performance impacts. Underlay transport modes and protocol examples include:

- IP-in-IP[1]
- Generic Route Encapsulation (GRE)[2]
- Internet Protocol Security (IPsec)[3]
- Virtual Extensible LAN (VXLAN)[4]

The development documented in this paper focuses on IPSec over IPv4 only and IPv6 only with varying levels of authorization and encryption. For simplicity, during development, pre-shared keys were used for IPsec authorization. However, for a production deployment of CLOVER-based solutions it is recommended that certificates be utilized minimally for IPsec authorization, mainly for simplifications related to automated resource creation and deployment. Several algorithms representing a wide range of encryption levels were used during development, including:

- Null or no encryption[5] – effectively no payload encryption
- AES-GCM256[6] – moderate encryption
- AES-256[6] – better encryption

Several additional algorithms for an IPsec payload encryption are available for use, but for the purposes of CLOVER development, it was of primary importance to determine the performance characteristics of good/better/best levels of encryption. Adopters of CLOVER are likely to select authorization techniques and encryption algorithms that are best aligned with their infrastructures.

1.3. Key Components

The following section outlines the key elements of a CLOVER-based solution. Each component of a CLOVER-based deployment can be implemented and engineered in a manner that best suits the adopter. For the purpose of this work, each component and its function is described autonomously, such that readers can determine independently the best utilization within their own infrastructures.

[1] https://en.wikipedia.org/wiki/IP_in_IP
[2] https://en.wikipedia.org/wiki/Generic_Routing_Encapsulation
[3] https://en.wikipedia.org/wiki/IPsec
[4] https://en.wikipedia.org/wiki/Virtual_Extensible_LAN
[5] https://tools.ietf.org/html/rfc2410
[6] https://tools.ietf.org/html/rfc4106

1.3.1. Virtual Network Appliance (VNA)

A virtual network appliance, or VNA, is, in essence, a virtualized network function where a software module that typically runs on specialized hardware is built to run within a virtual machine, or more specifically as a virtual machine within a private, public, or third party cloud environment. As it relates to CLOVER, the primary function that is being virtualized is one that provides IPsec-authorized and encrypted communications. IPsec VNAs will typically terminate communications against another network element, which is likely to be a physical device that is redundant and fault tolerant. While it is possible, it is not strictly required for VNAs to terminate communications against common implementations, meaning that VNAs and aggregators (see next section) are intended to be fully interoperable. CLOVER aggregators are typically deployed in a centralized manner leveraging a hub and spoke[7] or star topology[8]. The deployment model for CLOVER aggregation can be duplicated across a large network for increased capacity, improved performance, and localization.

The VNA is a primary communication path for all or a subset of network communications that are sent to and from hosts to where the VNA provides secure, network connectivity. As such, performance and IP transport implications are key considerations for deployment planning.

1.3.2. Aggregator

CLOVER aggregators are typically sets of redundant, high performance network elements that terminate secure network underlay communications for CLOVER virtualized networks functions like VNAs. While it is conceivable that aggregators too can be virtualized network functions, it is common for elements of this type to be dedicated hardware elements. Given the performance of modern computing platforms, commodity hardware is a valid consideration, from a platform point of view. Specifically, commodity computing platforms with high speed network interfaces running open source operating systems, like Linux or BSD derivatives, are legitimate alternatives to specialized, commercial hardware platforms. For example, Vector Packet Processing (VPP) under the Fast Data Plane Project (fd.io) supports packet forwarding on commodity hardware that is comparable to many commercially available platforms.

While there are capital and operational expenditures associated with specialized hardware platforms, open source software and commodity computing platforms are not without their own procurement and maintenance costs. The choice from an aggregator point of view truly boils down to adopter preference and capabilities.

1.3.3. Hosts

CLOVER hosts are the simplest element in the system. Hosts are just that, hosts, either bare-metal or virtualized, that are used to run proprietary or commercial applications and services. CLOVER is simply the mechanism by which network connectivity and communications are provided to them. Hosts can run on any operating system, and can essentially run a wide range of functions with a practically unlimited set of network configurations.

[7] https://en.wikipedia.org/wiki/Spoke%E2%80%93hub_distribution_paradigm
[8] https://en.wikipedia.org/wiki/Star_network

The origination of secure, dynamic underlay communications directly from hosts to aggregators, bypassing a VNA, is technically a valid mode of operation for CLOVER hosts, however, is out of scope of this document. This is future work and is in fact a valid construct in the context of a CLOVER system.

1.4. Communication Modes

Building on the definition of the basic components of a CLOVER system, the base communication modes illustrate how the elements of CLOVER can provide a flexible mix of communications paths that offer adopters alternatives compared to traditional cloud connectivity techniques -- private, public, or otherwise. The role of a CLOVER aggregator is to provide termination points for CLOVER VNAs into an area of a private network that may not be reachable over the Internet, or may intentionally be secure. CLOVER VNAs can be placed at multiple entry points in a serving network to provide granular, targeted access to different network segments. The VNA must obviously be able to terminate secure, underlay communications. In this case, the underlay described in this paper was built using IPsec. Operational, virtualized hosts in a third-party cloud environment can in turn utilize the VNA and the active, secure connection back to and through a CLOVER aggregator, over the Internet, for all or a subset of communications. The type, destination, or traffic source are governed by the chosen CLOVER connectivity model. CLOVER provides the flexibility to select a communication model that allows adopters to manage exactly how their applications and services are communicated with, including options to “bring your own” IP addressing.

1.4.1. Converged Model

In a converged communications model for CLOVER, a VNA is used to manage all host-based communications, regardless of IP version and other network communication properties. In a converged model, hosts in the cloud are not reachable through any other paths -- over the Internet or otherwise. From a management perspective, virtualized hosts may, in fact, remain reachable via virtual consoles, however, this is not entirely different than how physical hosts are managed today in a non-CLOVER environment.

The converged model is typically used in cases where an application or service is targeted for a cloud environment and is only required to be reachable from within an enterprise, and not over the Internet. The CLOVER converged model is analogous to a virtual data center connected back to a larger, centralized corporate network using an extension cord. Today, in many cloud environments, especially third party clouds, there are often options to establish direct connectivity to the provider. While these have some attractive properties, they can sometimes be costly, difficult to manage, and inflexible from a resiliency perspective.

In a converged model, all control, application, and service communications flow over the underlay via the VNA, since there are no alternate paths. As such, this does increase the throughput and bandwidth requirements for the VNA. Finally, increased volume of communication through the VNA can potentially decrease the overall quantity of hosts that can be served in this model. This in no way makes the model less usable or less desirable. It simply suggests that adopters must explicitly assess and profile applications and the respective deployment models that are best suited for a converged connectivity model. Access to a CLOVER-based converged cloud infrastructure allows for the expanded or extended use of cloud environments for applications or services that otherwise might be condemned to a life outside of the cloud.

1.4.2. Split Model

Unlike a converged model, a CLOVER "split" model introduces the notion of bifurcated communications. Essentially, control traffic for applications and services running on hosts are routed via the CLOVER VNA over the secure underlay, while actual application or service communications are routed or reachable via the Internet-facing addresses provided by the underlying cloud infrastructure. Again, the underlying cloud infrastructure could be private or public.

Analysis of control traffic for common environments suggests that the majority of communications to and from hosts are to support the currently running services or applications. As such, the use of split mode communications creates a positive imbalance in communications, such that control traffic capacity over the VNA can support a larger deployment of hosts. An assumption of 25-50% of communications to and from virtualized hosts for control communications represents an increase of 4 or 2 times, respectively, for a host's application or service capacity. In essence, the bifurcation of communications allows for a calculated over-subscription of the CLOVER VNA for network bandwidth and throughput. It is this over-subscription that is a key enabler of increased performance. Further, the use of Internet-facing addressing, provided by the cloud infrastructure, allows significantly enhanced localization and redundancy.

While there are interesting benefits to the use of the CLOVER split communications models, there are also some notable considerations. This model assumes that control information is the minority traffic for network communications. If this is not the case, then many of the gains, from a performance and capacity perspective, will be lost. Additionally, security considerations will need to be closely evaluated specifically from a network and a host perspective. Hosts in a split model may effectively straddle the Internet and secure network segments. This is not unheard of, however, done inadvertently this could compromise security for the adopter.

Observations

Modern day use of the cloud seems to have outpaced network engineering, and in many cases, the underpinnings of network technology. Further, traditional networking in many ways seems to limit the possibilities that the cloud offers to current and future adopters. The concepts explored with CLOVER push network technologies to their limits as they are applied to the cloud. CLOVER advances the cloud and virtualization to truly incorporate the network -- to the extent that data centers (large and small) can now be fully virtualized, in the cloud, while allowing for connectivity back to their parent network to gain secure access to protected resources.

Additionally, the CLOVER architecture more effectively enables adopters to leverage multiple cloud infrastructures simultaneously, which can yield significant cost benefits while maximizing redundancy and reducing the surface area of risk associated with the use of a single cloud provider. Conversely, CLOVER does represent a wholesale paradigm shift around how network and cloud technologies interact. The lines between the two blur, or certainly have the opportunity to blur. How fast, how slow, or if at all are up to the adopter.

Finally, through the use of CLOVER, application and service owners alongside their network brethren will more explicitly dissect network communication. A keen, in-depth understanding of the communications is instrumental in identifying the best CLOVER models for a given application or service set, while optimizing for performance, cost, and efficiency.

Deployment and Operational Considerations

Fully embracing virtualization and the cloud along with the enormous opportunities they represent, also introduces a significant opportunity for other changes. Change in the underlying technology and infrastructure, like CLOVER, can lead to changes in the scale and velocity around how networks, clouds, applications, and services are engineered and deployed.

Automation

CLOVER was built to be fully automated, leveraging the atomic, programmatic building blocks offered by the wide range of cloud technologies and platforms available today. A critical aspect of CLOVER is the automated creation, deployment, and enablement for all of the key elements of the CLOVER system, including:

- Creation of the CLOVER VNAs
- Bi-directional provisioning and licensing CLOVER VNAs with the CLOVER aggregator, including the communication mode (split or converged)
- Automated creation or cloning of virtualized hosts

It is possible, and perhaps desirable, initially, for existing functional areas with an organization to own and manage their respective tasks as noted above. However, it is conceivable that application and service owners or end-users can and will fully automate the creation of cloud-powered virtual infrastructure, end-to-end, including but not limited to the provisioning of the underlying network to enabling users to generate their own configurations. Processes and deployments that currently take days or weeks can now be completed in minutes or seconds. Generally speaking, automation is most often referred to in the context of creation, but decommissioning and resource recovery is equally as important to ensure that antiquated or defective technologies are managed and updated accordingly.

Monitoring and Telemetry

Monitoring and telemetry are critical for physical infrastructure that is largely fixed. The dynamic nature of the cloud and virtualized infrastructure makes this exponentially more important -- especially in a CLOVER-like model, where every element can automatically be created, decommissioned, or moved in a moment's notice. Automated management of CLOVER resources must include dynamic enablement and population of monitoring systems. Further, with such a dynamic virtual environment, it is essential for adopters to re-think their deployment models. Specifically, the opportunity to distribute applications or services more widely into larger quantities of smaller clusters becomes a reality, and in some cases is highly desirable.

Conclusion

CLOVER, or Cloud Overlay, was born out of real operational, technological, and commercial challenges associated with the aggressive use of and deployment into TPC infrastructures. As described, the use of the cloud and virtualization, and the associated business models, will continue to test the limits of the underlying network technologies.

Embracing virtualization to include network functions, coupled with automation, enables infrastructure engineers to keep pace with the evolution of the cloud -- and, in some cases, to fueling innovation. This applies to the use of third party clouds as much as it applies to next-generation private or on premise cloud infrastructures. The virtual landscape that is in front of us now lays the foundation for end-to-end automation and simplification, truly enabling adopters to streamline how the cloud is employed to power their business, products, services, and people.

Cloud utilization will continue to grow, from a sheer scale perspective, as will the related verticals, including the Internet of Things (IoT). This massive horizontal and vertical growth of the cloud will continue to drive innovation, potentially pushing architectures like CLOVER to quickly move beyond concepts like virtual network appliances, to host-based CLOVER models -- where hosts are dynamically and securely negotiating underlay communications paths. The possibility (likelihood) of host-based models thrusts open the doors of innovation, driving CLOVER to evolve to utilize technologies like IPv6 Segment Routing to statelessly and programmatically establish secure, redundant underlay communications over IPv6-only networks.

Abbreviations

GRE	Generic Routing Encapsulation
IETF	Internet Engineering Task Force
IPv4	Internet Protocol version 4
IPv6	Internet Protocol version 6
ISBE	International Society of Broadband Experts
IPsec	Internet protocol security
RFC	Request for Comment
SCTE	Society of Cable Telecommunications Engineers
SR	Segment Routing
TCP/IP	Transmission Control Protocol/Internet Protocol
TPC	Third Party Cloud

Bibliography & References

Generic Routing Encapsulation (GRE), RFC2784; Internet Engineering Task Force

Security Architecture for the Internet Protocol (IPsec), RFC4301; Internet Engineering Task Force

Internet Protocol, Version 6 (IPv6) Specification (IPv6); RFC8200; Internet Engineering Task Force

IP in IP Tunneling (IP-in-IP); RFC1853; Internet Engineering Task Force

