Network working group
INTERNET-DRAFT
Intended Status: Informational
Expires: April 27, 2015

P.A. Aranda
Telefonica
D. King
Lancaster University
M. Fukushima
KDDI R&D Labs
October 27, 2014

Virtualization of Content Distribution Network Use Case
draft-aranda-vnfpool-cdn-use-case-00

Abstract

This use case document provides requirements for moving Content Distribution Networks (CDNs) from physical servers to a virtualized environment. This new kind of CDN, known as virtualized CDN (vCDN), allows for new constructs that simplify the CDN architecture. The main elements of the CDN are analyzed with regards to the degree of elasticity demanded from them in terms of computation, storage and network resources.

This use case document provides resiliency requirements for virtualization of the Content Distribution Network, known as virtualized CDN (vCDN).

Table of Contents

1 Introduction

Delivery of content, especially of video, is one of the major challenges of all operator networks due to massive growing amount of traffic.

Growth of video traffic is driven by the shift from broadcast media to unicast delivery via IP. This is also complementary to the growth of today’s video on demand traffic.

Additional on-demand content services to Internet end-users, have similar quality constraints as video, high bandwidth and low latency, and stored as close to users as possible.

A Content Delivery Network (CDN) represents a group of geographically dispersed servers deployed to facilitate the distribution of information generated by content providers in a timely and efficient manner.

As physical functions, including CDN components, are migrated to virtual platforms, Virtual Network Functions (VNF), a critical aspect will be ensuring the VNF is resilient. Maintaining that resilience, especially, when virtual resources are dynamically migrated and managed will require co-ordination between VNFs.

This document discusses the key network resilience objectives for the virtualized CDN. It outlines the challenges and risks for the appropriate resilience requirements to negate or ensure minimal impact of CDN-based services.

1.1 Defining Resilience

In the context of this I-D resiliency will ensure the ability to provide and maintain an acceptable level of service or function to the user, in the event of faults and challenges to normal operation.

1.1.1 Resiliency for stateless services

In the case of services that do not require maintaining state information, it is sufficient to move the VNF offering that service to a new Virtual Machine (VM) or hardware entity.

1.1.2 Resiliency for stateful services

When a VNF is moved e.g., for failure mitigation, maintenance or workload consolidation, the offered service and its performance can be maintained, which is regarded as "service continuity" by those entities which are using it.

2 vCDN Use Case

2.1 Terms and definitions

CDN Provider: The service provider who operates a CDN and offers a service of content delivery, typically used by a Content Service Provider or another CDN Provider.

Content: Any form of digital data. One important form of Content with additional constraints on distribution and delivery is continuous media (i.e., where there is a timing relationship between source and sink).

Content Delivery Network (CDN): The network infrastructure in which the network elements cooperate at Layers 4 through 7 for more effective delivery of Content to Users.

Network Service Provider (NSP): Provides network-based connectivity and services to Users.

Over-the-top (OTT): A service, e.g., content delivery using a CDN, operated by a different operator than the Operator to which the users of that service are attached.

Service Continuity: ensure that if a service needs to be relocated to another site due to an anomaly event (e.g. CPU overload, hardware failure or security threat). The configuration of the VNF (e.g. IP address) is preserved; thus (ideally) there is no impact on the end user or node.

Users: The end user that interacts with a Content service. Such communication is not restricted to HTTP and may be via a variety of protocols. Examples of Users include: browsers, Set Top Boxes (STBs), dedicated content applications (e.g., media players), etc.

2.2 Content Distribution Network Components

A number of functional components exist for deployment and operation of Content Distribution Networks (CDN), these include:

- Content Cache Node to deploy content as close to each user as possible;
- Content Controller to route the users request for content to the closest available content store or content engine;
- Content Load Balancing to distribute user requests across one or multiple servers;
- Surrogate Servers for mirrored web content servers;
- Content Proxies;

- o Content DNS Servers;
- o GeoIP Information Servers;
- o Content Peering Gateways.

In many CDN deployments, CDN nodes are dedicated physical appliances or software with specific requirements on standard but dedicated hardware. Often physical appliances and servers for different purposes are deployed side-by-side. This comes with a number of disadvantages:

- o The capacity of the devices needs to be designed for peak hours (typically on weekend evenings). During weekdays and business hours, the dedicated hardware appliances and CDN servers are mainly unused.
- o It is not possible to react on unforeseen capacity needs e.g. in case of a live-event as hardware resources need to be deployed in advance.
- o The average peak utilization and resilience of CDN nodes for dedicated purposes or from different partners is lower as it could be if the hardware resources would be shared between virtual appliances on the same infrastructure.
- o Dedicated physical devices and servers from several parties drive the complexity of the operator network and increase the operational expenses.
- o Content delivery is a very volatile market driven by new content formats, protocols, device types, content protection requirements etc. Dedicated designed hardware hinders the necessary flexibility to react on these changes.
- o Content Delivery may imply some Value Added Services, e.g., for Security concerns or for optimizing Performances. It may be valuable for the Network Operator to rely on Outsourcing of a Partner’s solution rather than having to operate its own solution.

Therefore, it is important for CDNs to offer service continuity to users during partial failures of key CDN elements, including: load balancers, proxies and surrogate servers.

2.2.1 Cache Node

Operators to deploy their proprietary cache nodes into the ISP network. CDN cache nodes are dedicated physical appliances or software with specific requirements on standard but dedicated hardware.

2.2.2 CDN Controller

A CDN controller objective is to select a cache node (or a pool of cache nodes) for answering to the end-user request, and then redirect the end-user to the selected Cache Node.

The Cache Node shall answer to the end-user request and deliver the requested content to the end user.

The CDN controller is a centralized component, and CDN cache nodes are distributed within the Network and in multiple locations.

2.2.3 CDN Load Balancer

A CDN Load balancer objective is to distribute the demand to different nodes in the CDN taking into account different criteria including geographical proximity, network-wise proximity (number of hops, policies set by the operator like ASN, etc.) or load/performance parameters of the servers in the CDN. Additionally, the CDN load balancer also provides resilience and protection against contents server failure.

2.2.4 Surrogate Server

The CDN surrogate server interacts with other elements of the CDN for the control and distribution of content within it and with User Agents for the delivery of the content to the users. This behaviour corresponds with the surrogate in the WWW context as defined in [RFC3040]. The surrogate server provides resilience and protection against contents server failure.

2.2.5 Content Proxy

The content proxy is a server that acts as an intermediary for requests from clients seeking resources from the CDN content servers. The content proxy provides resilience and protection against contents server failure.

2.2.6 Content Peering Gateway

The content peering gateway is the element that interconnects different CDNs [RFC7337]. This element is a single point of failure.

2.3 vCDN Resiliency Requirements

[TBD - Pedro & Dan]

2.3.1 Automatic scale-out/scale-in for unpredictable traffic variation

One of the significant benefits of virtualization for CDN Providers is the elasticity of resource provisioning. This enables the CDN Providers to mitigate unpredictable traffic variation. Due to the unpredictability, the elastic resource management such as scaling out/scaling in should be automatically performed by Service Control

Entity and Pool Manager, rather than manually performed by human operators.

2.3.2 Quality assurance of content delivery

Since the quality of content delivery service is a key performance indicator of CDN Providers, it is crucial to assure the quality during the process of scaling as well as after the completion of scaling. In particular, geographical locations of added/remaining Cache Nodes should be taken into account. This is because these geographical locations have significant impacts on the quality of content delivery.

2.3.3 Minimum impact on interconnection interfaces

Interconnections between CDNs [RFC7336] have been recognized as a new opportunity of value creation for CDN Providers. In order for vCDN to foster this opportunity further, vCDN should minimize its impact on the interconnection interfaces between CDNs. In particular, scaling in/scaling out vCDN should not require any change of the architecture, protocols, and IP addresses/DNS names of interconnection points.

2.4 Service Degradation

CDN-based services will require suitable monitoring of performance metrics for delivering content. These include:

- Connection time
- DNS lookup time
- Download time
- First byte response
- Latency
- Page load time
- Response to request time
- Throughput
- Error Rate
- Packet Loss
- Uptime

In the event of failure or service degradation, the ability to switch between comparative VNFs will be required.

2.5 Applicability of Virtual Network Function Pool (VNFPool)

The use case reveals the potential of VFNPOOL in simplifying the CDN architecture. Today's cache farms could be simplified in the way they are deployed and handled. Imagine you deploy a cache for a certain contents (e.g. newspaper web site) as a VNF. (This, as such, is a compelling use case, because the granularity is much finer and you might lower the minimum requirements for deploying a cache.)

The VFNPOOL protocol would control the way the VNF is deployed onto the VNFPOOL. Then, during its lifetime, it would control how it scales in or out.

Finally, it would also control the way the VNF is decommissioned. Apart from scaling in and out, the VNFPOOL protocol would also check for integrity and control the way a VNF can jump into another for resilience reasons.

This second kind of control should include state transfer and synchronization from the life to the backup VNF in some cases.

2.6 Coexistence of Virtualized and Non-virtualized Network Functions

With a CDN designed as loosely coupled software components a variety of scenarios of co-existing virtualized and non-virtualized components are possible.

Given that the CDN Controller is able to control Cache nodes deployed on virtualized and non-virtualized server instances in parallel the following scenarios are possible:

- More centralized located Cache nodes can run on virtualized (Cloud) resources while Cache Nodes distributed deeper into the network might run on physical appliances for operational reasons.

- Centralized cache cluster might run on dedicated non-virtualized server for performance reasons while Cache node instances distributed within in the network are running on virtualized resources available in other network devices

- Within a migration scenario from non-virtualized to virtualized the legacy cache nodes can be kept in production until the end of their hardware life-cycle is reached (i.e. operation efficiency is still sufficient) while new capacity is added to the CDN by deploying the same software on virtualized resources.

3 Security Considerations

<Security considerations text TBD>

4 IANA Considerations

<IANA considerations text TBD>

5 References

5.1 Normative References

5.2 Informative References

[RFC3040] Cooper, I., Melve, I., and G. Tomlinson, "Internet Web Replication and Caching Taxonomy", RFC 3040, January 2001.

[RFC7337] Peterson, L., Davie, B., and R. Brandenburg, "Framework for CDN Interconnection", RFC7337, June 2014.

[RFC7336] L. Peterson, B. Davie, and R. van Brandenburg, Ed., "Framework for Content Distribution Network Interconnection (CDNI)", RFC 7336, August 2014.

[zong-vnfpool-problem-statement] Zong, N., "Problem Statement for Reliable Virtualized Network Function (VNF) Pool", January 2014.

[TRILOGY2] The trilogy2 Consortium, "trilogy2: Building the Liquid Net", http://trilogy2.eu/

6. Acknowledgement

This work is supported by the European FP7 Project "Trilogy2" [TRILOGY2] under grant agreement 317756.

Authors' Addresses

Pedro A. Aranda
Telefonica, I+D; GCTO Unit
Spain
Email: pedroa.aranda@telefonica.com

Daniel King
Lancaster University
UK

Email: d.king@lancaster.ac.uk

Masaki Fukushima
KDDI R&D Laboratories, Inc.
Japan

Email: fukusima@kddilabs.jp