



**IOWN**  
GLOBAL FORUM™

# PoC Project name: An Implementation of Heterogeneous and Disaggregated Computing for DCI as a Service

---

Classification: IOWN Global Forum Recognized PoC

Stage: SSF PoC Report

Confidentiality: Public

Version: 1.0

September 6, 2023

## 1. Introduction

In the Area Management Security Use Case of IOWN GF, it is required to dramatically improve the energy efficiency of the computing infrastructure for processing data in real-time image analysis of video streams from massive security cameras. As a solution to this problem, IOWN GF provides a DCI (Data-Centric Infrastructure) Functional Architecture document [1] that proposes the concept of “heterogeneous and disaggregated computing” in which a logical service node (LSN) is formed by combining various computing devices flexibly and optimally according to the workload. Also, the Reference Implementation Model (RIM) for the Area Management Security Use Case document [2] shows how to utilize in DCI in the same use case.

This PoC report describes, as an example of DCI implementation, a demonstration system employing an LSN configuration that combines multiple types of accelerators, such as GPU and FPGA, and shows the effects of applying the system to part of the CSP Area Management Security use cases.

## 2. PoC Project Completion Status and Project Participants

This PoC project is a single-phase project. A PoC demonstration was conducted at the event described in Section 3. The performance measurements and evaluations were conducted using the demonstration system described in Section 7. Section 7 also describes the relationships to the PoC Reference [4]. Although it is intended to be a single-phase project, as there might be some successor PoC projects in future, some additional performance evaluations would be considered using the PoC system in this PoC project.

- Overall PoC Project Completion Status: Completed
- PoC stage: Significant Step Forward (SSF)

PoC Team:

- PoC Project Name: DCI as a Service for CPS Area Management Security (DCI- # 2)
- Member A: NTT Contact: Ryosuke Kurebayashi
- Member B: Fujitsu Contact: Takayuki Uchihira

## 3. Confirmation of PoC Demonstration

This PoC demonstration was part of the exhibit, “NTT R&D Forum 2022 – Road to IOWN”. The forum was held in a hybrid format of online and offline. This section mainly describes on-site exhibition, i.e., physically performed demonstration. In addition, the flyer is presented in Appendix B as documented evidence for the exhibition.

The details of the event including the PoC exhibit are shown below. Figure 3 -1 is a photograph of the exterior of the exhibit in the demonstration. The photograph shows AI-processed footage of daytime and

overnight and graphs about power consumption. The subsystem providing the LSNs (Logical Service Nodes) is in a separate room and is not shown in the photograph.

#### PoC Demonstration Event Details:

- Event Name: NTT R & D Forum 2022 – Road to IOWN
- Event Web page: <https://www.rd.ntt/e/forum/2022/index.html>
- Event Venue: NTT Musashino R & D Center, Tokyo, Japan
- PoC Demonstration Name: A-N02: "White-box server technology for IOWN Smart City"
- PoC Demonstration Picture:



Figure 3 - 1: Demonstration system of this PoC

- Press release related to the PoC Demonstration: Developed next-generation computing infrastructure that enables optimal use of diverse computing resources - Four times power efficiency improvement confirmed in video camera analysis of smart cities - [3]

#### 4. PoC Goals Status Report

The goal of this PoC project is to improve the energy efficiency of the computing infrastructure by optimally configuring a variety of accelerators. In the PoC demonstration, we conducted the demonstration of the prototype of DCI-as-a-Service for the AI inference of video stream as the initial target. The DCI-as-a-Service includes the accelerator-related technology with high energy efficiency and the disaggregated computing technology which realizes the optimal combinations of the accelerators with energy efficiency according to the time of day.

- PoC Project Goal #1: To improve energy efficiency by optimally configuring accelerators  
Goal Status (Demonstrated and Met)

## 5. Supposed Use Case

### 5.1. Use Case Scenario Supposed in the PoC

Eight 4K cameras are used to detect people. The assumed scenario shows the reduction of the energy consumption when the configuration is changed by following the change of the flow of people according to the time of day for the processing of the human detection. Assumption is made that number of trespassing humans is high in daytime, hence all video streams require higher fps in daytime while there is no such requirement at night. To reflect this in the deployment, a different set of accelerators is assigned according to the processing and calculation load required for each inference activity.

### 5.2. Relationship to CPS Area Management Security Use Case

The Section 3 of the PoC Reference [4] lists three IOWN GF use cases as examples of reference cases. The PoC performed here corresponds to the use case of Area management with security monitoring.

In the Area management with security monitoring use case, footage recorded with multiple cameras are aggregated on a local aggregation node. The ingestion node then analyzes the video streams using AI inference to extract events. In addition, the data hub node stores the events and triggers necessary actions.

The target scope of this PoC is shown in Figure 5-1. In this PoC, the component corresponding to the ingestion node is implemented for the use-case scenario described above. The ingestion node receives the video stream of each camera as an input, then performs video decoding, filtering and resizing, finally outputs an event (human detection) as a result of the AI inference.

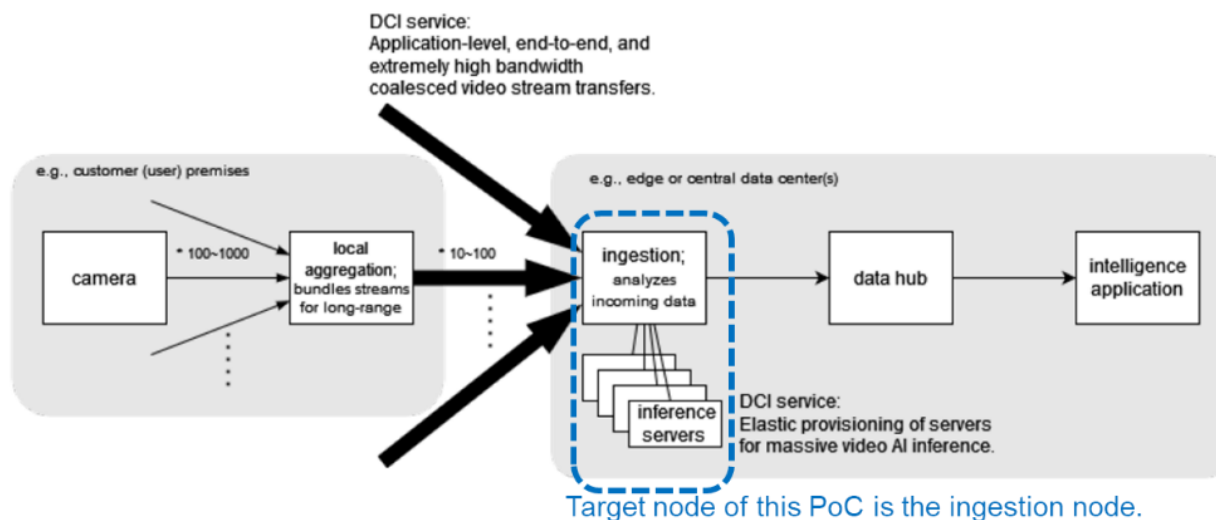
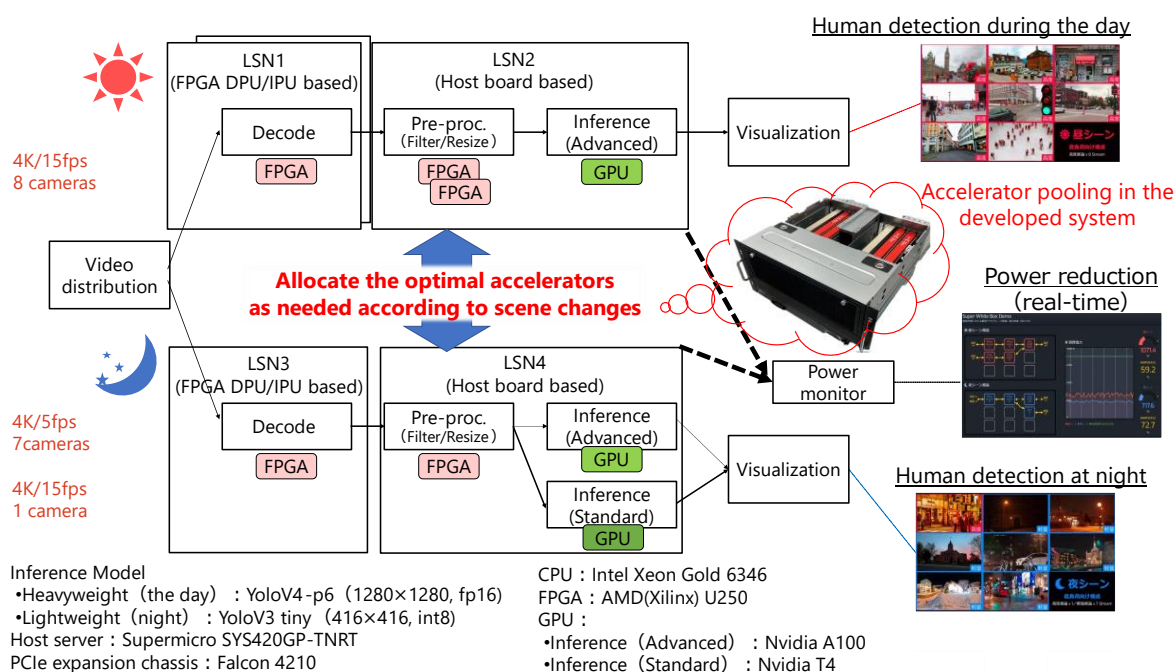


Figure 5-1: Target scope of this PoC

## 6. PoC System Configuration and Implementation

### 6.1. Overview of PoC System Configuration

An outline of the system configuration of this PoC is shown in Figure 6 -1. This system selects suitable ones from various accelerator pools and constructs a highly efficient pipeline according to the use situation. Concretely, two pipelines are constructed according to the daytime use scene and the nighttime use scene, respectively. Further, energy efficiency is improved by linking accelerators with high arithmetic efficiency with less involvement of the CPU. In this report, a group of accelerators and host boards that share a PCIe bus for user plane data communication is regarded as an LSN.



Note: The FPGA cards need to be configured through the host CPUs. Requirements on how LSNs should be configured is being studied for DCI Functional Architecture 3.0 or later.

Figure6-1 : An outline of the system configuration of this PoC

**Inference (Advanced):** Assume that humans are in the images and perform high-precision human detection.

- Choose high-precision inference models and high-performance GPU devices for detection accuracy
- Make inferences from high quality images (1280 x 1280)
- frame rate is 15 fps

**Inference (Standard):** It is assumed that there are no humans in the images, and it detects humans at a coarse confirmation level with less power.

- Prioritize energy efficiency, choose a low-power inference model, choose a low-power GPU device
- Make inferences from low-quality images (416 x 416)
- frame rate is 5 fps

## 6.2. Detailed System Configuration and Implementation

The overall configuration of the PoC system is shown in Figure 6-2. The system consists of 3 servers (SYS-420 GP-TNR) including a video distribution and a display server, and an accelerator pool using PCI expansion box (Falcon 4210).

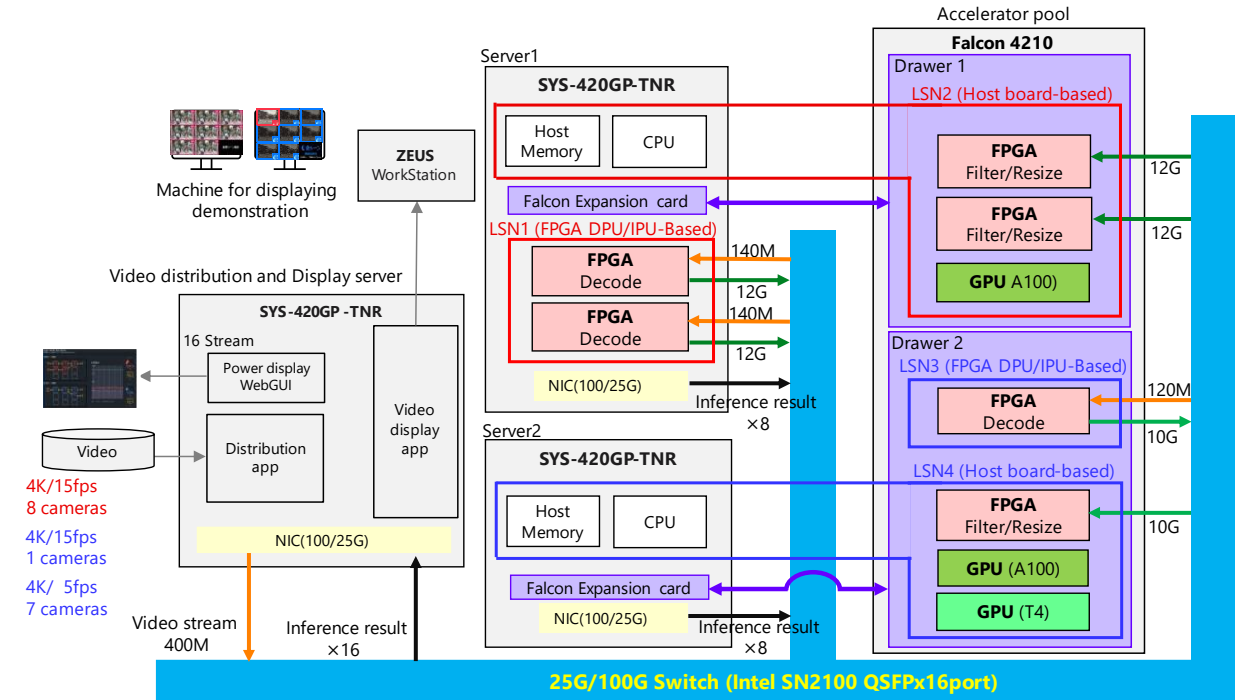


Figure6-2 : The overall configuration of the PoC system

The flow of data on the PoC system is shown in Figure 6 -3. The data pipeline in the daytime use scene consists of LSN1 (FPGA DPU/IPU-Based) and LSN2 (Host board-based). The LSN1 receives the video streams (H. 264) of the 4K/8 camera from the distribution server and performs decoding. The LSN2 receives the decoded video streams, applies preprocessing such as median filter and resizing to the video, and then executes inference processing for performing human detection. To accelerate each processing in the LSN, an FPGA for decoding is allocated as an accelerator in the LSN1, and an FPGA for preprocessing and a GPU for inference processing are allocated as accelerators in the LSN2. Similarly, the data pipeline is constructed using two LSNs, LNS3 (FPGA DPU/IPU-Based) and LNS4 (Host board-based) in the nighttime use scene.

Next, a communication method of the LSN will be described. In the daytime use scene, the video stream from the distribution server is inputted to the LSN1 using the RTP over UDP protocol. To the LSN2 at the later stage, the decoded video stream is transferred via SW using the 25 G AOC cable. In this case, it is transferred by the proprietary protocol over TCP/IP. Thereafter, the inference result by the LSN2 is transferred to a display server at a later stage via SW using the RTP over UDP protocol. The LNS3 and LNS4 for the night use scene also use the same protocols as in the day use scene.

The RDMA protocol itself is not used in the reception of video streams in the LSNs 1 and 3 and in the transfer of decoded video stream between the LSN1-LSN2 and the LSN3-LSN4. However, the protocol processing is terminated by the FPGA, and the required data is deployed on the memory for the user circuit of the FPGA (LSN1, LSN3) or on the host memory (LSN2, LSN4) to obtain the same offload effect of the protocol processing as RDMA. Furthermore, the LSN1 and LSN3 complete the decoding process and the

transmission process to the post-stage LSN in the FPGA and realize the autonomous communication not through the CPU.

Next, we describe a communication scheme between a preprocessing FPGA and an inference processing GPU in LSN2 and LSN4. A special DMA-based data transfer with less CPU overhead is performed between an FPGA and a GPU via a host memory.

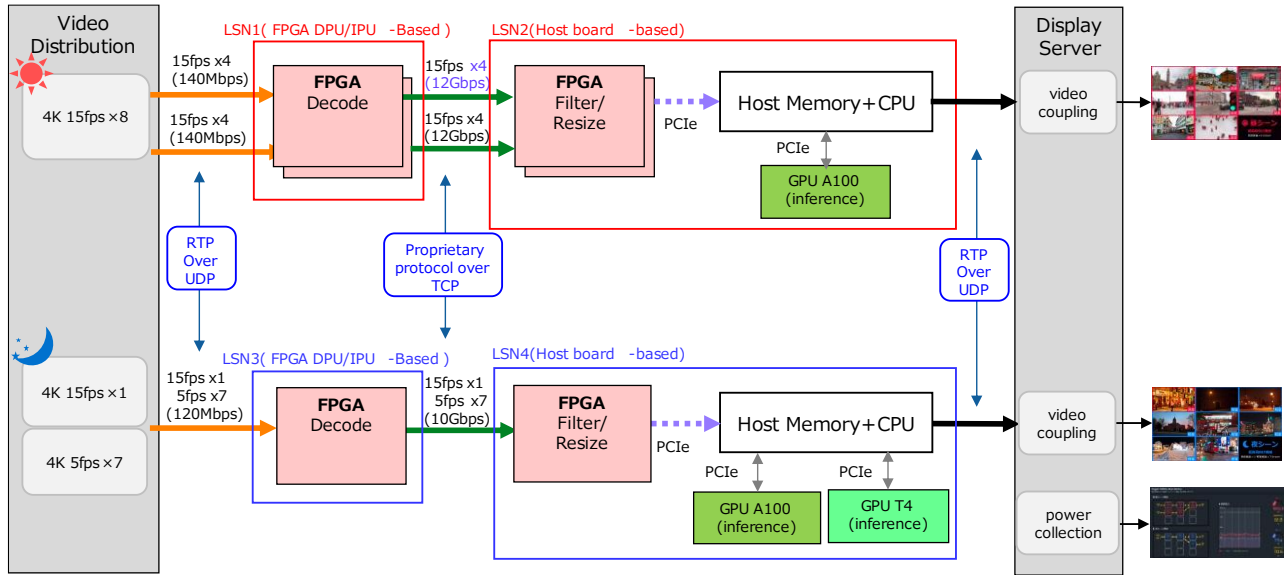


Figure6-3 : The flow of data on the PoC system



## 7. Performance Measurement and Evaluation of the PoC System

We selected the aspects of the measurement and evaluation with taking the PoC Reference [4] into account. The requirements and expectations in the PoC Reference are summarized in Appendix C.

### 7.1. Target of Performance Evaluation

Measurement points in the PoC system are shown in Figure 7-1. Frame loss measurement, frame rate measurement, and power consumption measurement were performed at the following points.

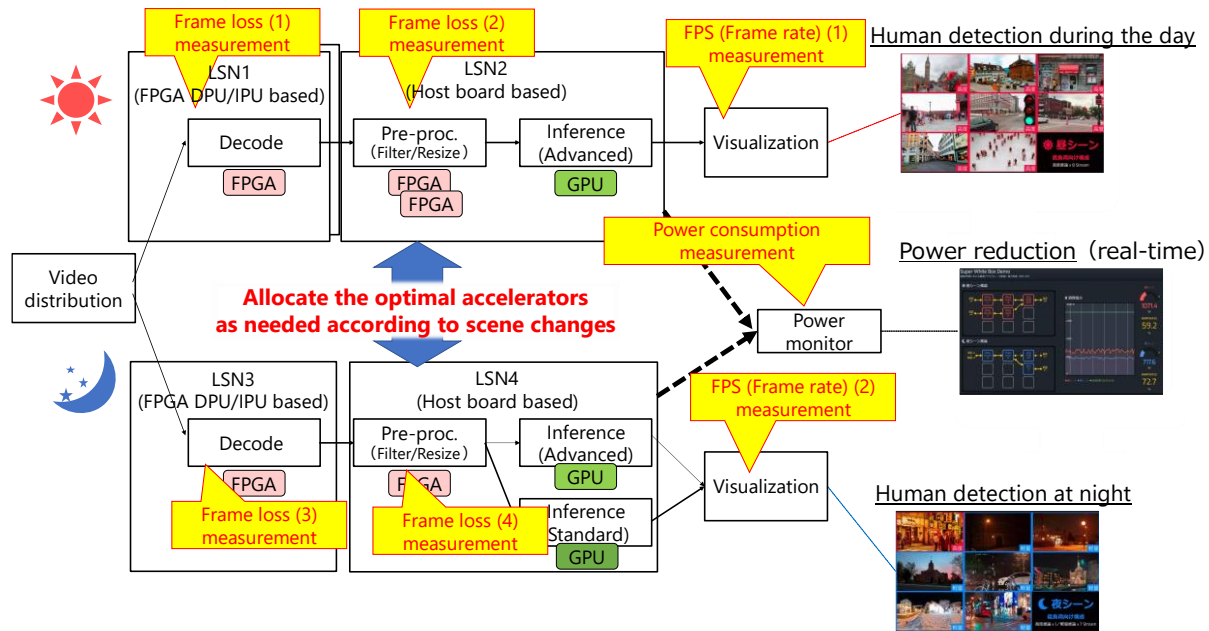


Figure7-1 : Measurement points in the PoC system

## 7.2. Measurement System and Results

### 7.2.1. Measurement of Frame Loss

It was confirmed that there is no frame loss at the measurement points of Frame loss (1) to (4).

### 7.2.2. Measurement of Frame Rate

It was confirmed that there was no significant decrease/increase at various fps measurement points. An excerpt of the frame rate measurements is given in Table 7-1.



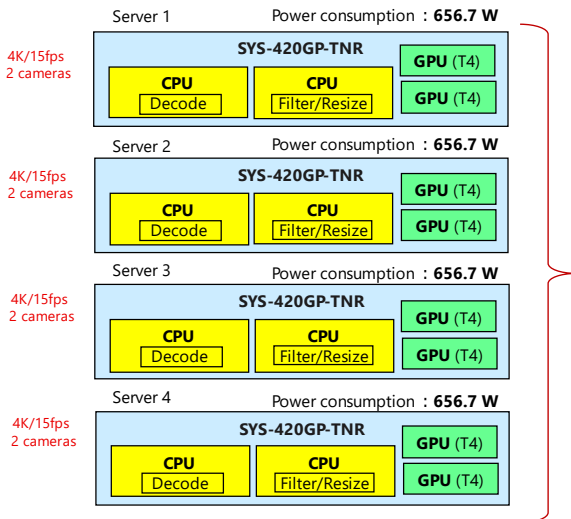
Table7-1 : An excerpt of the frame rate measurements (fps)

☀️ 4K/15fps/8 cameras	🌙 4K/15fps/1 camera 4K/5fps/7 cameras
09:52:22 perf: stream1; <b>fps: 14.977</b>	13:31:13 perf: stream1; <b>fps: 15.098</b>
09:52:22 perf: stream2; <b>fps: 15.230</b>	13:31:13 perf: stream2; <b>fps: 4.674</b>
09:52:22 perf: stream3; <b>fps: 15.514</b>	13:31:13 perf: stream3; <b>fps: 5.171</b>
09:52:22 perf: stream4; <b>fps: 15.027</b>	13:31:13 perf: stream4; <b>fps: 5.466</b>
09:52:22 perf: stream5; <b>fps: 14.996</b>	13:31:12 perf: stream5; <b>fps: 5.225</b>
09:52:22 perf: stream6; <b>fps: 15.048</b>	13:31:13 perf: stream6; <b>fps: 5.035</b>
09:52:22 perf: stream7; <b>fps: 14.988</b>	13:31:13 perf: stream7; <b>fps: 4.984</b>
09:52:22 perf: stream8; <b>fps: 15.683</b>	13:31:13 perf: stream8; <b>fps: 5.073</b>

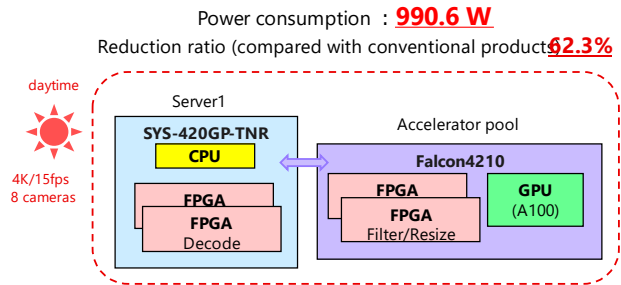
### 7.2.3. Measurement of Power Consumption

As a result of improving the efficiency of the pipeline by selecting the optimum accelerators from the dedicated accelerator pool, offloading to the dedicated accelerators in a wider range, and highly efficient data transfer between dedicated accelerators, approximately 60% reduction of power consumption was obtained compared with the conventional configuration. The FPGA circuit of this PoC configuration is a prototype implementation, and further reduction can be expected by improving the implementation. The details of the conventional configuration, the present PoC pipeline configuration, and the reduction of power consumption are shown in Figure 7-2.

#### Conventional configuration



#### Configuration of this PoC ( daytime )



Power consumption Total :  $656.7 \times 4 = 2626.8 \text{ W}$

Figure7-2 : Power consumption measurement (compared with conventional products)

Detailed results of power consumption measurement for the PoC system are shown below in Table 7 -2.

Table7-2 : Power consumption measurement results(Day Time)

Power[W] traffic	Server	CPU	FPGA				GPU	Other	Total
			Dec_0	Dec_1	F/R_0	F/R_1	A100		
15fpsx8	537.7	(216.0)	(49.0)	(50.1)	40.0	41.1	277.0	94.8	990.6

※The value in parentheses is included in the server power value.

In addition, by flexibly changing the pipeline configuration according to the use scene, about 75% reduction of power consumption was obtained. Figure 7-3 shows the details of the pipeline configuration change and the reduction of power consumption.

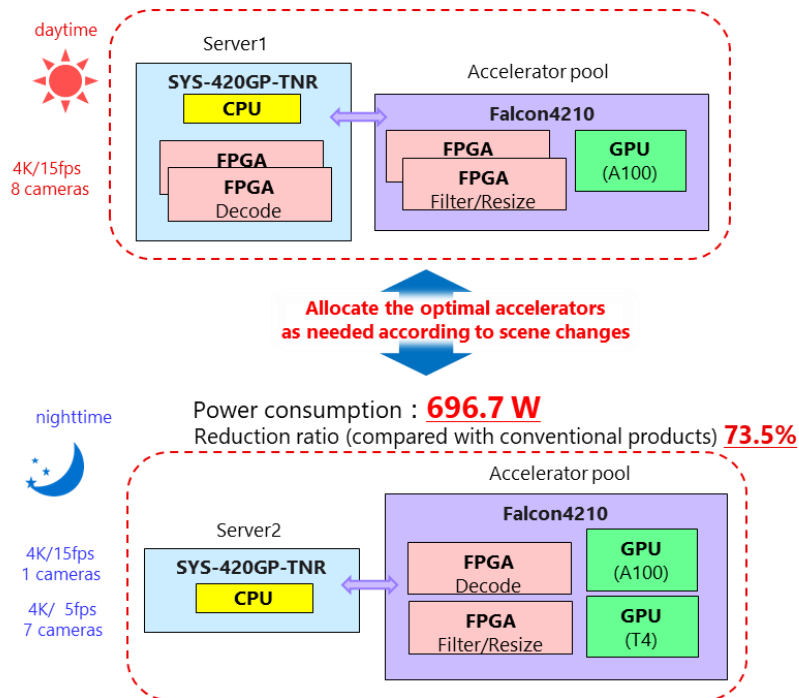


Figure7-3 : Power consumption measurement (when configuration is changed)

The detailed results of power consumption measurement when the configuration was changed are shown in Table 7 -3.

Table7-3 : Power consumption measurement results(Night Time)

Power[W] traffic	Server	CPU	FPGA		GPU		Other	Total
			Dec_0	F/R_0	T4	A100		
15fps x1 ,5fps x7	359.6	(131.5)	54.1	40.5	31.9	116.8	93.8	696.7

※The value in parentheses is included in the server power value.

### 7.3. Results of Performance Evaluation and Technical Finding

Objective Id:	AM1/S1/1
Description:	<b>Description of the PoC Demo Objective:</b> ✓ To show on-demand provisioning of LSNs (logical service nodes), ✓ To demonstrate enabling and release of LSNs [See PoC Reference [4] Section 3.1.3 and Section 4.1.1]
Pre-conditions	As on-demand provisioning of LSNs is not possible in this PoC, the evaluation gets started with the state that the LSN is ready. In addition, it is impossible to enable and release by remote command, so powering on and off the LSNs are deemed enabling and releasing of LSNs respectively.
Procedure:	1 Each environment of daytime human detection and nighttime human detection is defined as LSN.
	2 Enable and release by powering on/off individually for daytime and nighttime human detection environments
	...
Finding Details:	
Lessons Learnt & Recommendations	We have not been able to provide LSN on demand this time, but we have built an LSN environment that statically controls a fixed environment with accelerators installed in the Falcon (Accelerator pool) and accelerators installed in a server and realizes daytime and nighttime human detection by increasing or decreasing FPGA/GPU. In this case, manual operation occurs, and there are many problems in agility and man-hours. It was proven that LSN offer and enabling on demand were important elements for the practical operation.

Objective Id:	AM1/S1/2
Description:	<b>Description of the PoC Demo Objective:</b> ✓ To demonstrate two or more different LSN composition patterns [See PoC Reference [4] Section 4.1.1]
Pre-conditions	After the LSNs are composed by hand, the evaluation gets started with the state that the LSNs are ready.
Procedure:	1 Each environment of daytime human detection and nighttime human detection is defined as LSN.
	2 2-1. Power OFF the daytime and nighttime human detection environments 2-2. Power on the daytime Human detection environment 2-3. Multiple video streams from the video distribution server are sent to the daytime human detection environment, human detection is performed, and the detection results are passed to the display server. 2-4. Power of the daytime human detection environment, power on the nighttime human detection environment 2-5. Multiple video streams from the video distribution server are sent to the night human detection environment, human detection is performed, and the detection results are passed to the display server. 2-6. Power OFF only for nighttime human detection environment
	...
Finding Details:	

Lessons Learnt & Recommendations	As a result, it was found that the system can flexibly cope with the workload where load fluctuation occurs by appropriately performing processing in a plurality of LSNs, and that it has a large effect from the viewpoint of power consumption.
----------------------------------	--

Objective Id:	AM1/S1/3								
Description:	<p><b>Description of the PoC Demo Objective:</b></p> <ul style="list-style-type: none"> <li>✓ To demonstrate provisions and enables use of communication links with extreme traffic parameters between LSNs,</li> <li>✓ To demonstrate integrating and removing LSNs from the communication network,</li> <li>✓ To show measurement results of the KPIs (e.g. latency) of end-to-end connections between LSN ports (server NIC ports) and endpoints in application software</li> </ul> <p>[See PoC Reference [4] Section 4.1.1 and Section 5.2.2]</p>								
Pre-conditions	<p>In this PoC, there is no software stack capable of provisions and enables use of communication links. the evaluation gets started with the state that the communication links are ready.</p> <p>As some KPIs are measurable, the evaluation is conducted on the following KPIs:</p> <ul style="list-style-type: none"> <li>- maximum bandwidth,</li> <li>- corresponding packets drop rates (alternatively bit error rates), and</li> <li>- network service availability (calculated per different time quanta)</li> </ul> <p>On the other hand, some KPIs are not measurable, the evaluation is not conducted on the following KPIs:</p> <ul style="list-style-type: none"> <li>- minimum communication latency,</li> <li>- minimum jitter (variation of the latency)</li> </ul>								
Procedure:	<table border="1"> <tr> <td>1</td> <td>We define LSN (FPGA DPU/IPU-based) and LSN (Host board-based) as LSNs and measure the performance using 4K 15 fps/5 fps video streams.</td> </tr> <tr> <td>2</td> <td>The frame loss measurement was carried out by LSN (FPGA DPU/IPU-based).</td> </tr> <tr> <td>3</td> <td>In addition, the frame loss measurement was carried out by LSN (Host Board-based).</td> </tr> <tr> <td>4</td> <td>Frame rate measurements were performed at the Visualization location.</td> </tr> </table>	1	We define LSN (FPGA DPU/IPU-based) and LSN (Host board-based) as LSNs and measure the performance using 4K 15 fps/5 fps video streams.	2	The frame loss measurement was carried out by LSN (FPGA DPU/IPU-based).	3	In addition, the frame loss measurement was carried out by LSN (Host Board-based).	4	Frame rate measurements were performed at the Visualization location.
1	We define LSN (FPGA DPU/IPU-based) and LSN (Host board-based) as LSNs and measure the performance using 4K 15 fps/5 fps video streams.								
2	The frame loss measurement was carried out by LSN (FPGA DPU/IPU-based).								
3	In addition, the frame loss measurement was carried out by LSN (Host Board-based).								
4	Frame rate measurements were performed at the Visualization location.								
Finding Details:									
Lessons Learnt & Recommendations	<p>Only measurable points were evaluated this time. No frame loss occurred at four measurement points. The frame rates measured on the display side were 15 fps = (14.0 ~ 16.0 fps) for the daytime and 5 fps = (4.5 ~ 5.5 fps) for the nighttime.</p> <p>To perform many measurements, it is important to consider implementing measurement functions from the beginning of PoC development.</p>								

Objective Id:	AM1/S1/4
Description:	<p><b>Description of the PoC Demo Objective:</b></p> <ul style="list-style-type: none"> <li>✓ To show how the technology will influence the overall system energy efficiency</li> </ul> <p>[See PoC Reference [4] Section 5.1.1]</p>
Pre-conditions	

Procedure:	1	Subjects and methods of power consumption measurement (outline) are shown below. <ul style="list-style-type: none"> <li>✓ Server: obtained by ipmitool. * Supermicro sys -420 GP-TNR (CPU: Intel Xeon Gold 6346)</li> <li>✓ FPGAs: obtained with XILINX XRT</li> <li>✓ GPU: obtained with NVIDIA nvidia-smi tool</li> <li>✓ CPU: obtained with Intel pcm-power tool</li> <li>✓ Falcon 4210: obtained with redfish API</li> <li>✓ 100 G SW: obtained directly by connecting to the console port.</li> </ul>
	2	The above measurement results and the previous GPU device (as of 2020) were utilized in a general configuration and were compared with the case where accurate human detection was always performed.
	...	
Finding Details:		
Lessons Learnt & Recommendations	As a result of constructing a highly efficient pipeline by selecting optimum accelerators from the dedicated accelerator group, we were able to confirm about 60% reduction of power consumption compared with the conventional configuration. In addition, by flexibly changing the pipeline configuration according to the use scene, about 75% reduction of power consumption was also verified.	

Objective Id:	AM1/S1/5	
Description:	<b>Description of the PoC Demo Objective:</b> <ul style="list-style-type: none"> <li>✓ To show the maximum achievable service levels,</li> <li>✓ To show point-to-point streaming at extremely high bandwidths,</li> <li>✓ To demonstrate the viability of the proposed QoS-aware communication stacks in the context of heterogeneous servers (LSNs) up to application level,</li> <li>✓ To demonstrate provisions and enables use of communication links with extreme traffic parameters between the LSNs</li> </ul> [See PoC Reference [4] Section 2, Section 3.1.3, Section 4 and Section 4.1.1]	
Pre-conditions	In this PoC project, the PoC conductor did not impose the maximum network traffic (e.g., bandwidth) on the PoC system. It is due to the limit of accelerator's processing capacity, which is relatively lower than network's data transfer capacity. In this evaluation, the viability of the service using high-definition video stream is to be confirmed.	
Procedure:	1	4 K 15 fps × 8 video streams are received by 2 FPGAs (decode). 4 stream (140 Mbps) → FPGA1 (decode) 4 stream (140 Mbps) → FPGA2 (decode)
	2	The video streams decoded by an FPGA (decode) are transmitted to a filtering / resizing FPGA at a later stage. FPGA1 (decode) → 4 stream (12 Gbps) → FPGA3 (Fil/Resize) FPGA2 (decode) → 4 stream (12 Gbps) → FPGA4 (Fil/Resize)
	...	
Finding Details:		
Lessons Learnt & Recommendations	This time, due to the limited processing power of the accelerator card, streaming data transmission using the network's maximum bandwidth of 25 Gbps to 100 Gbps is not possible. However, it was confirmed that high-definition video stream can be transmitted at 12 Gbps and processed by an accelerator card to provide services. To realize broadband p-to-p	

	streaming, in addition to the optimization of the communication protocol stack, a method for enhancing accelerator processing is also an important element.
--	---

## 8. PoC's Contribution to IOWN GF and PoC Suggested Action Items

### PoC's Contribution to IOWN GF

Contribution	WG/TF	Study Item (SI)/ Work Item (WI)	Comments
Energy efficiency	DCI	n/a	This PoC demonstrated the reduction of power consumption. Concretely, about 60% reduction was presented.
Extreme QoS	DCI	n/a	It provides information that contributes to the feasibility of the extreme QoS of the inter-node interconnect of this PoC. Specifically, it stably transmits 4 k/15 fps video stream and 4 k/5 fps video stream. Between the LSN (FPGA DPU/IPU-based) and LSN (Host board-based), stable communication at 12 Gbps is realized with 4 streams of raw data.

### Gaps identified in relevant standardization

Since this PoC is constructed and verified as a system that operates on its own, when cooperation with other systems (orchestrators, etc.) becomes necessary, we plan to investigate and study standardization trends regarding interfaces.

### PoC Suggested Action Items

- It is expected to explore currently available hardware and software that can be used for disaggregated computing.

- In this PoC, LSN is constructed in an offline manner. It is expected that IOWN GF would show how to handle, on-demand delivery of LSNs and dynamic configuration changes.

**Note:** The FPGA cards need to be configured through the host CPUs. Requirements on how LSNs should be configured is being studied for DCI Functional Architecture 3.0 or later.

- In this PoC, we focused on the ingestion node of the Area Management Security use case RIM [2] and showed an example of the configuration of LSNs within a single site. On the other hand, in order to implement the use cases described in the RIM document in E2E, it is necessary to deploy workloads in multiple regional locations. Therefore, it is expected that IOWN GF would clarify the placement of LSN workloads in a distributed environment and how to link with APN, DCI-GW, and inter-node interconnect.

## 9. Next Steps

In this PoC, we demonstrated a DCI implementation for the use case of human detection using video, and confirmed its effectiveness from the viewpoint of power consumption. In the future, we will examine its effectiveness in other use cases where video is used, and in other workloads other than video processing and combinations thereof. In addition, we will strengthen our approach to business by expanding various operation functions for actual operation.

In this system, FPGAs are utilized. In addition, the accelerators are linked without using the CPU. This function is expected to reduce device cost, latency, and jitter compared to conventional CPU-centric architecture. Therefore, we will expand the study of the effects and application areas when these are effective.

## 10. Conclusion

IOWN is an integration of various technologies that will revolutionize social infrastructure. It is expected that this PoC and other PoCs will be organically combined and advanced. We look forward to the cooperation and participation of many companies in PoC activities in IOWN GF.

## Reference

[1] IOWN Global Forum, “Data-Centric Infrastructure Functional Architecture” (2022.01)

[2] IOWN Global Forum, “Reference Implementation Model (RIM) for the Area Management Security Use Case” (2022.01)

[3] <https://group.ntt/en/newsrelease/2022/11/07/221107a.html>

[4] IOWN Global Forum, “Data-centric-infrastructure-as-a-service PoC Reference” (2022.07)

## Appendix A: PoC Configuration Material (Spec)

Table A-1: PoC Configuration Material (Spec)

No.	Equipment	Spec
1	Server, Video Distribution and Display server	Supermicro sys-420GP-TNR (CPU : Intel Xeon Gold 6346)
2	Machine for displaying demonstration	ZEUS WorkStation
3	PCIe expansion chassis (Accelerator pool)	Falcon 4210
4	10Gb Switch (TOR SW)	Netgear XS748T
5	100Gb Switch	Mellanox SN2100-CB2F
6	FPGA Card	AMD (XILINX) Alveo U250
7	GPU Card Inference (Advanced)	NVIDIA A100 (80GB)
8	GPU Card Inference (Standard)	NVIDIA Tesla T4
9	100G NIC (Mellanox)	Mellanox ConnectX 5
10	AOC Cable	Mellanox MFA1A00-C003 compatible 100G Active Optical Cable




## Appendix B: Flyer provided in the PoC Demonstration Event

The exhibition event, which included this PoC demonstration, took place in a hybrid format of online and offline. The visitors to the exhibition could get a flyer about the PoC demonstration as shown in Figure B-1. The flyer has description including the background and features of this PoC.

**A-N02** White-box server technology for IOWN Smart City

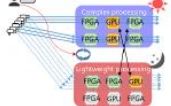
**Improve power efficiency by optimally configuring accelerators**



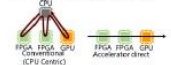
Sustainable Smart City

Two approaches to improve energy efficiency in Smart City


Enhance the flexibility of configuration change  
Policy driven




Enhance the accelerator connectivity  
GPU



High load apps run during daytime  
(object detection, people count)



Low load apps run during overnight  
(object detection)



**Background**

In order to implement IOWN's smart city services, it is necessary to improve the power efficiency when processing a large number of camera images. In particular, the system must be able to flexibly change the computing resources used for video processing as the scene changes.

**Summary**

In order to implement IOWN's smart city services, it is necessary to improve the power efficiency when processing a large number of camera images. In particular, the system must be able to flexibly change the computing resources used for video processing as the scene changes.

**Feature 1**

Flexible configuration of combinations of various processors and function-specific accelerators

**Feature 2**

Improved power efficiency by changing to the optimal configuration for the scene, using the video AI application as an example

**Feature 3**

Improved power efficiency by enabling cooperation between accelerators with high computing efficiency without involving the CPU

**Future benefits**

We will realize a sustainable society by applying white-box server technology as a power-efficient platform to social infrastructure such as smart cities that process vast amounts of sensor information.

**Exhibiting Company**

NIPPON TELEGRAPH AND TELEPHONE CORPORATION

**Contact**

rdforum-exhibition@ml.ntt.com

NTT R&D FORUM – Road to IOWN 2022 Copyright © 2022 Nippon Telegraph and Telephone Corporation All Rights Reserved.

Figure B-1: Flyer about the PoC demonstration

## Appendix C: Requirements and Expectations in the PoC Reference

The table C-1 shows the requirements and expectations in the PoC Reference [4] related to our PoC scope, and how they were treated in our PoC Report.

**Table C-1: Requirement and Expectations in the PoC Reference**

#	Requirements and Expectations	Our PoC Report
1	(Section 2) PoC Reports should report the maximum achievable service levels	<ul style="list-style-type: none"> <li>Described in AM1/S1/5 in 7.3</li> </ul>
2	(Section 3.1.3) PoC conductors should be particularly aware of these two DCIaaS features <ul style="list-style-type: none"> <li>point-to-point streaming at extremely high bandwidths</li> <li>on-demand provisioning of logical service nodes</li> </ul>	<ul style="list-style-type: none"> <li>Point-to-point streaming: Described in AM1/S1/5 in 7.3</li> <li>On-demand provision: Offline configuration of LSN was used in this PoC. See AM1/S1/1 in 7.3</li> </ul>
3	(Section 4) PoCs should aim to demonstrate the viability of the proposed QoS aware communication stacks in the context of heterogeneous servers (LSNs) up to application level. Any DCIaaS PoC Report should provide evaluation results	<ul style="list-style-type: none"> <li>Described in AM1/S1/5 in 7.3</li> </ul>
4	(Section 4.1.1) PoC should demonstrate two or more different LSN composition patterns, together with the enabling and release of such LSNs	<ul style="list-style-type: none"> <li>Described in AM1/S1/2 in 7.3</li> </ul>
5	(Section 4.1.1) a software stack should be demonstrated that <ul style="list-style-type: none"> <li>provisions and enables use of communication links with extreme traffic parameters (e.g., latency, jitter, bandwidth) between these LSNs and</li> <li>allows integrating and removing LSNs from the communication network</li> </ul>	<ul style="list-style-type: none"> <li>Offline configuration of communication links was used in this PoC. See AM1/S1/3 in 7.3</li> </ul>
6	(Section 4.1.1) PoC Reports should explain how LSNs are realized	<ul style="list-style-type: none"> <li>Described in section 6</li> </ul>
7	(Section 5.1.1) PoC conductors are encouraged to report how the technology will influence the overall system energy efficiency	<ul style="list-style-type: none"> <li>Described in 7.2.3 and AM1/S1/4 in 7.3</li> </ul>
8	(Section 5.1.2) PoC implementers are expected to report about the benchmark methodology. [energy efficiency]	<ul style="list-style-type: none"> <li>Described in 7.1</li> </ul>
9	(Section 5.2.1) conducted benchmarks should demonstrate that the realized PoC communication mechanism removes some boundaries of today's technologies regarding IOWN GF use cases [Network service quality]	<ul style="list-style-type: none"> <li>Performance measurement and evaluation results are described in 7.2 and AM1/S1/3 and AM1/S1/5 in 7.3. Note that we mainly focused on the improvement of the energy efficiency of 10Gbps-class data handling.</li> </ul>
10	(Section 5.2.2) PoC Reports should include measurement results of the following KPIs <ul style="list-style-type: none"> <li>minimum communication latency, minimum jitter (variation of the latency), maximum bandwidth, corresponding packet drop rates (alternatively bit error rates), and network service availability (calculated per different time quanta) of end-to-end connections between               <ul style="list-style-type: none"> <li>LSN ports (server NIC ports) and</li> <li>endpoints in application software;</li> </ul> </li> </ul> measurements should be performed between different LSNs	<ul style="list-style-type: none"> <li>Performance measurement and evaluation results are described in 7.2 and AM1/S1/3 and AM1/S1/5 in 7.3. Note that we mainly focused on the improvement of the energy efficiency of 10Gbps-class data handling.</li> </ul>