INTRODUCTION TO CPS

Cyber Physical Systems are multidisciplinary systems that conduct feedback control on widely distributed embedded computing systems through combination of communication, computation and control technologies. Modern CPS are able to realize the real-time, dynamic, safe and reliable collaboration with physical systems represented via embedded system. They are Integral mixture of existing network systems and traditional embedded systems. Where, Physical system data modules collect data by distributed field devices in CPS system, then pass data to the information processing layer as per the complete given tasks and demands of services by information processing technologies such as statistical signal processing, feedback control, data security processing and data uncertainty management. CPS interact with physical system through networks, the end system of CPS is normally traditional centralized tightly coupled embedded computing system, which contains a large number of physical systems composed of intelligent wired/wireless actuators & sensors.

The potential benefits of the convergence of 3C technologies for developing next-generation engineered systems that can be called Cyber Physical Systems are wide ranging and highly transformative via efficient computation, distributed sensing, high-level decision-making algorithms, control over wireless / wired communication networks formal verification technologies and multi-objective optimization; engineered cyber physical systems are in many societal critical domains such as construction, energy, transportation, and medical systems. Scientists and Engineers in this field have deep understanding of system and branches of mechatronics, biology, computer science and chemistry. Physical systems & Technical systems are developed and designed to be more & more reliable, efficient, smart, robust and secure.

With such high notions, the scope of CPS and integration of Cloud computing is about to bring the next big Industrial Revolution, Industry 4.0. The complete factory can be made into digital twin in the cyber space. Any changes in the cyber twin will reflect in the physical world which could boost the production & efficiency in the field of Avionics, Distributed robotics, Energy conservation, Process control, Smart structures, Defense systems, Critical infrastructure control, Assisted living, Environmental control, medical systems, Manufacturing and Traffic safety & control

INDUSTRY 4.0

The term ‘Industry 4.0 stands for the fourth industrial revolution, the next stage in the organization and control of the entire value stream along the lifecycle of a product. This cycle is based on increasingly individualized customer wishes and ranges from the idea, the order, development, production, and

delivery to the end customer through to recycling and related services. Fundamental here is the availability of all relevant information in real-time through the networking of all instances involved in value creation as well as the ability to derive the best possible value stream from data at all times. Connecting people, objects and systems leads to the creation of dynamic, self-organized, cross-organizational, real-time optimized value networks, which can be optimized according to a range of criteria such as costs, availability and consumption of resources." Hermann et al. defined Industry 4.0 as "a collective term for technologies and concepts of value chain organization.

Within the modular structured smart factories of Industry 4.0, CPS monitor physical processes, create a virtual copy of the physical world and make decentralized decisions. Over the IoT, CPS communicate and cooperate with each other and humans in realtime. Via the IoS, both internal and cross-organizational services are offered and utilized by participants of the value chain". Both of the definitions above deem Industry 4.0 as the next stage of value chain organization and management. Industry 4.0 is characterized by the integration along three dimensions: vertical integration together with networked manufacturing systems, horizontal integration through value networks, and end-to-end digital integration of engineering across the value chain of a product's life-cycle. Smart factory is a core concept component as well as a key feature of Industry 4.0, which is where the vertical integration takes place. Horizontal integration refers to the integration of multiple smart factories through value networks, occurring both within a smart factory and across different smart factories. Vertical and horizontal integration enables the end-to-end integration across the entire value chain. Smart product is another critical concept component in Industry 4.0's concept system. In a smart factory, products and machines communicate with each other, cooperatively driving production. Smart products can refer to objects, devices, and machines that are equipped with sensors, controlled by software and connected to the Internet. Industry 4.0 will give rise to novel CPS platforms geared toward supporting collaborative industrial business processes and the associated business networks. CPS platforms are where the specific requirements for horizontal and vertical integration of CPS, applications, and services arise in business processes.

It should be noted that Industry 4.0 stands for the fourth industrial revolution, which necessitates the consideration of many other issues that may occur in the upcoming new era, including standardization, safety and security, resource efficiency, new social infrastructure, work organization and work design, training, regulatory framework, etc. Technologies of nine aspects that power the transformation of the current industrial production to that of Industry 4.0 have been

identified, which more or less have something to do with CPS. CPS provide a critical support to the vertical and horizontal system integration. The combination of CPS and the industrial IoT enables the creation of the IoT and IoS. CPS will surely bring about the cybersecurity issue. Moreover, the widespread application of CPS means the generation of industrial big data, which requires cloud technology and big data analytics for storage and analysis. The virtual world of CPS consists of a great variety of models of the production facilities, for which simulation can play an important role. Augment

reality technology is required for operators to interact with CPS. Additive manufacturing and robots are essential parts of the CPS-based manufacturing systems of Industry 4.0. Within a smart factory, all the physical production elements in the physical world have a cyber twin in the virtual world. The physical and virtual worlds as well as the physical assets and cyber twins in them are seamlessly connected to achieve the global optimization of production within a smart factory. Moreover, within a value network, multiple factories are horizontally integrated, i.e., the physical assets and the cyber twins are, respectively, integrated to enable optimized decision-making across the value network. The integration through the value network will give rise to CPS platforms, within which things, services, "data," and "people" are connected over the Internet.

CLOUD MANUFACTURING

The term "cloud manufacturing" and defined it as "a new networked manufacturing paradigm that organizes manufacturing resources over networks (manufacturing clouds) according to consumers' needs and requirements to provide a variety of on demand manufacturing services via networks and cloud manufacturing service platforms". "a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable manufacturing resources (e.g., manufacturing software tools, manufacturing equipment, and manufacturing capabilities) that can be rapidly provisioned and released with minimal management effort or service provider interaction". In the cloud manufacturing mode, providers supply their manufacturing resources, which will be transformed into services and then pooled into the cloud manufacturing platform. In line with the market economy theory, operators are introduced to manage the platform so that high-quality services can be guaranteed and provided. Customers can submit their requirements to the platform for requesting services ranging from product design, manufacturing, testing, management, and all other stages of a product life-cycle. The core of cloud manufacturing lies in the establishment of the cloud manufacturing platform, which relies on many technologies. A cloud manufacturing platform has a multilayer architecture, including resource layer, virtual resource layer, global service layer, application layer, and interface layer.

The implementation of different layers requires different technologies. IoT, virtualization, and servitization technologies are needed to sense manufacturing resources and transform physical resources into virtual resources in the virtual resource layer. The core technologies for global service layer are cloud computing, service-related technologies (including service-oriented technologies and service management-related technologies), and semantic Web technology. In the interface, it is the human–machine interaction technology that plays an important role. Certainly, the high-performance computing technology and advanced manufacturing model and technology are also essential. Many other supporting technologies such as big data are also necessary for the complete implementation of a cloud manufacturing platform

Industry 4.0, has been compared with other manufacturing concepts such as “Advanced Manufacturing” or “Lean Production”. This multifaceted term comprises of various interdisciplinary concepts without a clear distinction. Distinguished by its huge technological potential, comparable to technical innovations which led to the industrial revolutions: the field of mechanization, the use of electricity, and the beginning of digitization.

Smart factories operating according to Industry 4.0, on the basis of collaborative CPS, represent a future form of industrial networks. Manufacturing systems can be transformed into digital ecosystems via Industry 4.0. One of the paradigms of Industry 4.0 is the Smart Product. From among these, the guiding idea of the Smart Product is to extend the role of a common work element to one which becomes an active part of the system. The product acquires a memory in which operational data and requirements are stored directly as an individual building plan. In this way, the product itself requests the required resources and orchestrates the production processes. This is a prerequisite to enable self-configuring processes in highly modular production systems. Industry 4.0 is still somewhat in its initial stages of development, especially related to decision-making systems and smart solutions

Industry 4.0 is named after the industry revolution, while cloud manufacturing follows from the advanced manufacturing models and technologies. Therefore, Industry 4.0 needs to be able to describe the landscape of the manufacturing industry in the upcoming era and present solutions to issues that need to be dealt with (such as the resource and energy efficiency, urban production, and demographic change). Industry 4.0’s CPS-based manufacturing systems are capable of providing effective means for solving the problems. Cloud manufacturing is an advanced manufacturing business model that focuses on issues that are directly related to manufacturing (i.e., resource sharing and collaboration in the cloud) and pays less attention to issues like urban production and demographic change, etc. In terms of concept system, Industry 4.0 encompasses both vertical integration and horizontal integration, but cloud manufacturing concentrates on the integration in the cloud manufacturing platform in the form of service composition (integration in the cloud corresponds to the horizontal integration concept in Industry 4.0). At the factory end, cloud manufacturing is more concerned with how to connect the manufacturing resources into the cloud manufacturing platform and pays less attention to the issue of internal organization and operation (e.g., vertical integration) within a cloud manufacturing factory. But this does not mean that the implementation of the cloud manufacturing factory is unimportant. Exploring the requirements of the cloud manufacturing factory and building it are important future research issues in the area of cloud manufacturing. In this aspect, the CPS-based smart factory of Industry 4.0 can provide important reference. Based on the analysis above, we can conclude that Industry 4.0 is a broader concept system than cloud manufacturing in terms of the issues involved and the completeness of the concept systems.

Industry 4.0's fundamental idea is to integrate manufacturing systems of different smart factories along a value chain (or a value network) in the form of CPS so that real-time data and information across the entire value chain can be obtained, which enables real-time and accurate decision-making. Industry 4.0's CPS-based manufacturing systems have high flexibility, adaptiveness, realtime capability, and can achieve the transparency of production processes. Industry 4.0 is therefore capable of producing increasingly individualized products (of even batch size one) with higher quality, lower costs, and high productivity, etc.

Cloud manufacturing's fundamental idea is to connect and integrate manufacturing resources of different factories (or enterprises) into the cloud so that large-scale resource sharing and collaboration can be realized in the form of services and their composition. Cloud manufacturing's cloud-based resource sharing method can also bring enormous benefits and advantages to enterprises, such as financial flexibility, business agility, and instant access to innovation. Both Industry 4.0 and cloud manufacturing converge to better meeting increasingly customer's individualized requirements. In fact, Industry 4.0 represents a highly digital and networked manufacturing paradigm that satisfies enterprises' digital manufacturing and collaboration requirements between them and business partners, while cloud manufacturing can effectively satisfy the sharing and collaboration requirements of enterprises in a convenient and agile way. Overall, the technologies for Industry 4.0 have something to do with CPS, while the technologies for cloud manufacturing are mainly for the implementation of a cloud manufacturing platform. As far as the inter-factory integration is concerned, Industry 4.0 relies on CPS platforms, and cloud manufacturing relies on cloud manufacturing platforms. Process and Digital Twin Modeling Solutions. With modern developments of communication protocols and connection architectures, numerous initiatives have attempted to model physical machines, processes, and results with computational methods.

These "Digital Twins" of the physical world have been developed to different extents, each modeling and predicting different areas of the production process. The concept of Digital Twin modeling is defined as "virtual machine tools of physical machines for cyber-physical manufacturing by using sensory data and information fusion integration techniques". Digital Twin initiatives have been implemented to address a wide variety of modeling scenarios. Simple solutions range from open-loop monitoring of process data while more complex implementations include AI-powered feedback to correct errors in a manufacturing process.

Many Digital Twin methods integrate models of physical systems with information and result from both computer simulations and CPS connected data. These two sources of information are combined to provide a more accurate method of predicting the system's behavior and production results. The concept by summarizing the technology needed for a complete Digital Twin to predict microstructure

development, residual stresses, and part defects in additive manufacturing. Through the integration of temperature and velocity fields from both numerical simulation and experimental measurements.

Digital Twin model for directed energy deposition in additive manufacturing provides a more accurate cooling rate and temperature gradient prediction than traditional conduction calculations. Looking to the future of Digital Twin modeling where complete factories of CPSs can be connected and monitored provides exciting opportunities for process development. With more advanced capabilities to link and correlate manufacturing data from different sources, demonstrate a future application for Digital Twin modeling in partial and parallel disassembly sequence planning for products. Implementation of the near real-time analysis for product information, timing information, and upstream downstream events would provide beneficial flexibility and adaptability for assembly planning methods. These two types of platforms are different in the aim, core technologies, operational mode, and platform architecture.

Aim: Cloud manufacturing platforms aim to support full sharing and efficient collaboration of social manufacturing resources through centralized management and operation. Cloud manufacturing platforms are open that allows enterprises to freely join or exit. CPS platforms also aim to provide support for collaborative industrial business processes and the associated business networks for all the aspects of smart factories and smart product life-cycle. Core technologies: For a CPS platform, it is certain that CPS and IoT are core technologies. Depending on business objectives of CPS platforms, other technologies may also be needed. While for cloud manufacturing platforms, the core technologies include IoT, virtualization and servitization, cloud computing, service-related technologies, etc.

Operational mode and business model: The operational mode of the CPS platform in Industry 4.0 is an open problem and there is little discussion about it. By contrast, the operational mode of a cloud manufacturing platform has been explicitly defined. Nevertheless, there are some common issues to be solved regarding these two types of platforms in terms of business model, such as dynamic pricing, fair benefit sharing, broader regulatory requirements, intellectual property and know-how protection, monitoring of business processes, legal issues, as well as safety and security issues. Architecture: To date, there has been no specific discussion with respect to the architecture of the CPS platform, whereas the cloud manufacturing platform adopts layered serviceoriented architecture, on which researchers have largely reached a consensus.

The common parts between CPS platforms and cloud manufacturing platforms lie in the implementations of IoT, IoS, Internet of devices (IoD), and Internet of platforms (IoP), (or the Internet of users (IoU)). Moreover, service is an important concept for both Industry 4.0 and cloud manufacturing. There are some differences in the meaning of services in these two concepts.

In Industry 4.0, services are closely related to CPS. The most frequently mentioned service concept in Industry 4.0 is productrelated services (such as condition monitoring and preventive and predictive maintenance). Cloud manufacturing embraces the concept of manufacturing-as-a-service, taking everything (including manufacturing resources and processes encompassed in the entire product life-cycle) as services. This is the broadest service concept (including the product- and process-related services in Industry 4.0). Hence, the scope and connotation of services in cloud manufacturing are much broader than that in Industry 4.0. Although much research has been done on both Industry 4.0 and cloud manufacturing, it is still in the very early stage. Existing research on Industry 4.0 is more about CPS, smart factories, big data, etc. Although inter-factory networking and integration (i.e., horizontal and end-to-end integration) are also key constituents for the concept of Industry 4.0, the related studies are actually scarce. In this regard, only a couple of works discussed the issue of inter-factory integration. In contrast, cloud manufacturing research mainly focuses on how to implement a cloud manufacturing platform and its associated technologies, among which the large body of work addressed resource and service-related issues, such as resource classification, perception and connection, virtualization, and servitization.

SIMULATED PRODUCTION SYSTEM

In a first step, a simulation is used to acquire production data of a virtual factory. All of the factory's components like the machine tools and the product workpieces are virtual as well. Each machine has a machine type and each product has a product type, so there are different types of virtual products and machines. The work plan, i.e., the order of operations required to produce a final product, is given externally and cannot be changed. Thus, the product type defines which operations need to be processed sequentially to finish the product, while the machine type defines the operation the machine is capable of The ability of machines of a certain type to perform an operation with specific requirements is encoded into the machine type. For example, if the machines of a certain type are able to drill holes, but their accuracy cannot be guaranteed to be high enough for a certain operation, their machine type marks this operation as not performable. Though, other machines might be able to perform this operation with the required accuracy. Thus, the technological capabilities are encoded into the machine types. The material removal rate (MRR) may vary during operations depending on material type, cutting speed and depth, cutting aids, tool type, or other factors.

This leads to different process times, even for operations with the same production technology. The required setup times for each operation are included in the resulting processing times that are given in Table 1. To finish the production of a product, all of its operations need to be processed in order, while each operation takes a certain time. Since in the presented example no machine is capable of performing all operations, the products have to be processed on different machines sequentially. So, the current operation of a product is first finished on one machine, and then the product is transported to another machine. Since this new machine might be busy, the product is enqueued. To do so, each machine has

a queue of waiting products that are processed in order of arrival (first in–first out). If there are different machines of the same type, the question arises, which of these machines should process a certain product. This question cannot be answered in a perfectly optimal way for real-life sized problems due to its high-computational complexity.

So, an optimal solution cannot be calculated in a feasible time, but heuristics can be evaluated efficiently to come close to an optimal solution of product distributions. This work uses the heuristic of always choosing the machine that will have processed the product's individual operation first. Other heuristics could be considered as well, but since the choice of heuristic is not important for the demonstration of the presented analysis tool, the described heuristic is chosen out of simplicity reasons. Another issue is the optimal arrangement of machines in the factory. This problem is also computationally very expensive and cannot be solved optimally in a feasible time for a larger number of machines. Therefore, the arrangement of machines in the presented example was chosen as demonstrated by simply distributing groups of identical machines within the factory.

The transportation times of products between the machines depend highly on the arrangement of those machines. Since the production batches in the used example are very large, the resulting transportation times are very small in comparison. Therefore, the transportation times are visually disappearing in this example. Still, the methodologies that are presented are easily extendable in a straightforward way to also visually include transportation times, as will be seen shortly. In the presented example, a free transportation model is used. Naturally, other simulations could use restricted transportation routes to implement conveyor belts or other transportation methods. The production data used for the analysis describe which product and which operation are performed on which machine at which point in time. To acquire this data, each product type is virtually produced 30 times in a simulation, while starting with a product of type A, then B and C, and then repeating this loop 30 times with a temporal gap of 10 min in between the products. This means, a new product of a specific type starts its virtual manufacturing every 30 min. To analyse the gathered production data, the visualizations and the methodologies described in the following are used.

INDUSTRIAL DATA ANALYTICS

Industrial data analytics play an essential role in achieving the smart factory vision and improving decision-making in various industrial applications. Five main industrial data methodologies are generally studied including highly distributed data ingestion, data repository, large-scale data management, data analytics, and data governance. Industrial data processing offers valuable information about various sections of industrial applications including inefficiencies in industrial processes, costly failures and down-times, and effective maintenance decisions. The industrial data analytics are generally deployed for improving factory operations through improving machinery utilization and predicting production demands, improving product quality by analysing market demands and reducing defective products, and enhancing supply chain efficiency by analysing risk factors and

making accurate logistic plans and schedules. The methods of industrial data analytics can be split into different categories such as descriptive, diagnostic, predictive, and prescriptive analytics. Descriptive and diagnostic analytics are responsible for analysing historic data and the causes of events and behaviours. Predictive and prescriptive analytics require more processing power, anticipate the trends of data, and deploy the historical data in making decisions to achieve production goals. Examples of industrial data analytics frameworks can be found. A platform for performing industrial big data analysis is presented where the performance requirements are introduced to achieve a cost-effective operation.

A manufacturing big data solution for active preventive maintenance in manufacturing environments is proposed. Various other frameworks for industrial data analysis can be found, where the importance of using data analysis in decision-making is emphasized. The data available in a cyber-physical production system can be used to make production systems more flexible. In this context, flexibility can be understood, on the one hand, as the transformability of the system to engineering changes on medium- or long-term perspective. On the other hand, flexibility can be understood as achieved by decentralized production control on a short-term view. It is obvious that the vast amount of data is not useable without refinement. Therefore, user friendly tools for data analysis and visualization are needed. Examples for engineering changes are the reconfiguration, addition, substitution, or removal of production equipment, e.g., machine tools, in a manufacturing system. They usually have extensive impacts on the manufacturing system due to the manifold interrelationships among production objects and hence need careful analysis and planning before implementation.

The change in one element might result in the disruption of the process chains, material flow, or information flow. Therefore, tools are required that can analyse the effects of envisaged engineering changes in a fast and comprehensive manner. Tools that support the planning and analysis of changes in manufacturing systems can be found within the concept of the digital factory. Simulation and evaluation software for products and material flows can be applied in order to analyse processes and their changes. However, such software tools require specific know-how and qualified personnel to use them and to keep them up to date, and are not specialized on engineering changes. A framework specialized for analysing impacts of engineering changes to existing manufacturing systems is proposed. Here, the alternative solutions for engineering changes are visualized in a 3D virtual environment where effects on factory layout and material flow can be seen in a spatial context. Although a three-dimensional virtual environment displays information intuitively and thus gives a realistic feeling of the modelled factory, it shows only partial views of the factory and does not guide the user to the information needed. For the fast and effective analysis of impacts to engineering changes, both, aggregated and detailed views, are necessary.

To enable the overall evaluation of the given situation, e.g., to examine process chain and information flow consistency, the available data need to be visualized in an aggregated manner. On the other hand, scalability of the data is required to allow the user to focus on single products or machines and to well

defined time steps of special interest. An essential requirement is to guide the user to the most interesting features of the regarded system and to show critical issues. Therefore, comparative and interactive data highlighting integrated in the spatial context of the factory are needed. Different perspectives focusing on machines, products, and material flows within the visualization tool need to be distinguished and interlinked in an interactive manner. In contrast to the planning of engineering changes, decisions in production control need to be taken in real-time with limited information. The concept of self-control in a decentralized production system is based on the ability of several elements of the system (e.g., machine tools or work pieces) to act and decide autonomously. In contrast to that, in the centralized approach, planning is accomplished by a superordinated planning entity. Therefore, especially for decentralized production control with a multitude of decision-makers, a fast recognition of data patterns is necessary to adapt the behaviour and decision rules of the acting elements. The applicability of different self-organization concepts is tested in several research projects by using prototype factories. As the implementation of such prototypes with real machinery involves considerable effort and expenses, they are therefore not meant for real scale experiments. Thus, the amount and complexity of data can still be managed manually, so tools for visual evaluation and optimization of the concepts are missing. propose a test field based on a multi-agent system to test several self-organization concepts against each other in a real sized but virtual environment.

Here, several different decision routines, e.g., for machine tool selection or production order, are possible. This case shows that in the analysis of decentrally controlled production data, the impacts of different decision routines need to be visualized. Further, there is a need to identify patterns on an aggregated data level to derive the system's sensitivity to changes of decision routines. As a consequence, aggregated views displaying the overall performance in a spatial context, and detailed views representing the perspective of single elements, are necessary to understand an entire system. To summarize, one major issue for data visualization is to be intuitively understandable. Therefore, an interactive guidance for the user is required, which makes it easy to find interesting features in a data set. To get a quick but comprehensive overview of the status of the production system, different perspectives on an aggregated level are needed. These have to be interlinked to navigate through the perspectives. Beside the aggregated views, scalability is a further required functionality that enables to select single hotspots and establish detailed comparisons between machines, products, or time steps. Embedding the data into the spatial context of the factory is needed to give the user a realistic and intuitive understanding of the factory and its performance.

VISUALIZATION

A tool for the user-guided visual analysis of simulated production data was designed as described below. The tool is a linked view system, visualizing the manufacturing process under different aspects. This means that there are different views, each showing the same data but having a focus on different aspects. The presented tool contains a flow view, a workload view, and a production view. Additionally, the

views are interlinked by transferring user interactions like selection and highlighting of products, product types, or machines from one view to all views. Another user interaction is to choose a time window by manipulating a point in time and an interval size in all views. Then, this time window will be considered for visualization. This enables the user to zoom in and out onto certain interesting points in time. All of the views of the system only show the data that occur during this chosen time window, thereby treating this window consistently for all views. By doing so, the user is enabled to focus on certain features, while the overall picture is preserved. This helps the user in building a mental map of the production data. Since all views of the presented system always show data for the same time window or selection, a cognitive transition from one view to the others is straightforward. After the virtual manufacturing system is simulated once, the whole tool and its views work in real-time to provide flexibility of interactions to the user. The presented tool can be used to analyze virtual factories, provide user guidance for later optimization or comparison, and help in decision making.

The overall goal is the optimization of the production process with respect to a diversity of parameters. Still, this optimization cannot be done fully automatically because of its highcomputational complexity. This stresses the importance of the presented tool to support users in their analysis tasks. Although the optimal solution is unknown to users, the presented tool can be used to iteratively improve factory settings. By that, users are enabled to approximate an optimal solution, thereby finding a sufficient solution and gain a certain confidence in their production process.

GRAPH DATABASE APPROACH

The advantages of deploying GDB include having a more natural approach of data modeling and keeping data properties connected to nodes and relationships. Moreover, GDBs offer graphical and visualization interfaces to data and are able to keep the time-related information of events through various graph paths. Also, an extended list of applications and implementations of GDBs is presented in to show their use on enterprise data, social networks, and determining security and access rights. It was found that GDBs provide the much-needed structure for storing the data and incorporating a dynamic data model. In general, the use cases, in which GDBs perfectly improve the data management, include path finding with weighted and time-related path properties, mapping dependencies of various system components to capture potential weak points, and communications between various networked elements. Conversely, query languages are used to extract data, including traversing the database, comparing node properties, and subgraph matching. The performance of different GDB tools and methodologies are analyzed and compared.

Various aspects of functionality differentiate the performance of query languages such as subgraph matching, finding nodes connected by paths, comparing and returning paths, aggregation, node creation, and approximate matching and ranking.

Graph Database for Industrial Data Analysis, Due to their advantages including scalability, efficiency, and flexibility, GDBs are widely adopted in various industry-related applications and use cases such as

network operations, fraud detection, and asset and data management. The authors have proposed a new object tracking approach for surveillance applications. The GDB approach is selected to contribute to the scalability of the proposed scheme and support the required connectivity analysis for the object tracking. Moreover, relationships in social networks have been modeled using a GDB for structural information mining and marketing. Conversely, GDBs are also deployed in business solutions for scenarios with multiple large data sources, which require distributed processing in decision-making for various problems such as fraud detection, trend prediction, and product recommendation.

GDB and Neo4j can be used in the network security-related applications because the network characteristics are in compliance of the GDB concept of nodes and relationships. It was stated that Neo4j is selected to efficiently query and analyze the data where the query results can be visualized directly. Neo4j was also used to build a model for a power grid network analysis were experimental results compared the performance with an example relational database system. An efficient and secure information retrieval framework for content centric networks used the Neo4j graph database to improve the efficiency of storing and processing large-scale data. Neo4j was used to analyze the network vulnerability to guarantee the accuracy of the attack graph generation and analysis process.

Moreover, the use of GDB, and more specifically Neo4j, in analyzing time stamped data logs has been demonstrated. A GDB approach has been used for analyzing the network log files from different sources in real time. The data from different network layers have been exported and combined in a single graph to detect anomalies in network performance.

The business event logs monitoring is demonstrated where a loan application was exported to a GDB to facilitate the business decision-making process based on the available data. In this study, we introduce the application of a GDB approach for analyzing industrial CPS, which achieves the one-to-one mapping between the network activities and the corresponding physical actions.

Machine Tending

A GDB was built to manage data collected from testbed measurements of both network traffic and physical operations. In this section, we briefly introduce graph components developed for our testbed and the data processing flow that transforms measurement results to graph entities. To justify using the proposed approach, we start by stating and defining the collected data characteristics and the requirements for the deployed database approach in handling the data for the goal of our study as follows:

Heterogeneous data: The collected data from industrial wireless communications system is heterogeneous in different aspects as follows:

- Different sources: We collect network data at various network nodes in the system. Also, collected data using wireless sniffer describe the wireless physical environment. Data from the supervisor controller are also collected, which include the system states and the supervisory commands. Data from the robots are used to describe the physical actions taken.

- Different formats: The data include different file formats such as packet capture (PCAP) files, and data that come from different PLC and robot controllers are stored in the format of comma separated value (CSV) files. Another example is the time stamp format from different devices.
- Different rates: Data packets can be both periodic and event driven. Also, the robot state feedback is periodic with a different update rate than the update rate of the PLC state.

Entities are interrelated: This is the main requirement and challenge in this work where the goal is to obtain the direct one-to-one connection between physical actions and their corresponding entities including network activities, the physical wireless environment through sniffer reports, and the physical system state.

Various entity types: The data model will consider two types of system entities, namely, dynamic and static. The class of static entities covers testbed setup profiles, which contain testbed components, network interfaces, and their settings. These entities are normally predetermined or collected in the initialization of each measurement. The class of dynamic entities captures various system events such as machine status reports, network traffic, and information flows in the testbed. These entities are dynamically added into the data set whose quantities and properties are determined by the measured data.

Data model with multiple abstraction mechanisms: The considered data model and the corresponding queries should encompass multiple levels of abstraction including traffic data level, physical hardware level, physical environment level, physical actions level, and the interactions between these various levels. The network database system must allow for the categorization and labeling through these levels.

Time travel queries: The data model and the resulting database should allow for direct querying for temporal variations of the studied entities. Hence, temporal relationships between data packets and the corresponding physical actions should be stored and directly accessible.

Efficient path and relationship queries: Given the requirement of having interrelated nodes, the query language should allow for path and relationship queries to directly extract this information. These types of queries are used for calculating various system metrics and hence should be performed in an efficient manner

ML

////To Be Filled////

TYPES OF SECURITY MECHANISMS

A product realization process (PRP) involves assets such as software, physical systems, and humans, which are interconnected to each other. In a PRP, it is possible that the security of a noninformation

based asset such as a physical system could be compromised while maintaining the security of information-based assets. Such compromises can result in victims beyond simply the asset owner. Hence, the security challenges in a PRP are much stricter than in traditional information security.

A sPRP provides security to information-based assets, noninformation-based assets, and the network connections among them. The network connections between the assets are based on how the different stakeholders involved in the PRP contribute to product development. The security mechanisms deployed in the sPRP can be classified into the following four categories: assetcentric security, network-centric security, data-centric security, and user-centric security. Asset-centric security refers to the protection of hardware and software associated with the asset.

Network-centric security covers security related to communication channels between two or more assets. Data-centric security refers to the protection of data throughout its lifetime (creation, transmission, storage, updating) in a PRP. Clearly, the lifetime of protection in data-centric security is longer than network-centric security and this depends on the design stage in a sPRP. Usercentric security refers to the different security requirements from the stakeholders in a PRP.

The major security objectives in a sPRP are as follows:

- Confidentiality refers to the protection of the asset from unauthorized stakeholders.
- Integrity refers to the protection of the asset from unauthorized modifications.
- Availability requires the asset to be accessible, operational when it is needed.
- Accountability refers to the identification of the responsible entities for the various actions. In this section, we briefly discuss the common security mechanisms in each category and identify the security objectives achieved by each.

ASSERT-CENTRIC SECURITY

Asset-centric security subsumes all aspects that relate directly to the security of a device or asset in a PRP. The security mechanisms in this category are further classified into hardware security and software security.

Hardware Security focuses on establishing the validity of physical components, for example for traceability, counterfeit detection, or liability reasons, typically through leveraging randomness that is extrinsic or intrinsic to the part. Hardware security methods may be extrinsic, such as imprinting or adding an embossing or another marker to a part; or intrinsic, leveraging the inherent randomness of the part as in many physically unclonable function (PUF) designs. Other security concerns involve sabotage or tampering detection/prevention methods, which must be designed into a part.

- Physical Marking/Watermarking: This technique refers to marking a product for future reference in verifying point of origin, authenticity, and counterfeiting detection. This may involve stamping/embossing, barcode or radio-frequency identification (RFID) tagging, or the use of physically unclonable functions (PUFs). Watermarks may be fragile (designed to “break”

if tampered with or altered) or robust (designed to remain detectable in the face of some tampering or damage), depending on the use case.

- Crypto-processors: This is a processor packaged with multiple physical security mechanisms that are aimed at protecting it and its contents from attacks involving physical tampering. Such processors have been used in payment devices, trusted computing applications, and military applications.

Software Security solutions are aimed at protecting software against attacks such as tampering or causing run-time misbehavior through, e.g.,) invalid inputs that the software erroneously treats as legitimate. The latter exploit the existing software bugs and vulnerabilities which are inadvertently introduced during the software development cycle, whereas the former (tampering) take the form of unauthorized modifications to software that initially did not contain such vulnerabilities. The following are some of the widely used software security solutions.

- Antivirus Software: Antivirus software detects malevolent functionality (malware) by running the program in a virtual environment. Depending on the actions logged, the antivirus engine determines the existence of a malware. This is a computationally intensive approach. Recently, researchers proposed a machine learning based approach to detect malware. These approaches are based on the software's observed behavior.
- Tamper-resistant software: Software can be protected from tampering by increasing the difficulty of reverse engineering and/or modifying it, where attackers either add to it unauthorized functionality or disable functionality that the legitimate software publisher deems essential. Recent research in this area is aiming at not only detecting tampering but also correcting its effects.
- Digital watermarking: Digital watermarking is similar to physical watermarking in terms of achievable security features (Fragile/Robust). However, they are prone to different forms of attacks. Attacks can aim to remove watermarks, to transform watermarks to make them useless, or to determine the secret random pattern used in embedding the watermark so as to apply the watermark to illegitimate (counterfeit) media.

NETWORK-CENTRIC SECURITY

Network-Centric Security aims at providing secure communication among different assets in the network. These approaches are particularly important given the rise of cloud-based design and manufacturing. The common security mechanisms in this category are discussed as follows.

HTTPS is an extension of HTTP protocol with messages encrypted at transport layer security or secure socket layer. This implementation is mainly used for data integrity and confidentiality.

Secure Multicasting enables a designer to securely duplicate their information to the authorized receivers within a group of designers. This is a widely accepted and well-established technique, and we omit exhaustively listing and detailing its applications.

Virtual private networks enable network members to communicate over a public network such as the Internet as if the members were directly connected to a private network. This too is a well-understood and established technique, and we also omit listing and detailing its applications.

Blockchains may be used to validate transactions in a way that is difficult to modify after the fact. Blockchains link users through their hash-based identities. This can also be used for achieving accountability.

Intrusion detection systems are used to detect policy violations or malicious activity in a system or network. They do so by using two mechanisms, looking for evidence of specific known intrusions using a database describing the telltale signs ("signatures") of such known intrusions and looking for anomalies such as unusual sequences of events or deviations from expected traffic or usage patterns. Note that, in the signature-based approach, the database of attack signatures needs updating as new attacks are discovered, whereas no such need exists in the anomaly-based approach and that can detect new and previously unknown forms of attack. On the other hand, the anomaly-based approach can generate false alarms, whereas the signature-based approach generates none. Researchers have extended these to intrusion prevention systems as well.

Onion routing is a technique where the private data is wrapped with successive layers of encryption, similar to an onion (one layer for each node on the path from the source to the destination, so a node knows only its predecessor and successor and no node can correlate the path's source and its destination). This technique is used for anonymous communications.

DATA-CENTRIC SECURITY

Unlike network-centric security, the main aim of data-centric security solutions is to protect data throughout its lifetime starting from its creation, transmission, storage, and updating. Here, we briefly discuss a few popular approaches.

Steganography consists of hiding a secret message in a digital or physical object, such that it is not feasible for an adversary to determine, from examining the object, whether it contains such a message. Note that, in order to win, it is not necessary for the adversary to learn the secret message: determining that it exists is enough.

Encryption consists of encoding data for storing and transmitting it securely, such that only authorized parties can decode it (i.e., decrypt it). There are two broad categories of encryption techniques:

- Symmetric Encryption: The same key is used to encrypt and decrypt data. Rijndael's encryption algorithm is being used in the advanced encryption standard selected by National Institute of Standards and Technology (NIST). Diffie and Hellman's key exchange protocol can be used to exchange the keys required for symmetric encryption. Currently, commercially available CAD software such as AutoCAD-2016 and Onshape use AES-256.

- Asymmetric encryption: Different keys are used for encrypting and decrypting the data, neither of which can be inferred from the other in a practical amount of time. The encryption key is referred to as a public key and the decryption key is referred to as a private key. Commonly used asymmetric encryption schemes are RSA, Elgamal, and elliptic curve cryptography [40,41]. AutoCAD 2016 uses RSA-2048 bits to protect drawings.
- Secret sharing: This is also a cryptography-based approach. In (k, n) schemes, a secret is shared among n parties such that any k of them can collaborate to obtain the secret, but fewer than k cannot learn anything about the secret.
- Secure multiparty computation (SMC): This is a framework for multiple parties to cooperatively carry out a computation whose inputs are distributed among them, without revealing any party's input to the others. It can be built on secret sharing and encryption schemes that have specific properties. These computations use modular exponentiation and are resource intensive, which has impeded their use in practical applications.
- Digital certificates: Digital certificates are used for proving ownership of a public key, and rely on asymmetric encryption. They are issued by certificate authorities, well-known companies that charge for issuing such certificates. Anyone who has a certificate authority's public key can verify the ownerships asserted by the certificates it issued.
- Cryptographic hash functions: Cryptographic hash functions are message reducing functions which are one-way (i.e., difficult to invert) and collision resistant. They are used in digital signatures and for authentication purposes. Because changing any bit in the input to such a function causes huge changes in its output, it is challenging to use such functions in physical systems that are noisy or naturally change over time (e.g., through wear and tear). To enable the use of cryptographic techniques to such noisy physical data, the technique of fuzzy extractors has been proposed for reliably extracting a bitstring from the noisy data in a repeatable fashion (it is to this stable bitstring that the cryptographic techniques are then applied).

USER-CENTRIC SECURITY

Advancements in sensor technologies and communication methods have made it possible to monitor product use in real time. The users of the product could be either individuals or enterprises. Monitoring the product use could raise concerns related to privacy and data ownership.

Authorization is a process of granting permission to a user with a set of actions on a resource including read, modify, share and print. OAuth 2.0 is one of the popular authorization frameworks.

Authentication is a process of verifying the identity of a user. The verification technique can be based on the following three factors: something the user knows; something the user has; and something the user is.

Recently, two-factor authentication has evolved into an industry standard in many business fields including cloud computing.

Access control is a policy-based approach used for authorizing and authenticating users to provide access to resources. It is built on the partitioning of the information based on different user-credentials. Researchers have designed access control techniques based on roles, attributes, and other classifiers.

Anonymity; Hiding the identity of the user or, in data anonymity, the identity of the person corresponding to a record in a dataset consisting of many such records. An early anonymity technique is K-anonymity, in which the data are modified to protect users' identities while maintaining the usefulness of the dataset for statistical analyses.

Combination of security mechanisms may be necessary to simultaneously achieve multiple security attributes. The strength of security in such combinations depends on its weakest security mechanism. In order to overcome such issues, security experts suggest to offer system security in multiple independent layers, i.e., integrate different types of security mechanisms for each security objective. Designers need to be careful while combining security mechanisms, as it may lead to undesired redundancy, or security gaps that make the system vulnerable to attacks. Note that there could be ancillary uses of the stated security mechanisms in order to achieve security objectives.

The security mechanisms help designers to come up with a security architecture that prevents an attack. There is another branch of security that deals with situations after an attack is discovered. Forensic analysis includes damage estimation, identifying the method of attack and its source. In this paper, we are focusing only on the use of security mechanisms in design.

The process of combining security mechanisms poses a complex challenge for achieving the sPRP as it requires a proper combination of security mechanisms at every stage in a PRP. One of the possible ways in which different security mechanisms can be integrated into multiple independent layers while achieving multi-objective security where each member of the network has valid credentials. Network-centric security should also be deployed in the user, data, and asset-centric security layers, but this is omitted in the figure for readability. In this paper, we review the application of security mechanisms discussed in two important stages in the PRP: design, and manufacturing.

GAME THEORY

Game Theory is primarily a mathematical framework analyses the decision-making of a player based on how they expect other players to make a decision i.e., determining optimal rational choices given a set of circumstances which can be applied in many fields such as economics, politics, computer science, biology, philosophy & so on. Game theory depicts the game played between different players and the strategies of each player.

A game can be defined as interaction of different players according to a set of rules. Players may consist of individuals, machines, parties, companies or associations. The results of game theory depend upon behaviour of every other player present in the game and not only the current player.

Due to this reason, this approach is extremely scalable and versatile. The outcome of game theory also depends on the estimated payoff by each player before making decisions, which is a measure of the satisfaction obtained by each player by making that decision. Therefore, the players will perform actions and take decisions that would provide them the maximum payoff.

Types of games

In game theory, there are different types of games that help us analyse different problems. They are categorised in the basis of number of players involved, cooperation among players & symmetry of the game.