CSDT in Collaboration with a DT: A Solution for Disturbance Prediction In IoT Systems

Meriem Smati[1,2], Vincent Cheutet[1], Jannik Laval[1], Christophe Danjou[2]

[1]) INSA Lyon, Université Lumière Lyon 2, Université Claude Bernard Lyon 1, Université Jean Monnet Saint-Etienne, DISP UR4570, Villeurbanne, 69621, France
[2]) Laboratoire Poly-Industrie 4.0, Département de Mathématiques Et Génie Industriel, Polytechnique Montréal, Montréal, Québec, Canada
meriem.smati@insa-lyon.fr, vincent.cheutet@insa-lyon.fr, jannik.laval@univ-lyon2.fr, christophe.danjou@polymtl.ca

1 Context

With the evolution of Digital Twins (DTs) since its inception in 2002, there has been difficulty in achieving a consensus around the DT concept. Many research proposals have been put forth without clear objective evidence of scientific or technical advancement.

In light of this, an Internet of Things (IoT) system has been put in place as the Physical Twin (PT) that is constituted of two subsystems. The replica of the first subsystem is a DT with cognitive capabilities, making it a Cognitive Digital Twin (CDT). In addition to that, data disturbances have been injected to the CDT leading to the emergence of Cognitive Super-Digital Twin (CSDT). The replica of the second sub-system is a DT that predicts the physical root-cause problem in real-time.

Problematic: In this data-driven world and the increasing use of IoT devices, a central challenge reside in guaranteeing uninterrupted functionality despite potential disruptions. The focus of this paper is on the data anomalies, also known as outliers as well as foundational factor or root-cause of a a disturbance. Since IoT systems are susceptible to a range of vulnerabilities stemming from both software and hardware failures considered as obstacles and challenges that encompass the uneven distribution of data in IoT systems, with a prevalence of normal data that complicates anomaly detection; the variability in defining normal and abnormal behavior, which is affected by factors like season, location, or context; and the dependence on historical data in traditional DT, requiring a data anomaly to occur before making predictions in future cases.

Therefor the research objective if the answer the following question: **How can a DT be used as a tool for detecting root-cause problems and preventing data disturbance in an IoT system?**

2 Contribution

2.1 CSDT Proposal

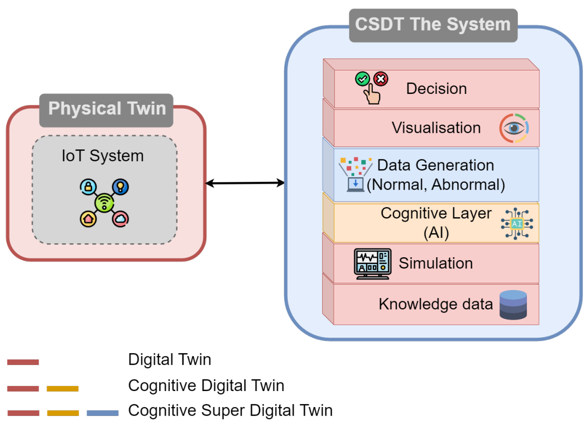

Figure 1: CSDT Layers

As it can be seen in Fig. 1, the chosen definition of a DT is a set of layers illustrated in pink: The first layer is the "database" which plays a crucial role in storing the vast amount of data required to create and maintain DTs. DTs can have multiple sources of information, such as information from processing the measured data and information collected previously from various data sources along the life cycle of a product. In this context our database is considered to be constituted of two repositories, the "dataset acquisition" repository that collect dynamic time-series data and the "Vault" repository that preprocess the collected data in the first repository and make the archive of the system. The layer that is positioned above of it uses the database layer to simulate and replicate exactly what the PT does. The visualisation layer in the DT is an important aspect since it allows users to see a virtual and real-time representation of systems or components. And depending on all these layers, the final layer of the DT is the Decision layer, helping the human being to make the decision. But to make predictions and make the decision-making more automated a cognitive layer is added represented in Orange, making the DT a CDT. Shifting focus on the contribution at hand which answers the problems mentioned: the data generation layer in Blue. This stratum makes a CDT a CSDT.

2.2 Generic Architecture Solution

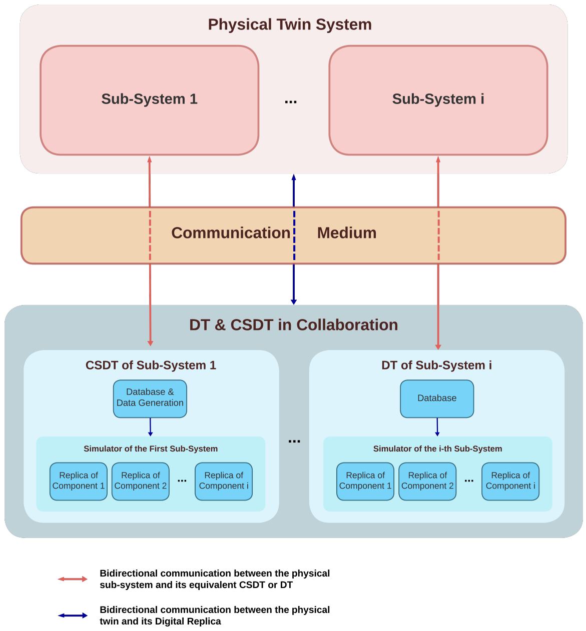

Figure 2: Generic Architecture

Fig. 2 represents the versatile architecture of our contribution. As any DT, the architecture must be constituted of the three essential parts: a Physical asset, a communication medium and the digital asset. In this paper, the PT is considered as an IoT system, composed of several sub-systems, each potentially featuring distinct IoT architectures. To facilitate the bidirectional communication between the two assets, an intermediary communication medium plays a pivotal role in establishing a seamless and real-time tunnel between the previously mentioned components. As it can be seen, everything in the architecture is considered a component, i.e a sensor is considered a component, a sub-system is considered a component, the physical twin is a component, the sole distinction lies in the magnitude and scale of the component. That is why the communication can be done between the two components physical and digital asset, as well as between the subsystem and its equivalent replica. Moving on to the digital part, it represents the communication between a DT and a CSDT, which means that we can choose the type of DT that we want, i.e either a DT, CDT or CSDT, in the figure we see a communication between a CSDT and a DT. The data generation is represented in the "Database & Data Generation" module. After the data generation, an operation of data acquisition will start, since the origin of the data processed by the CSDT can come from different sources (the generated data, the data collected from the IoT system, etc) so it serves to collect the data and process it to combine it and make it one global source of data or dataset. The over element of the CSDT is the simulator, which is the core of the DT which is a faithful replication of the PT's functionality. a DT can represent the whole system, on sub-system or many sub-systems. Each sub-system can communicate with its own replica directly and get bidirectional updates periodically.

3 Software and Hardware Implementation

A case study has been put in place to test the contribution and finding suitable Machine Learning (ML) or Deep Learning (DL) models for the physical and digital counterparts. This use case focuses on detecting room occupancy by gathering environmental data. The Physical Twin (PT) comprises two subsystems, the first of which utilizes an Edge/Fog IoT architecture featuring a Raspberry Pi 3 Model B and three sensors: Grove DHT11, Grove Light, and Grove SGP30. Data on temperature, humidity, CO2, and light levels are transmitted to the Fog for model training, augmented with an external dataset[1]. Predictions from this model, alongside collected data, are relayed to the second subsystem—a Poppy Ergo Jr robot—aiding decision-making. The Poppy Ergo Jr has its own DT which tells us if the robot arm is making the right movements, if not, it mentions the root cause. On both subsystems, the data is stored in InfluxDB, a time-series database, via the Communication Medium. Once received, the CSDT preprocesses the data, amalgamating DT-generated, external, and collected data to form a comprehensive dataset. The detailed architecture is available along with a demonstration[2].

[1] https://archive.ics.uci.edu/dataset/357/occupancy+detection
[2] https://www.intranet.disp-lab.fr/s/arbqpMqesnBtACp