Testing Scalability of Data Center Timing Services

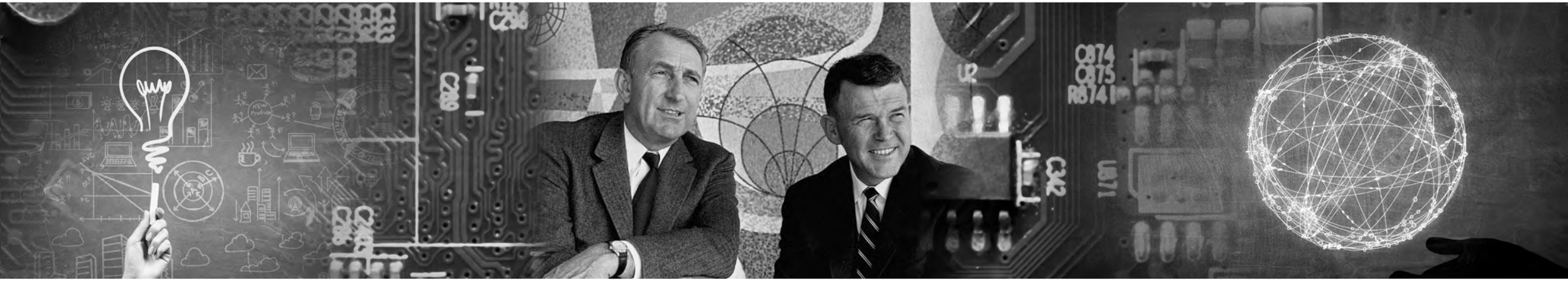
Alon Regev

Architect, Keysight Technologies

2021.11.03

Keysight Helps You Get to Market Faster

We help you create. Innovate. And deliver what's next.



The innovation leader in electronic design and test for over 75 years

Founded in 1939 by Bill Hewlett and Dave Packard as HP with an ongoing mission to help create new markets

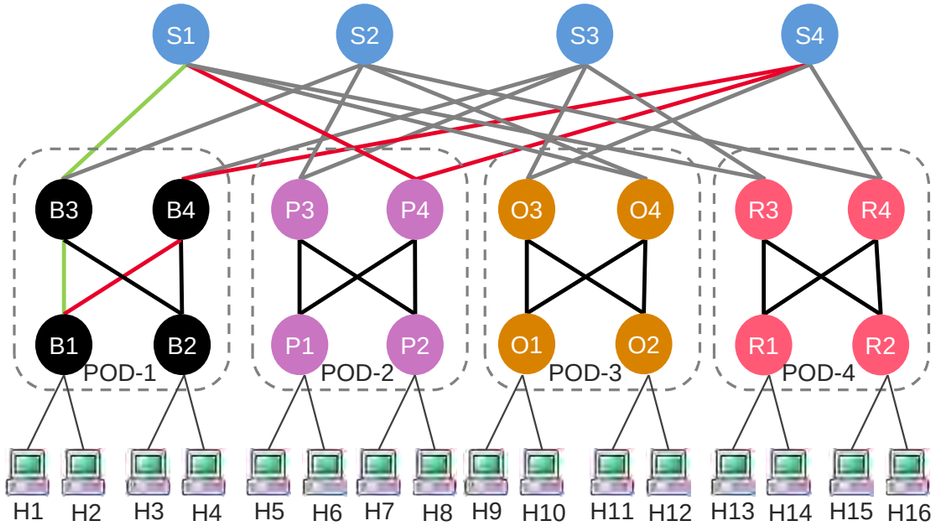
Trusted hardware, innovative software and a global network of experts

Introduction

- Precise time synchronization is needed by data centers
- PTP Data Center Profile has been defined by OCP TAP
- Data Centers have unique requirements requiring new solutions
- To match the data center requirements / implementations, new testing approaches are needed

Hyperscale data centers and their networks

- Hyperscale data centers are massive, highly optimized facilities
  - 10,000+ servers (often much more)
  - 10,000+ sq. ft
  - Optimized for power, compute, and network efficiency
- Using Spine-Leaf (aka Clos) organized switches
  - Multiple redundant paths between each source/destination
  - Reverse path typically different than forward path
- Layer 3 (IP) switching used rather than layer 2
  - IPv6 is typically used (rather than legacy IPv4)
- Multicast is often not used / supported
- Many VMs can run on each machine
- Scalability to millions or VMs and beyond



Data Center PTP Profile

- Existing PTP profiles did not meet DC requirements
  - Scalability to millions of PTP sessions per GM
  - IPv6 unicast
  - Using switches as TCs
- Data Center PTP profile developed by the Open Compute Project (OCP, opencompute.org) Time Appliance Project (TAP).
  - An IEEE Std 1588 profile designed to meet the data center requirements
- In addition to the Data Center Profile, the Open Time Card, Open Time Server, and related open software were developed by OCP TAP
  - Described in other ITSF presentations
  - See https://www.opencompute.org/wiki/Time_Appliances_Project

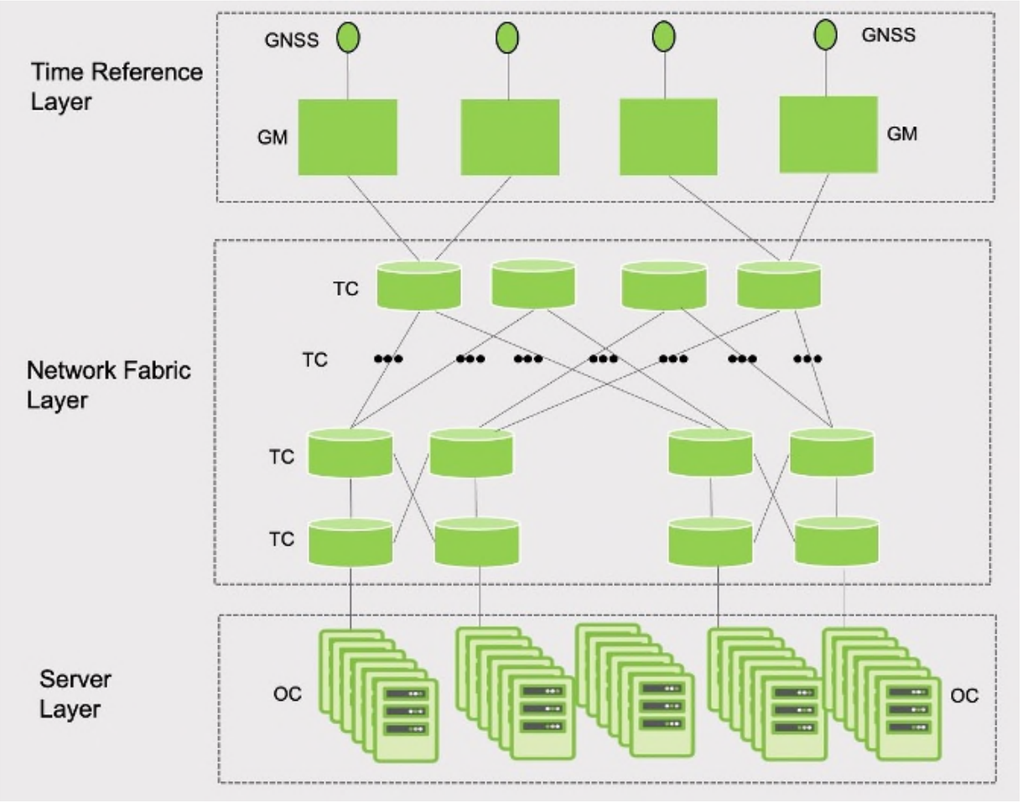
Data Center PTP Profile № 1

Abstract

This document defines a PTP profile for time-sensitive applications within a data center environment. The document is developed within the Open Compute Project (OCP) Timing Appliances Project community [1]. The PTP profile is based on IEEE Std 1588 TM-2019 [2]. When applicable, the profile also references and reuses information from other PTP profiles or other industry specifications. The document provides a set of requirements for implementing, deploying and operating timing appliances within a data center. A timing appliance is an element that is PTP-aware such as a switch/router, time server, NIC card, software module, timing card, monitoring device, etc.

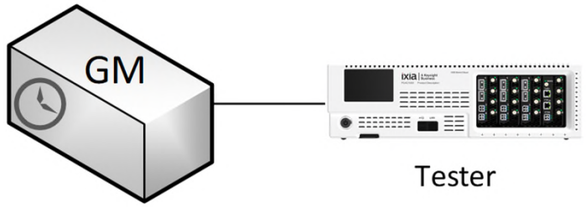
Data Center PTP Requirements

- From the OCP TAP Data Center Profile No 1 *
  - In normal operating mode, an OC has connectivity to more than 1 GM.
  - There are a number of GMs that are either active or standby.
  - The maximum time error between any two OCs must be within ± 5 microseconds, i.e., |TOC,*j* – TOC,*k*| <= 5 µs for $k \neq j$.
  - The maximum time error between a GM and any OCs must be within ± 2.5 microseconds. i.e., |TGM – TOC| <= 2.5 µs.
  - The maximum time error between any two GMs must be within ± 100 nanoseconds, i.e., |TGM,j – TGM,k | <= 100 ns for $k \neq j$.
  - The maximum time error generated by a TC must be within ± 100 nanoseconds, i.e., |TTC,j | <= 200 ns.
  - In normal operating conditions, each OC has connectivity into multiple GMs. Under failure of a GM, an OC must be capable of having connectivity to at least another GM.



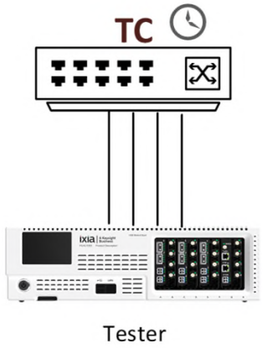
* see https://github.com/opencomputeproject/Time-Appliance-Project/tree/master/DC-PTP-Profile

Sample Data Center Testing Topologies



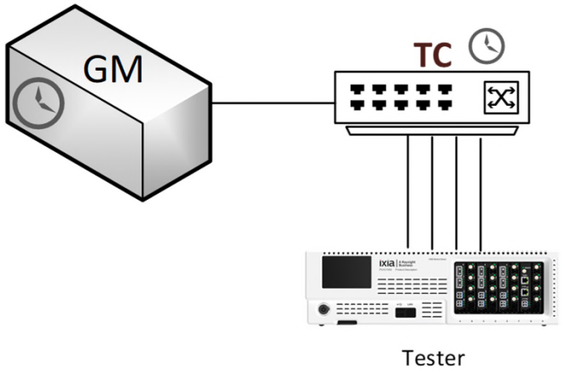
Test a GM

- Emulate millions of sessions per port
- Generate background traffic
- Accurately measure clock quality on all sessions
- Resilience & security test



Test a switch (TC)

- Emulate multiple GMs and millions of sessions utilizing the switch under test (acting as a TC)
- Generate other traffic types and patterns to validate switch behaviors with the traffic
- Measure clock quality on all sessions, characterizing the switch effect on the time quality



Test a system

- System under test can contain any mix of GMs and switches
  - Example: SUT can be single pod
- Test system emulates other parts of the system such as clients
- Test system can test under different traffic & protocol scenarios

Data Center PTP Testing Requirements

Emulate millions of OC clients to test scalability

Scale message rates

Test discovery & negotiation rates

Measure TE, Max |TE|, MTIE, TDEV across all OCs

System behavior when GM or POD disappears

Security testing (i.e. DoS resilience, fuzz testing, etc.)

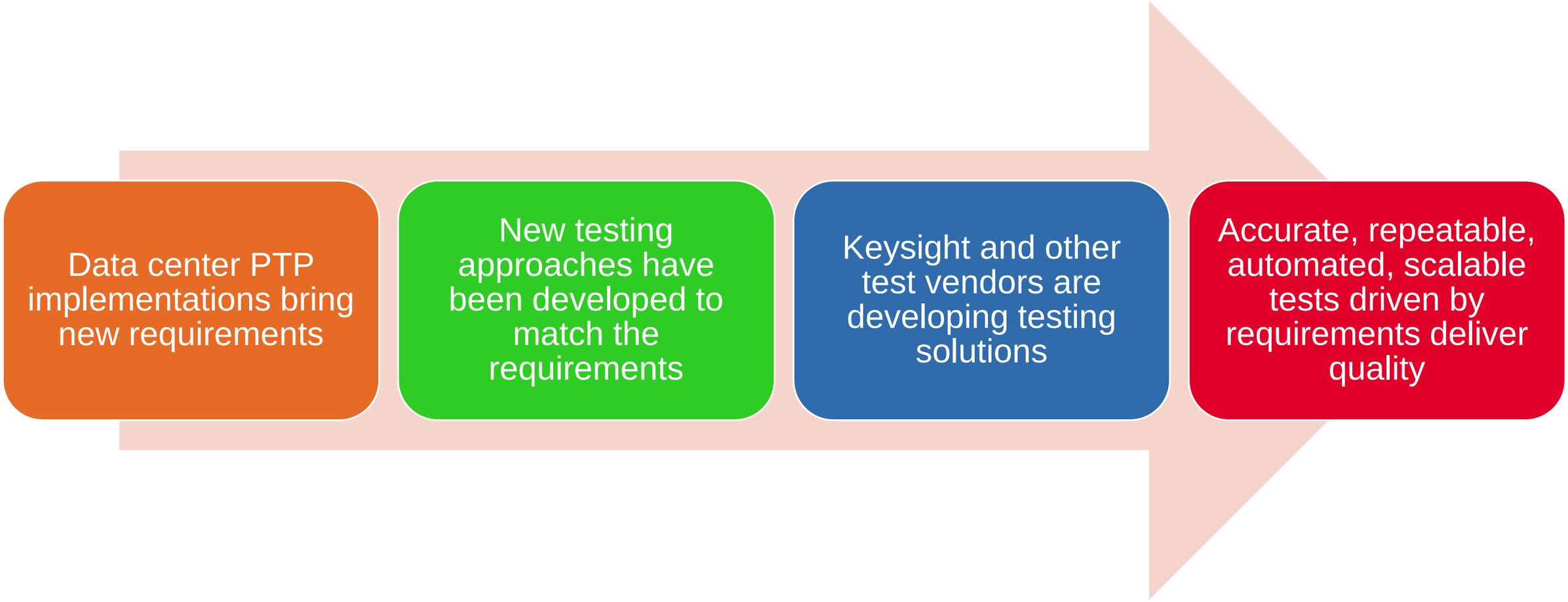
Tests need to support automated, regular, repeatable regression runs

Sample test case summary: measure announce interval

- Purpose: To verify that the DUT's Announce message interval remains within the permitted range as the number of sessions is scaled
- Test Procedure
  - 1. Begin emulating 100 000 clients
  - 2. Ensure all clients successful connect to the GM
  - 3. Measure at least 100 Announce Intervals, preferably consecutively.
  - 4. Increase client count by 100 000 and repeat steps 2-4 until requested scalability is met or test fails.
- Observable Results:

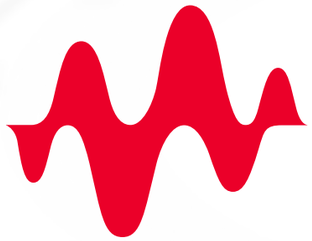
Step	Status	Description
3	FAIL	One or more of the clients did not receive an Announce
3	FAIL	On one or more of the emulated clients, the mean ± the sample standard deviation of the observed Announce intervals is outside the range of ± 30% of the $2^{currentLogAnnounceInterval}$
3	FAIL	On one or more of the emulated clients, more than 10% of the observed Announce intervals are outside the range of ± 30% of the $2^{currentLogAnnounceInterval}$
3	INFO	The number of Announce intervals observed, their max, min, mean, sample standard deviation and median are reported
3	PASS	All Announce intervals are within acceptable tolerances

Conclusion



Questions or comments: email Alon.Regev@keysight.com

Thank you!!



KEYSIGHT
TECHNOLOGIES