BIG DATA MANAGEMENT OF A CYBER-PHYSICAL MULTI-LOCATION CHEMICAL FACTORY

[1]ABHISHEK N. KULKARNI, [2]ANITA S. JOSHI, [3]SATISH R. INAMDAR

[1]Electronics & Telecommunications Engineering Department
[2]Electrical Engineering Department, [3]Chemical Engineering Department
[1,2,3]Vishwakarma Institute of Technology, 666, Upper Indiranagar, Bibwewadi, Pune 411037, Maharashtra State, India.
E-mail: [1]abhi.bp1993@gmail.com, [2]satish.inamdar@vit.edu

Abstract- Big data is becoming a common technical term wherever scientific computing, factory automation, information databases are to be managed for control and operation of facilities. The big data problem occurs in factory automation when the processing rate, decision making and computing algorithms slow down due to relatively big size of data, its complexity, variety and a number of operations to be done to operate any cyber-physical multi-location chemical factory. This is relatively a new area in control and automation field, although many such features exist in real life. However, the topic deserves further research to seek new solutions with growing complexity that arises due to nature and size of data.

Keywords- Factory automation, Big data, Signal processing, Cyber-physical plant, Data retrieval and processing, Concurrency and parallelism.

I. INTRODUCTION

Today the advent of a large range of digital sensors, Internet resources for communication of data in various forms, business dealings for sales, marketing and inventory control, supply chain management are few important examples why a large amount of data is being communicated for business purposes, technology and development, financial data communication, and an array of data communication activities in many fields of activity. This data can be identified as having a scale for size of data to be processed in a finite time interval based on different time scales and we are going to face new challenges in near future, which are going to transform communication systems, market, and our lives too! The terms system integration for technology and the related issues, and social integration that covers medicine, general items for people and banking systems, travel and a host of areas which the Internet and many communication systems have covered is affecting the life style of people at large.

In this paper we will consider the problem of a multi-location factory that has facilities, warehouses and business centers in various cities and regions. Later, we will point to a specific problem that we will identify for cyber-physical activities in a production facility and see howa new face of control and automation is emerging and will occupy a place as a part of social integration.

We state that this becomes a problem in terms of social integration as the consumer satisfaction level and product quality needs to be monitored as a part of global economic policy, as we are going to apply stringent control measures to operations management and get control plots and data representation graphics for the production facility as a three stage flow shop that delivers goods as a product range based on customer orders received. Clearly, the problem falls more or less within the domain of operations research.

A. Control and automation activity

We assume that the data that is transmitted by a redundant sensor network can be in the form of signals, digital binary data, or a time series from various instruments that is collected and sent to automation consoles for the purpose of data reconciliation, quality control, control over product function/ effect that determines the consumer satisfaction level for a range of products. Hence, the occurrences that may be described as discrete events are sensor failures leading to systemic disturbance upsetting the equilibrium and quality control lines and many such discrete events that cross the limits for disturbances introduced during operation of plant. Re-setting the performance then can be analyzed as a probabilistic view of plant management personnel, which is being monitored, will be needed frequently. If maintenance team has to look for replacement of actuators or any control component, a parallel standby system alternative should be available, although some cost needs to be borne as a statistical process control measure.

B. Signal processing activities and data reconciliation

The instructions for re-setting performance specifications in the event of plant component failure or pre-emptive disturbances in production schedules due to various reasons come from operations control. This operations control is connected to automation management personnel via ERP interfaces. The ERP interfaces are in turn connected to central management control at another location via Internet resources. The operations control is provided by clear view of signal processing activity that monitors the plant operation and is able to re-set plantoperation specification as per any production slate that is computed based on orders placed applying methods in operations research. It is then clear from above discussion that a number of hierarchical control layers can be designed and are forming an open connectivity for the big data (which we are yet to prove it to be so) to be processed and that decisions are made in tandem with multi-location management control and expert advices.

II. PROBLEM STATEMENT

The data analysis can be seen to be constrained in many ways. In order to begin analysis of problem, we first set objectives based on previous discussion. The objectives identified are given here.

1. To make changes in plant performance specifications based on eliminating any violation of constraints.
2. The saturation of actuators, crossing the readability range of process data requires interlocks and this may result into piecewise linear control systems requiring a different kind of adaptability.
3. The control of divergence in network sections that are a result of systemic disturbances affecting a specific section of plant wide dynamics needs to be identified

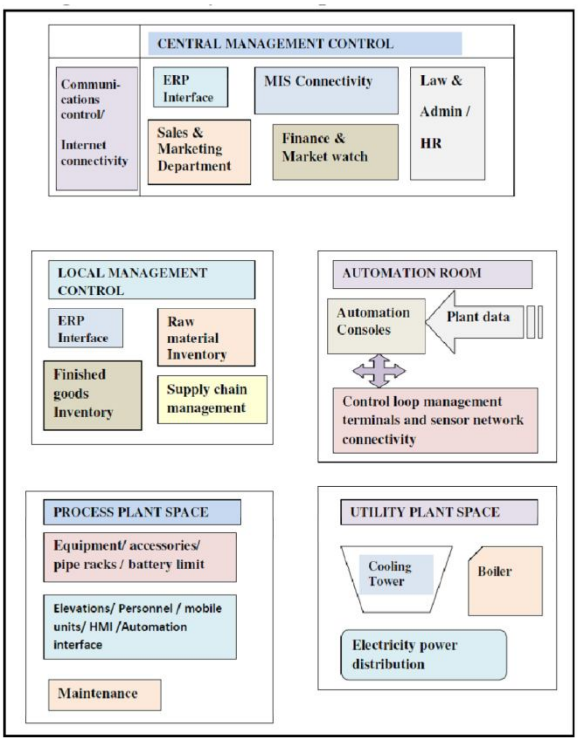

Figure 1Plant layout and operations view

4. and new network control measures that account for memory of controllers need to be defined.
5. The overlapping of concurrent control action sequences result into incorporation of parallelism into automation control decision search tree for time being considered as a logic value and mapped over a real time coordinate for subsequent corrective actions.
6. The cyber-physical activities are based on communication between personnel, intelligent mechatronics, image sensors, mobile robotic units and a few human operators having a communication set up that will coordinate and end up in effecting a time-continuous synergistic activities that are a result of fetch-instructions for every node in an affected network.

In this paper, we will limit ourselves to the activities that are related and will result from meeting the objectives by designing a layered logic system that can set up communication between multi-location agents who will be part of network monitoring and decision implementing system for the cyber-physical control set up.

III. SYSTEM REPRESENTABILITY AND ANALYSIS

The computing and communication network is the decisive system administrator that connects the operations management within each location and its open connectivity with the central command via ERP interface at another location.

This is a limiting description of the cyber-physical factory automation set up. We will make use of hybrid automata, state transition diagrams, and communication networks having an ingenuous connectivity that will reduce the probability of having communication problem, completion of task sequences being possible with time buffers provided as setting up flexibility limits. The outcome of every trace we put on signal pairs those existing within cyber-physical factory is turned into a nodal connectivity for communication as explained earlier. On the face of it two main activities get defined from statements we have made.

1. Recovery from systemic failures leading to divergence from plant performance specification; and
2. Implementation of changes to be made into production slate according to dynamic manipulations to be done in schedules that were computed as a preceding one.

A clear analysis that will represent the system operation and decision making set up needs to be a systematic network representation of the multi-location facility.

Figure 1 shows equipment arrangement and operations view of a multi-location factory and central management control, which is the simplest example we can consider in this analysis.

Our objective is to find out how the problems with big data arise in such a factory automation system. Let us begin with activity enumeration and see how their connectivity can be represented and its complexity can be summarized using probabilistic views. The signal processing problems form the basis of the analysis from which we stretch into domain of data processing, sensor network used for data collection, and how the data analytics can be explained in terms of slowdowns caused by concurrency and parallelism that sets into the operations control.

IV. CYBER PHYSICAL ACTIVITIES IN PLANT

Analysis of signal processing activities in terms of probabilistic evaluation of occurrences of certain events is one way of solving the destabilization problem. The linear matrix of the interconnected system can be given as

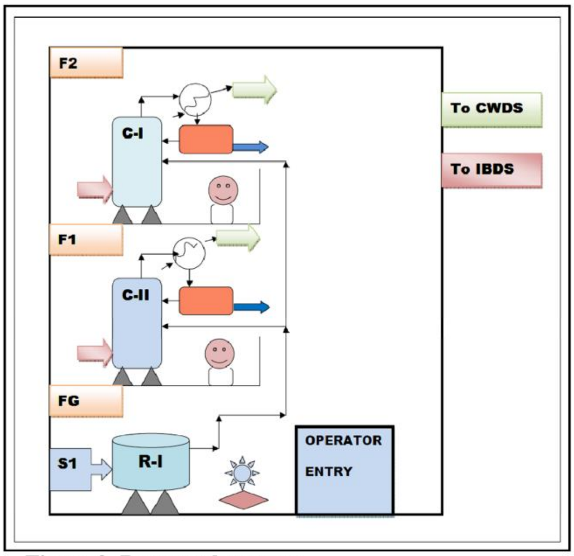

Figure 2. Process plantspace.

$$\widetilde{\Psi} = \begin{bmatrix} H & \phi & \phi \\ \vartheta & h_1 & 0 \\ \vartheta & 0 & h_2 \end{bmatrix}$$

Where the structure of each component in the interconnected system is given as

$$\Sigma_i \quad : \quad \begin{aligned} \dot{x}_i &= f_i(x_i) + g_i(x_i)u_i \\ y_i &= h_i(x_i) \\ u_i &= u_{e,i} - \sum\nolimits_{j=0}^{m} \Psi_{i,j} y_j \end{aligned}$$

It may be noted that, $i = 0, \cdots, m$; u_i is input to subsystem *I*, $u_{e,i}$ are exogenous inputs to system dynamics. Thus any exogenous input to any subsystem can upset the closed loop stability of interconnected subsystem. We will discuss more about this at the end of this Section.

For the interconnected system, for the star-shaped symmetric system consisting of a single reactor and two distillation columns in parallel, it will be possible to recover the product. The rest details are not shown as only the preliminary concept of cyber activities has to be explained. The exogenous inputs into system dynamics of columns at various time instances disturbs the stability of closed loop control system. This demands an on-line re-tuning and switching to new domain of operation. The time delays imposed by plant component failure and switching to stand-by units also are accounted into the piecewise stable closed loop dynamics. Assuming that added instrumentation and redundant sensors are available for re-setting dynamics in the event of fault occurrences, the problem then reduces to the problems of inventory control and data analytics for cyber physical controls. The type of cyber-physical system we consider here is the one having distributed computations and networked control architecture of the system. The system includes real time inputs into various components and control subsequences being computed to reach new set points for the interconnected dynamics.

The other problems that are faced during re-setting and re-tuning the network control system are variable time delays and other variations that need distributed decision support systems. Only such systems can have characteristic design features to satisfy highly reliability and necessary security requirements for data network as well as internet connectivity for operating the cyber-physical plant setup. There will be a lot of computer science related expertise needed that will check the control hardware and software that is termed as verification and validation of patchy software codes that do the job of shifting the interconnected system smoothly to another set points for all plant components. The real challenge lies in design and development of next generation software architecture that bring together big data problems via network data communication protocol, dedicated application software to create required modules (e.g. Hadoop) to solve identified problems.

Even the semantic description of cyber systems can express well the time instances and time continuity to initiate and execute control subsequences that are interdependent.

The event-based semantics too can be chosen to handle fault occurrences and switching, as it can improve the scope and versatility of control system by bringing together interaction between components and observations which is a discrete event, and that all such codes can be integrated into higher levels of sophistication easily. The topics that come up for design of cyber control systems are dependability, design and modeling as well as implementation of control action sequences as concurrency exists and any delays in scheduled task sequences will end up in parallelism being introduced into the control architecture.

The concept of cyber systems usually looks for new control system architectures and automated overall systems that have features of smart manufacturing, smart logistics and mobility, smart utilities and smart grid supply, and applies well to smart supply chain management.

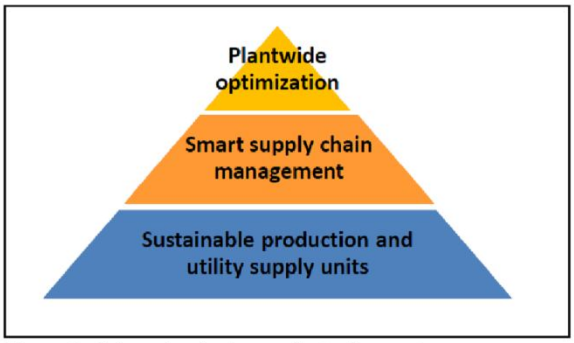

Figure 3. Cyber physical manufacturing system

The smart manufacturing features are connections and communication set-ups between plant floors (see ground (FG), first (F-I) and second (F-2) floors in Figure 2) and communications between mobile operator (see smilies on two floors and a mobile robot unit on ground floor). The communication and implementation of changes at local level or through the local management control by technical team via automation room personnel support results into a nicely defined cyber set up. The Figure 3 depicts some interesting features for the cyber physical systems.

V. CONCURRENCY LEADING TO PARALLELISM

The concurrency i.e. initially independent task subsequence starting simultaneously with short time differences, usually exists in all manufacturing plants. This needs to be accounted for when a cyber physical control systems are being designed. Any overlapping (that is absent initially) due to time delays introduces parallelism into cyber physical control operations.

The point we wish to discuss here is the presence and occurrence of bottlenecks of various kinds that will pose difficulties in re-scheduling and re-starting cyber control operations. Figure 4 (a)-(c) show diagrams of different causes of bottlenecks to occur or be present as inherent problems in the manufacturing system set ups.

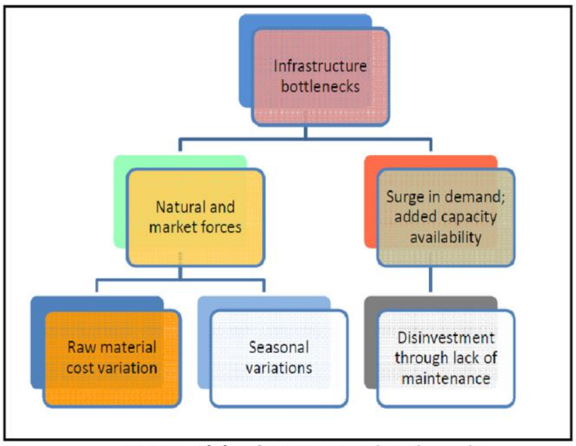

Figure 4(a) Infrastructure bottlenecks

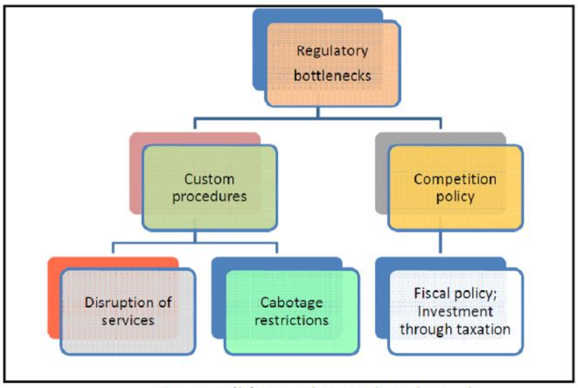

Figure 4(b) Regulatory bottlenecks

The other types of bottlenecks that can occur or can be present are regulatory and supply chain bottlenecks. These are shown in Figures 4(b) and Figure 4(c) here. The regulations in logistics often cause delays in goods movement. If they are at international level, more delays may be expected. So in addition to facility inventory and warehouses, wait time for cargo to be lifted can be added. Even though the intention is not to cause and carry forward various delays in logistics operations, the regulations inevitably cause delays as well as disruptions. These sources of bottlenecks cause delays as created by the indirect effects of regulation. The cabotage restrictions will prevent foreign carriers to carry the goods within country.

There are also capacity limitations for freight within a country. Competition policies can effect bottlenecks. This happens by virtue of either supporting a monopoly (here rent seeking strategies are sought), or a complete deregulation as many carriers will compete of the similar transport segments. Also, fiscal policies can affect investments that come through taxation and end up in causing bottlenecks.

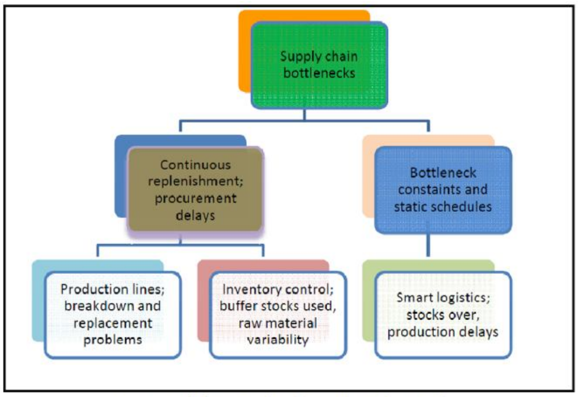

Figure 4(c) Supply chain bottlenecks

The bottlenecks trigger into supply chain system due to various reasons. There are bottleneck constraints in static schedules formed as a plan. Also, the continuous replenishment causes procurement delays. The static schedule gets upset due to constraints and limitations on capacity. Examples for delays in static schedules are labor unavailability for work shifts. The production delays and storage capacity limits may impose time dependent capacity shortages. At times, some firms may effect bottlenecks with an objective of rent seeking strategy. Production line breakdowns and replacement of accessories can upset schedules. The inventory control measures show procurement delays using up buffer stocks and raw material variability increasing processing times. These can cause bottlenecks too.

How all these issues are measured as a model that quantifies the data to be processed, stored and communicated is the focus of this discussion. The operations to be done on data in various forms is limited by several factors such as processor speed and capacity, size of data, control sequences scheduled but not completed leading to parallelism being set into data operations.

This will require more number of serves that can be assigned parallel tasks and off-line processing requirements. We will see how a qualitative statistical analysis can shed light on the problems of big data management and analytics.

VI. STATISTICAL ANALYSIS OF BIG BYTES

In order to begin analysis of big data, we need to define why it is termed a big data problem! Let us find out how many types of data are to be stored and classified to get a clear picture of big data. Although there are many inherent reasons of time delays in operations, we will try to focus on operations research view of bottlenecks causing data accumulation for processing and fining remedy to minimize the tardiness in the schedule as a dynamic strategy. We classify the data sources into three main types. These are data related to, 1. Inventory and logistics control, 2. Plant control and automation, and 3. Communication and Internet data transfer.

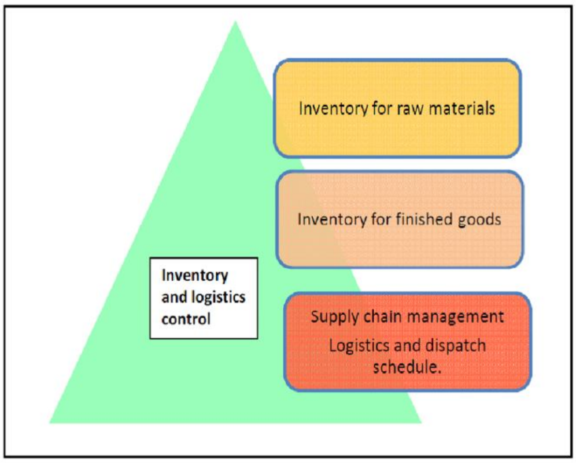

Figure 5 (a) Inventory and logistic control big data

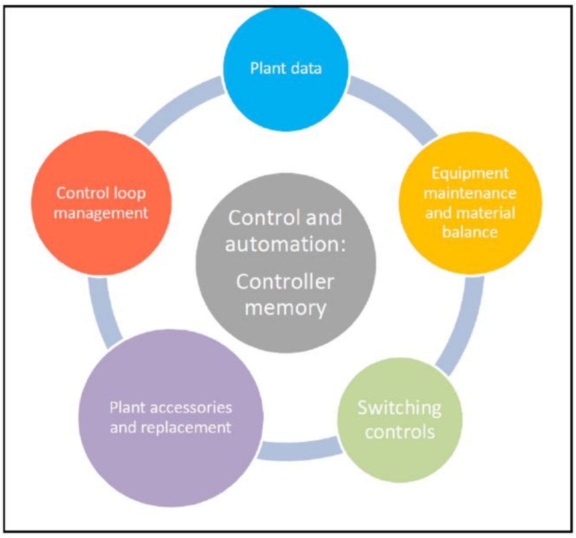

Figure 5(b). Controller memory and big data sources.

Figure 5(a) shows inventory and logistics control and big data sources. The inventory of raw materials and finished goods focus onto data accumulation sources that can cause bottlenecks in data processing, retrieval and storage. It also shows that supply chain data and logistics scheduling data accumulates and also needs backup, warehousing and preparing current data files, i.e. it points to stacking up data for some period that should not hamper speed of processing for current orders and deliveries. The Figure 5(b) depicts the controller memory that uses multiple processors to take of concurrent processes that may start almost simultaneously at times and are initially independent of each other at start instances. Later if precedence may be existing that introduces delays, wait times and ends up in stackpile of previous history, when new computations set up new targets and control action sequences. These are relatively large bytes of data and frequency w.r.t time scales is frequent. It also considers theshort breakdowns and partial standby support for replacement. Other sources of big data are HMI databases, local data inventory of mobile units, automation consoles and databases retrieval and a host of other operations on various types of data; and lastly we add the local databases storage for sensor networks and individual sensors

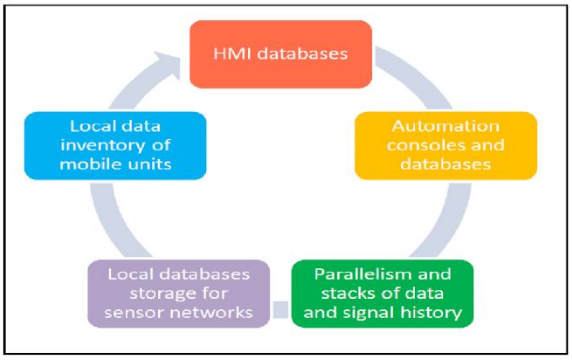

Figure 5(c). Communications and big data sources.

Figure 5 (c) shows communication network including Internet connectivity. Putting the three sources together, it gives rise to a big data problem.

CONCLUDING REMARKS

Thus, we have observed that the capacity considerations, processing speed with minimum number of processors, and Internet communication speed and problems and problems in decision making end up in causing various bottlenecks from operations research point of view and at the same time sets up concurrency and parallelism into control system operations. This big data problem will be applied to a concrete case study and will be published elsewhere.

REFERENCES

[1] Binod Timilsina, Removing bottleneck from a manufacturing unit: A case studies to BET-KEROY, Ylivieska-84100, Finland.

[2] Jos´e A. Joao et. al., Bottleneck Identification and Scheduling in Multithreaded Applications, ASPLOS'12 March 3–7, 2012, London, England, UK.

[3] E. Ebrahimi, R. Miftakhutdinov, C. Fallin, C. J. Lee, J. A. Joao, O. Mutlu, and Y. N. Patt. Parallel application memory scheduling. In MICRO, 2011.

[4] Ray, A., Sarkar, B., & Sanyal, S. (2010). The TOC-based algorithm for solving multiple constraint resources. IEEE Transactions on Engineering Management, 57 (2), 301-309.

★★★