Data Collection Requirements and Technologies for Digital Twin Network

draft-zcz-nmrg-digitaltwin-data-collection-00

Cheng Zhou (zhouchengyjy@chinamobile.com, **Presenter**)
Danyang Chen(chendanyang@chinamobile.com)
Pdero Martinez-Julia (pedro@nict.go.jp)

Scope & Objectives

- **Scope**
  - Describe the requirements of data collection for building digital twin network (DTN) sytem.
  - Provide data collection methods towarding building DTN data repository.

- **Objectives**
  - Identify the data collection requirements and principles for DTN
  - Call for more efficient data collection methods suitable for DTN system
  - Reach a consensus on selecting data collection methods for various network data

Change Log

Versions:

00

- **History**
  - Initially, the draft was just a specific data collection method for DTN
  - Promote the daft to extend the scope to general data collection requirements and methods for DTN
- **Major Changes in this version**
  - Re-scope the draft
  - Add section of data collection requirements for DTN
  - Refine the text, and some editorial changes

Contents

Table of Contents

1. Introduction 2
2. Definitions and Acroyms 3
3. Data Collection Requirements for Digital Twin Network 3
   3.1. Target Driven and On-demand Collection 3
   3.2. Diverse Tools for Various Data 4
   3.3. Lightweight and Efficient Collection 5
   3.4. Open and Standardized Interfaces 5
   3.5. Naming for Caching 6
   3.6. Efficient Multi-Destination Delivery 6
4. An Efficient Data Collection Method for Digital Twin Network 6
   4.1. Overview 6
   4.2. Efficient Data Collection Mechanism 6
   4.3. Data Collection Process 8
5. Summary 9
6. Security Considerations 10
7. IANA Considerations 10
8. References 10
   8.1. Normative References 10
   8.2. Informative References 10
Authors' Addresses 10

Data Collection Requirements for DTN

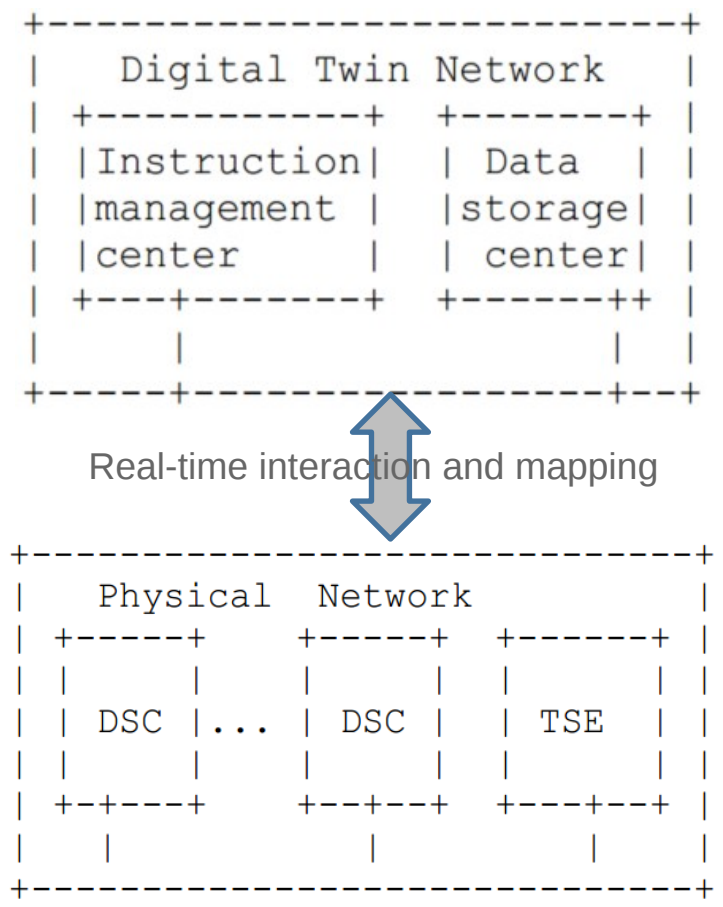
- **Target Driven and On-demand Collection**
  - **Complete** data is NOT affordable
  - Just collect data toward the target to build specific model for specific application
- **Diverse Tools for Various Data**
  - **Widely-used tools**, such as SNMP, NetConf, Telemetry, INT (In-band Network Telemetry), DPI (Deep Packet Inspection), etc.
  - New innovative directions
    - High-performance data collection technology based on programmable circuits
    - Measurement methods for complex network data such as network performance and network traffic
    - Collaborative data collection technology for multiple data sources and complex network data.
- **Lightweight and Efficient Collection**
  - To improve efficiency of execution, reduce the cost of computing, storage and communication bandwidth
  - Redundant data should be avoided or minimized
  - Data compression and aggregation technology are recommended

Data Collection Requirements for DTN - Cont.

- **Open and Standardized Interfaces**
  - Support configuration management, including the data collection protocol, frequency or period, etc.
  - several speed options (e.g. minute-level, 10-second level, second level (near real time), and real time level) to accommodate different data requirements from applications.
  - Be able to provide secure and reliable information exchange mechanism
- **Naming for Caching**
  - Give a unique identifier or "name" to each data or knowledge item for references
  - The name will be used by caching mechanisms to store the data and provide it for clients
- **Efficient Multi-Destination Delivery**
  - The telemetry system [RFC9232] can be used to deliver the requested data or knowledge items to the requesters.
  - Items will be provided by the closest cache to the destination of the data.
  - And, items will be replicated in the best nodes, following an efficient multi-cast spanning tree.

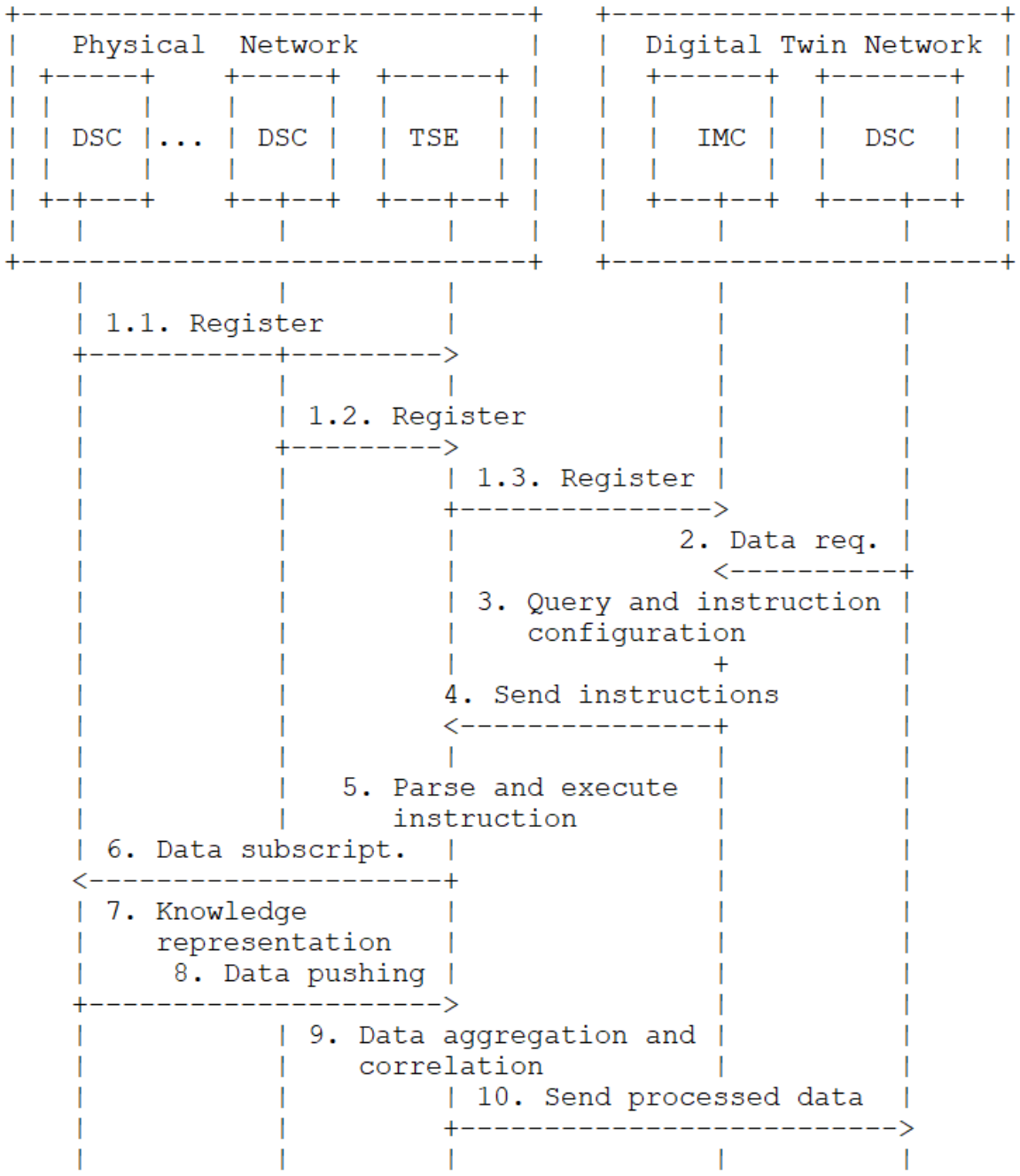
An Efficient Data collection method for DTN

- Current collection methods collect raw and full data from Physical Network, and have the problems:
  - Time-cost
  - Insufficient storage resources
  - Low computational efficiency
  - Waste of bandwidth resources caused by data transmission.
- This draft proposes an efficient and lightweight data collection, aggregation and correlation method.
  - Twin Network sends instructions to Physical Network to collect data on demand
  - Physical Network completes instructions such as knowledge representation
  - Telemetry Streaming Element (TSE) of Physical Network completes data aggregation and correlation.
  - Finally TSE sends the represented data to the Twin Network.



An Efficient Data collection method for DTN - Cont.

- Data collection Process
  1. Regisger: DSC in the PN side registers into the TSE. The TSE registers into the IMC.
  2. DSC in the DTN: send data Request
  3. IMC: query and config
  4. IMC: send Instructions
  5. TSE: Parse and excute instructions
  6. TSE: send data subscription
  7. DSC in PN: represents the data semantically in RDF form or sends the data in raw form to the TSE for it to make the semantic representation.
  8. DSC in PN: pushes the data or knowledge item to the TSE
  9. TSE: aggregates and correlates the collected data or knowledge.
  10. TSE: sends the processed data or knowledge to DSC in TN



Next Steps

- Further investigate and categorize the legacy data collection tools toward various data for building DTN.
  - Verify the data collection methods in a DTN demo system.
- Call for more efficient data collection methods suitable for DTN system to enrich the draft.

- **Looking forward to the comments, suggestions and questions**.