OPEN POSSIBILITIES.

Precision Time Protocol Profile for Data Center Applications & Related Network Requirements

OCP GLOBAL SUMMIT

NOVEMBER 9-10, 2021

TAP

PTP Profile for DC Applications & Related Network Requirements

Michel (Meta) and Thomas (NVIDIA)

OPEN POSSIBILITIES.

OPEN PLATINUM™

OCP GLOBAL SUMMIT
NOVEMBER 9-10, 2021

TIME APPLIANCES

Agenda

Why Synchronization?
Application & Time Error Requirements
Time Synchronization Service
PTP Profile Aspects
Call to Action

OCP GLOBAL SUMMIT
NOVEMBER 9-10, 2021

Why Synchronization in Data Centers

- Provide a reliable time synchronization service across the infra of a data center
- Enable set of new applications
- Improve set of current applications
- Using Precision Timing Protocol (PTP), increase the level of accuracy by 2 to 3 orders of magnitude beyond what NTP infra offers today

> Nanosecond-level clock synchronization can be an enabler of a new spectrum of timing- and delay-critical applications in data centers
> - Stanford / Google paper

OCP-TAP Applications

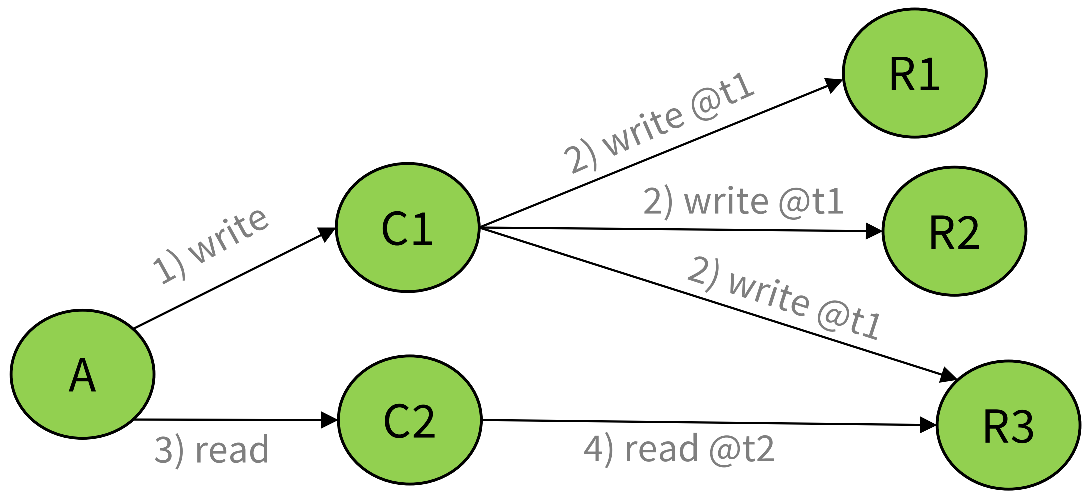
- Applications discussed within OCP-TAP:
  https://www.opencompute.org/wiki/Time_Appliances_Project

Distributed database systems	- Increase system transactions via stringent time bound guarantees - Reduce guard band between read-write operations across machines
Network monitoring	- Early detection of anomalies (e.g., latency, loss, congestion) - One-way delay (OWD) measurements between any given machine
Telco Cloud Radio Access Network	- 5G air interface synchronization - Time distribution across midhaul/fronthaul networks
Enterprise	- Cloud & enterprise applications - Financial services

Distributed Database - Example

- If t2 < t1 due clock skew, application will see wrong information

- Ordering of operations is necessary, but not always sufficient
- Ordering in time leads to improve performance but requires strict clock skew guarantees between machines (e.g, to enable property of linearizability)



OPEN POSSIBILITIES.

OCP GLOBAL SUMMIT
NOVEMBER 9-10, 2021

Time Synchronization Service & PTP Profile

- Significant advances in time distribution & synchronization across packet networks in the past decade
- Multiple industries rely on high accuracy & reliable time distribution
- Many products and networks around the world run PTP today
- IEEE 1588 (PTP) is a large specification with many capabilities and options to choose from
- A “PTP Profile” defines the capabilities that are required to support a particular use case scenario

A "PTP Profile" is an essential part of a "Time Synchronization Service"

PTP Profiles in the Industry

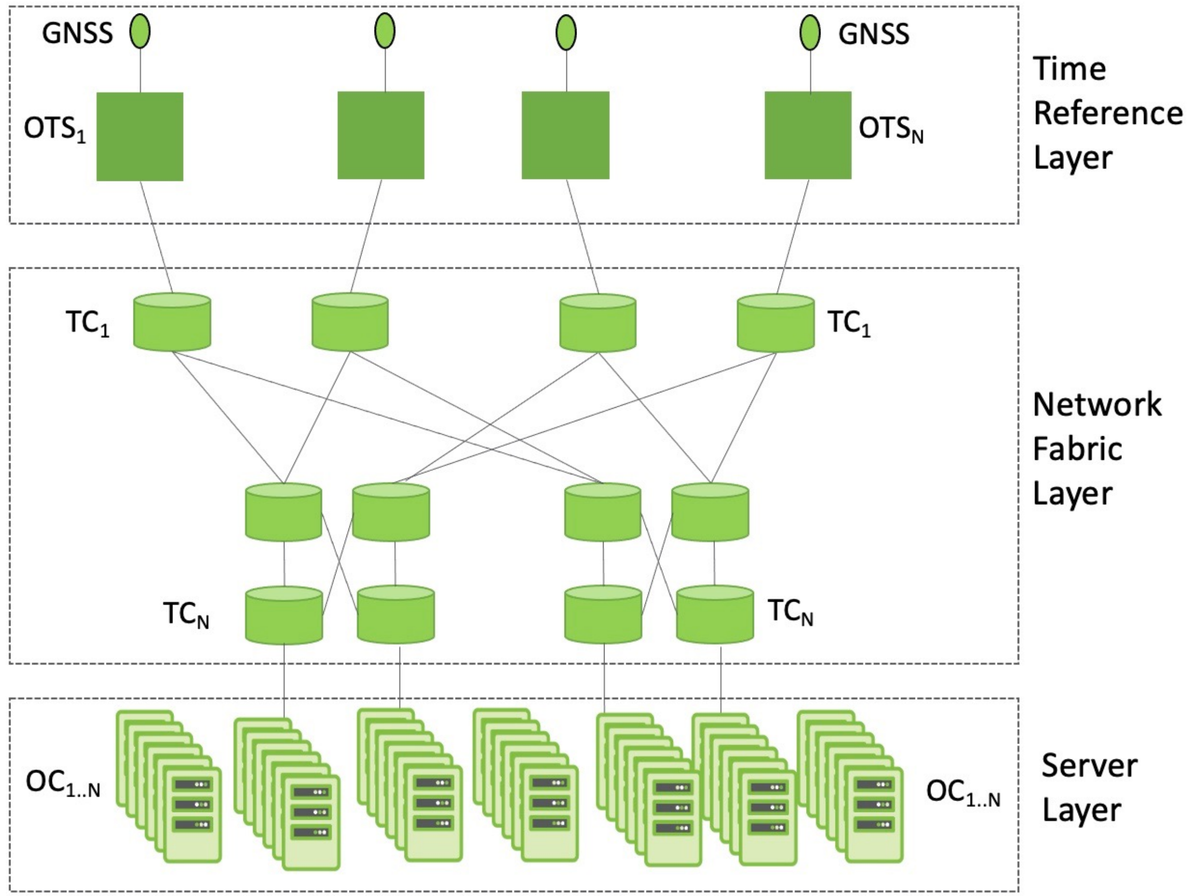
Industry	Application	Specification
Telecom & Mobile	Sync for 2G/3G/4G/5G base stations & fronthaul networks	ITU-T G.8265.1 ITU-T G.8275.1, G.8275.2
Professional Audio/Video	Sync for video/audio feeds between sources and receivers	SMPTE ST 2059-2
Power	Sync for substation sampled values, synchrophasor, power protection	IEEE C37.238-2017 IEC 61850-9-3 & IEC 62493-2 Annex A.2
Audio/Video, Industrial, Automation, Automotive	Sync of A/V applications with high QoS/QoE demand and time sensitive networks	IEEE Std 802.1AS-2020
Industrial Automation	Sync for industrial plants, machine-to-machine real-time control	IEC 62439-3 Annex B IEC 62439-3 Annex C
Enterprise/Financial	Sync of time tagged and packet latency measurements	draft-ietf-tictoc-ptp-enterprise-profile-21
Data Center	Sync for time-sensitive applications within data center	OCP DC PTP Profile #1 (released Sept 2021)

“Time Sync Service” Reference Model

Time Reference Layer:
- Rootftop antennas, GPS system
- Open Time Server (OTS)

Network fabric Layer:
- Large set of PTP-aware switches
- e.g., Transparent Clock (TC)

Server Layer:
- Very large set of server machines
- End applications requiring time
- HW timestamping

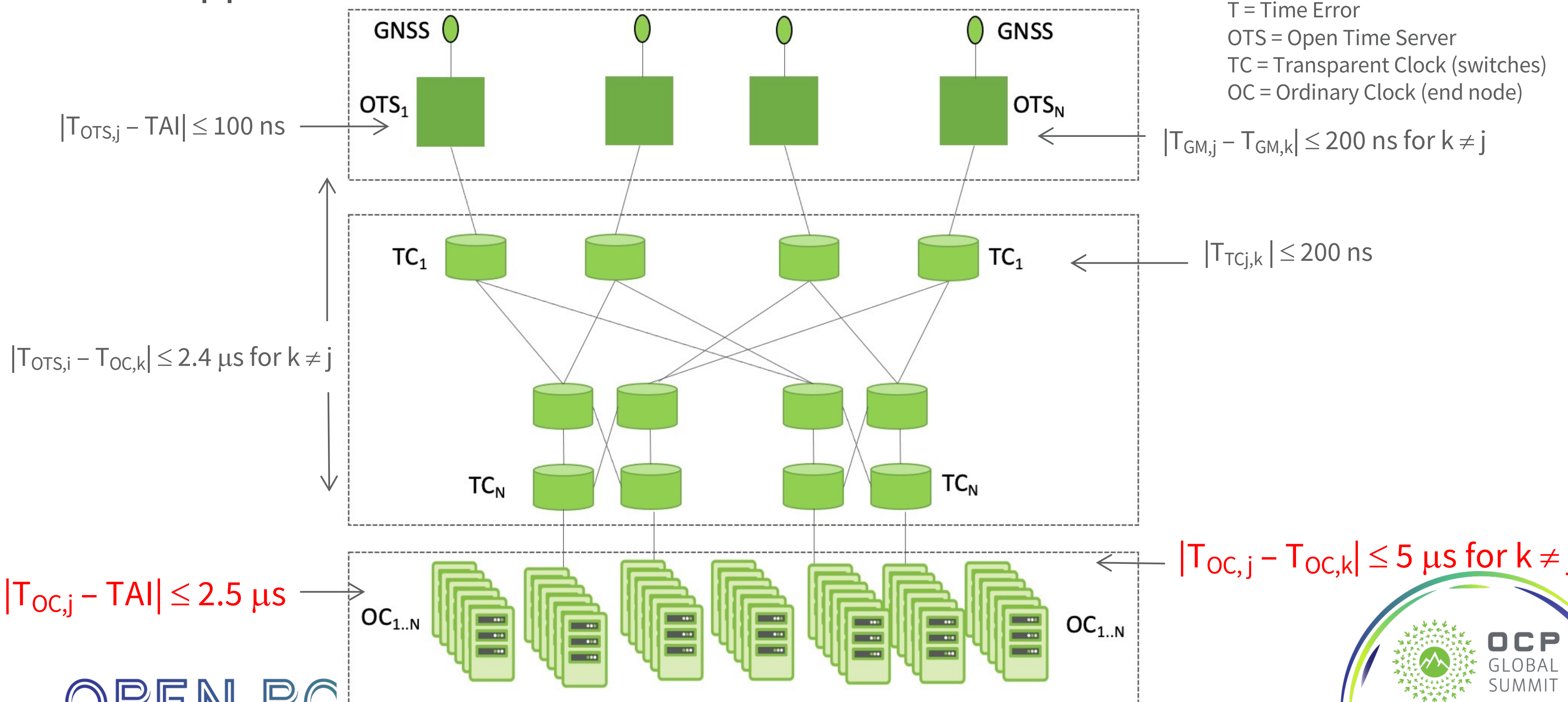


OPEN POSSIBILITIE

“Time Sync Service” Error Requirements

- End Application & Infra

T = Time Error
OTS = Open Time Server
TC = Transparent Clock (switches)
OC = Ordinary Clock (end node)



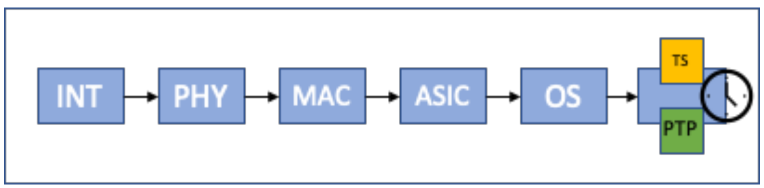
OPEN PC

OCP GLOBAL SUMMIT
NOVEMBER 9-10, 2021

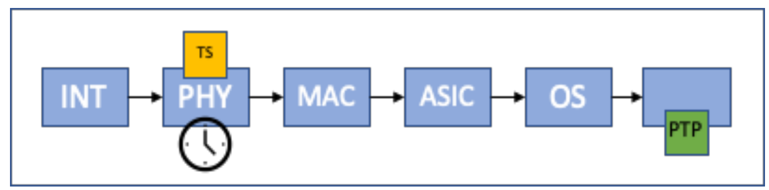
Software vs. Hardware Timestamping

Software timestamping don’t provide a high accuracy and deterministic behaviour (10 to 100 microseconds) due to system noise, latency, scheduling

Hardware timestamping pulls timestamps as close as possible to the MAC with minimal overhead (sub 10ns in modern implementations)



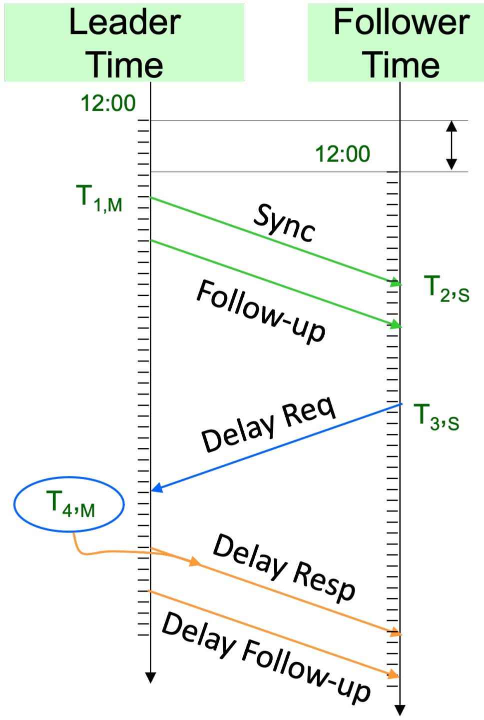
SW timestamping: TS, Clock & PTP



HW timestamping: TS & PHC vs. PTP

1-step vs. 2-step clock

- Historically (1588v1 era), 2-step SW implementations were available (HW limitations/availability)
- 2-step doesn't write timestamps to PTP messages as they egress. Older HW couldn't encode 1-step timestamp messages at higher interface rates (10 Gbps+), this isn't the case anymore
- 1-step guarantees each sync message is correctly linked up with its timestamp. Especially when there may be multiple possible network routes (Sync vs. Follow_Up) (OoO, packet spraying across links)



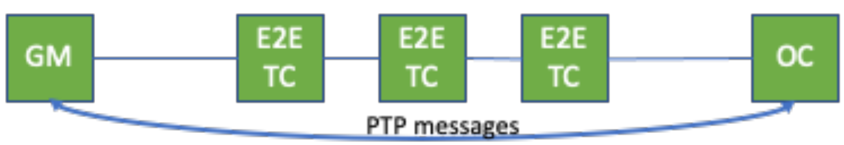
OPEN POSSIBILITIES.

OCP GLOBAL SUMMIT
NOVEMBER 9-10, 2021

Switches & capabilities – Model 1 & 2

Model 1: Transparent Clock

- E2E PTP msgs sprayed across links
- 1-step HW TC to avoid OoO msgs
- Network routing for failure recovery

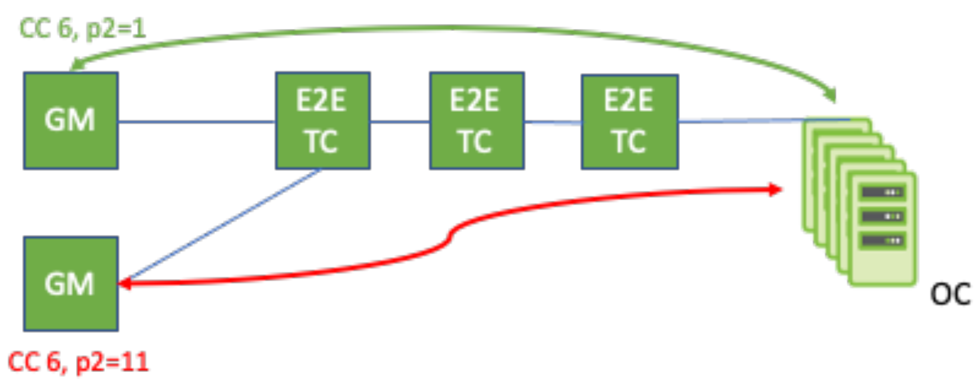


Model 2: Boundary Clock (proposed)

- PTP messages processed hop-by-hop
- BC is SW based and handles:
  - PTP dataset analysis
  - BMCA for failure/path recovery
- BC chain may impact overall settling time
- 1-step or 2-step clock may be used

Failure/redundancy mechanism

- An OC Follower can have PTP sessions configured to multiple GMs (Open Time Server)
- Over IPv6 Unicast transport in Model 1 (E2E Transparent Clocks)
- PTP Stack running on the OC compares PTP Dataset via BMCA
- Select 'Best GM' based on P1, ClockClass, ClockAccuracy, ClockVariance, P2, SrcID
- Network reachability with candidate GMs relies on IGP



OPEN POSSIBILITIES.

OCP GLOBAL SUMMIT
NOVEMBER 9-10, 2021

Data Center PTP Profile

Attributes, Parameterization & Configuration

…putting the end-to-end PTP pieces together to meet performance objectives

PTP Attributes	OCP Specification
OCP Company ID (CID)	7A-4D-2F
Timing awareness	PTP aware in all elements (switch, NICs, open time server, etc.)
PTP Clock types	GM, TC, OC
PTP Messages	Announce, Sync, Follow_Up, Delay_Req, Delay_Resp, Signaling, Management
Network transport	IPv6, IPv4
PTP path delay measurement	End-to-End
Domain Number	0
Clock Operations	One-step and Two-step for GM, OC One-step for TC
PTP Message rates	Announce {0, -4} Sync {+3, -7} DelayReq/Resp {0, -7}
Network Communication	Unicast discovery & Unicast negotiation Multicast is prohibited
Redundancy	Active-Standby Active-Active (future)
Open Time Server Clock Class	6 (traceable) 7 (holdover, within spec) 52 (holdover, out of spec)
Open Time Server Clock Accuracy	0x21 (±100 nanoseconds)

OPEN POSSIBILITIE

Call to Action

- Data Center PTP Profile v1 was released on Sept 21
- Get involved in OCP-TAP workstream #2 to further develop PTP Profile v2
  - Reference Model with Boundary Clock
  - Security aspects
  - Load balancing of PTP unicast sessions
- Project Wiki: https://www.opencompute.org/wiki/Time_Appliances_Project
- DC PTP Profile v1: https://github.com/opencomputeproject/Time-Appliance-Project/tree/master/DC-PTP-Profile

Open Discussion