Detecting disturbances in supply chains-The case of capacity constraints

<table>
<thead>
<tr>
<th>Journal:</th>
<th>International Journal of Logistics Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID:</td>
<td>IJLM-12-2015-0223.R2</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Original Article</td>
</tr>
<tr>
<td>Keywords:</td>
<td>Supply chain risk, capacity constraint, disturbance detection, clustering</td>
</tr>
</tbody>
</table>
Detecting disturbances in supply chains-The case of capacity constraints

Abstract

Purpose – The ability to detect disturbances quickly as they arise in a supply chain helps to manage them efficiently and effectively. This paper is aimed at demonstrating the feasibility of automatically, and therefore quickly detecting a specific disturbance, which is constrained capacity at a supply chain echelon.

Design/Methodology/approach – Different supply chain echelons of a simulated four echelon supply chain were individually capacity constrained to assess their impacts on the profiles of system variables, and to develop a signature that related the profiles to the echelon location of the capacity constraint. A review of disturbance detection techniques across various domains formed the basis for considering the signature based technique.

Findings – The signature for detecting a capacity constrained echelon was found to be based on cluster profiles of shipping and net inventory variables for that echelon as well as other echelons in a supply chain, where the variables are represented as spectra.

Originality/value – Detection of disturbances in a supply chain including that of constrained capacity at an echelon has seen limited research where this study makes a contribution.

Keywords Supply chain risk; capacity constraint, disturbance detection, clustering

Paper type Research paper

1 Introduction

Supply chains are vulnerable to disturbances. Described as unplanned and unanticipated events that disrupt the normal flow of goods and materials in a supply chain and expose associated firms to operational and financial risks (Craighead et al., 2007), disturbances could be catastrophic such as earthquakes, fires, hurricanes and terrorist attacks, or recurring, such as delays in material deliveries, process losses and inappropriate information processing (Chopra and Sodhi, 2004). Together with risks which can be characterized as potential occurrence of disturbances, disturbances in supply chains have been a subject of intense research in recent times (Juttner, 2005; Wagner and Bode, 2008; Thun and Hoenig, 2011; Manuj et al., 2014). The important area of disturbance detection however, appears to have been overlooked.
Detecting a disturbance, and doing so early/quickly, enables it to be more efficiently and effectively mitigated as the greater reaction time causes more mitigation options to be available. As an example, consider loss of production capacity at a supplier which could delay incoming supplies at a production facility. An early detection of this disturbance could mean mitigation options such as using alternative suppliers or rescheduling production to produce alternative products or preparing customers for late delivery being available; a delayed detection on the other hand could cause most or all of these options to be unavailable resulting in idling of resources and/or delayed deliveries to end customers with associated penalties, and thereby higher economic consequences (Bodendorf and Zimmermann, 2005).

Though the advantages of early supply chain disturbance detection are well recognized (Craighead et al., 2007; Blackhurst et al., 2011), actual interest in this subject as per the extant literature has been limited. In fact, disturbance detection itself is discussed in only a few studies; the discussion is also largely conceptual with the effectiveness of the detection approaches discussed also being questionable. For example, the practice of physically detecting and communicating disturbances as discussed in Svensson (2000) and Norman and Jansson (2004) is mostly between adjacent and not multiple echelons across supply chains, and whose subjective nature leaves scope for errors in classifying a disturbance as such or disturbances being not reported/misreported.

Other more objective and automated approaches such as track and trace that are used in logistics contexts also have limitations, such as disturbances of only a specific kind like delivery delays being detectable, the nature of detection being transactional i.e. lacking in intelligence, and the operational scope of detection being limited to a few echelons (Karkkainen et al., 2004). Similarly, supply chain event management (SCEM) systems, which are an adaptation of track and trace for supply chain contexts and detect discrepancies in ordering and fulfillment processes are also transactional in nature and lack multi-echelon analysis ability i.e. the ability to consider the impact of disturbances in a cascaded structure (Otto, 2003). A need therefore exists for:

- an information-centric disturbance detection approach (to avoid subjectivity/bias in detection),
- which uses information from multiple echelons (to enable a supply chain wide detection perspective),
- with the information being automatically processed through the use of protocols (to enable speed, scale and intelligence in detection), and
- which could be applied for a variety of disturbances.
Such an approach has been successfully applied in other domains (Davenport and Harris, 2005) but not pursued for supply chain disturbances despite calls to that effect (Christopher and Lee, 2004; Blackhurst et al., 2005). Demonstrating the feasibility of this approach and suggesting a framework for its practical use forms the focus of the exploratory research that we present here. It is aimed at developing a protocol based on system state information, i.e. time series data of variables such as orders, inventory and shipping for automatic, and therefore quick, detection of disturbances in supply chains, with the focus being on a specific disturbance, the capacity constraining disturbance. In theoretical terms, the research contribution can be explained through systems theory (Bertalanffy, 1951) which argues for systems to be treated holistically rather than in terms of its constituent parts. Though this theory has been widely used in generic supply chain work due to the interconnected nature (involving material and information flows) of supply chain entities (Sterman, 2000; Frankel et al., 2008), here it is being used in the context of supply chain disturbance detection and management which is not seen previously. Specifically, negentropy or tendency of systems to reduce entropy, one of the key aspects of systems theory is relevant here, with a ‘disturbed’ system leaving a trail as it transitions to a more stable/ordered state over time.

The rest of the paper is structured as follows. In the next section we highlight the need for capacity constraint detection in supply chains. In Section 3, we explore techniques being used in different domains to detect disturbances so that related learnings could be applied here. In Section 4, we discuss findings from simulation of a four echelon supply chain system incl. evolution of a protocol for detecting capacity constraint at an echelon. Finally, the practical aspects of applying the protocol based quick disturbance detection approach are discussed in Section 5, which is also the concluding section.

2 Problem statement and capacity constraints

Capacity constraints in supply chains arise frequently and cause significant economic loss through associated delivery delays/short supply to customers (Chopra and Sodhi, 2004; Wagner and Bode, 2008). They are also difficult to avoid given that knowledge of echelon capacities is dispersed across the supply chain (Norrman and Jansson, 2004), and market demand and operations failures that cause capacity constraints are unpredictable in nature. Their choice as a disturbance for developing the protocol based quick disturbance detection approach is therefore appropriate. The protocol in this case should help detect whether a supply chain is capacity constrained, and if so, its echelon location so that appropriate demand and/or supply
management related countermeasures could be applied. Such a protocol is not seen in the previous academic literature.

Development of the above protocol requires an understanding of the impact of capacity constraints on system variables, which could be best achieved through controlled experimentation. Simulation, specifically control engineering based system dynamics simulation is therefore considered, with a well known four echelon system used in several previous supply chain dynamics studies incl. Sterman (1989) and Shukla et al (2009) being simulated. Although some work on capacity constrained systems using system dynamics modelling has been done previously, the nature of investigation in those studies is different to what is done here. For example, in Wilson’s (2007) investigation of system behaviour under loss of capacity, the focus is on the dynamics of inventory and backlogs, while in Spiegler and Naim’s (2014) case the emphasis is on the dynamics of shipments. Investigations in these studies are therefore limited to a few/select variables unlike the multiple variable analysis covering orders, shipments and inventory attempted here, and which is also more appropriate given that these variables are all systemically linked. Moreover, previous studies have used the dynamics of variables to understand/improve system performance, while the focus here is to use them and their characteristics to develop a detection protocol.

Therefore to summarise, our aim in this study is to develop a protocol based on system information for quick detection of capacity constraint (at an echelon in a supply chain), and thereby demonstrate the feasibility of this kind of an approach for detecting supply chain disturbances in general. Usefulness of this approach however depends on the availability of mitigation options that could utilise the advance disturbance information. The choice of these options however, do not form a part of this study and can be referred to elsewhere (Tomlin, 2006; Schmitt and Singh, 2012; Shao and Dong, 2012).

3 Generic disturbance detection techniques and their applicability

Our goal here is to briefly outline alternative techniques and assess their suitability for developing the capacity constraint detection protocol. Several domains, where disturbances are automatically (and quickly) detected through techniques based on system information, were identified in the literature. The underlying philosophy is based on Ansoff (1975), who first advocated the use of early warning or weak signals to manage unpredictable, dynamic and hence difficult to plan contexts, and Haeckel’s (1999) sense and respond approach, that involves
sensing changes in dynamic contexts through probes/data and filters/models/data mining methods, and then responding with corrective action/s, if needed. Table 1 below gives these domains, the disturbances encountered therein, the technique/s used for detecting the disturbances and the variable data input to operationalise them. Sub-techniques are not highlighted as our intention is to focus at a generic level.

Looking at Table 1, it is clear that in most domains, predominantly two kinds of techniques termed as anomaly based and signature based are used for disturbance detection. Anomaly based ones involve modeling the behaviour of system variables in statistical terms such as by fitting a probability distribution or a time series model and then monitoring deviations (vis-à-vis usual/normal behaviour), with an alarm being triggered in case of a specified threshold being breached (Basseville and Nikiforov, 1993; Garcia-Teodoro et al., 2009). Though successfully applied in manufacturing contexts for process and quality control (Montgomery, 2005), these techniques have not seen significant application at a supply chain level (MacCarthy and Wasusri, 2002), which could be due to challenges such as: 1) Difficulty in stochastic modelling of variables’ dynamics for multistage systems such as supply chains (Batson and McGough, 2007; Tsung et al., 2008), 2) Difficulty in specifying thresholds due to the non-stationary nature of the supply chain variable profiles (from the continuous changes in the internal and external environment), 3) Involvement of a large number of system variables in a typical supply chain assessment, where these techniques are known to be less effective (Woodall and Montgomery, 1999), and 4) Difficulty in effecting optimal mitigative responses post-detection, as a ‘disturbed’/’not-disturbed’ kind of detection (rather than indication of the specific disturbance impacting a system) is provided by these techniques.

Signature based techniques on the other hand involve formulating signatures, or characteristic profiles of variables associated with specific disturbances, that could subsequently be used to detect them in test cases (Han and Kamber, 2006). These techniques are versatile and not constrained by either the number of variables involved or the nature of their profiles (as gauged by detailed examination of the studies in Table 1 that have applied these techniques). Moreover, the fact that these techniques operate on a pre-specified set of disturbances (for which signatures have been formulated), enables them to determine the exact disturbance affecting a system, thereby enabling specific mitigative actions to be applied. Given these advantages, use of the
signature based techniques to develop the capacity constraint detection protocol is therefore further explored.

Signatures could be developed through several alternative methods including classification, association and clustering (Ngai et al., 2011), among which the last, i.e. clustering, is the most preferred for time series data (Liao, 2005). Clustering is also relatively easy to interpret and therefore, the signature developed through this method can be expected to have more practical utility. Clustering involves partitioning of data sets into a few homogenous groups or clusters, whose characteristic profiles for individual disturbances could be used as signatures for their detection. Application of this method however, requires choices to be made in terms of: a) System variables whose time series data is to be input, b) Domain in which to represent the data.

The system variables whose time series data is to be input could vary considerably for different collaborative scenarios. However, orders, inventory and shipping are the variables on which information is most commonly shared between organizations (Lee and Whang, 2000; Huang et al., 2003). A signature involving these variables would therefore not only be less complex given the few variables used, it would also encounter less data access problems. Also, as per researchers such as Tang (2006) and Elkins et al. (2008), disturbance detection is possible through such information.

Another important consideration in applying the cluster profiles based signature technique is the domain to be used for representing data. Alternatives include time or features such as autoregressive model coefficients, autocorrelation function, spectra or amplitudes of constituent sinusoids (derived through Fourier transform or FT) and wavelet coefficients derived through discrete wavelet transform of the time series (Liao, 2005). While time domain findings are easier to interpret, data in this domain tends to be noisy which makes signature formulation and application error prone. Representation in the feature domain is therefore increasingly preferred, with spectra considered to be among the most popular features (Keogh and Kasetty, 2003), and which has also been previously tried for supply chain contexts (Shukla et al., 2012). Specific advantages of the spectra include its invariance to time delays that are typically encountered in supply chain contexts and tend to mask underlying profile relationships (Thornhill and Naim, 2006), and its lower sensitivity to missing values. We therefore explore the development of a signature based on cluster profiles of order, inventory and shipping variables, where these variables are represented as spectra, to detect capacity constraint (as the disturbance) in a supply chain.
4.1 System simulated

The well known four-echelon supply chain system of the Beer Game (Sterman, 1989) was simulated for the investigations. Retailer (R), Wholesaler (W), Distributor (D) and Factory (F) are the four echelons in this system where orders flow upstream from the Retailer through to the Factory, and in response, products are shipped downstream. There are delays (also referred to as lead times or LT) in order transmission, shipping and production and a provision for backlogs in case the stock level is inadequate to meet demand. The system is triggered by market demand at the Retailer with individual echelons controlling stock and ordering in each time period. The simulation involved translating the Beer Game system structure and mechanics into difference equations, with the Automatic pipeline, inventory and order based production control (APIOBPCS) heuristic (John et al., 1994) used at each echelon to decide the order quantities.

APIOBPCS, which mimics human behaviour (Sterman, 1989) and is representative of industrial practice (Evans et al., 1997) has been used in previous studies (e.g. Shukla et al., 2009 and Spiegler and Naim, 2014). It can be described as: the order placed is equal to the average sales rate plus a fraction $\left(1/T_i\right)$ of the inventory error or the difference between desired and actual inventory plus a fraction $\left(1/T_w\right)$ of the work-in-process (WIP) error or the difference between desired and actual WIP, where $T_i$ is the time to adjust inventory and $T_w$, the time to adjust WIP. The average sales rate is based on exponential smoothing, where $T_a$ or time to adjust demand is related to the exponential smoothing parameter $\alpha$. While $T_p$ is average delay/lead time between order placement and delivery/production output and is therefore a physical parameter, $T_i$, $T_w$ and $T_a$ are decision parameters whose values are decided on the basis of criterions such as lower order and inventory variance and greater system responsiveness. For the investigations, ‘hard’ systems based rational parameter values as per John et al (1994) were considered which are: $T_a=2T_p$, $T_i=T_p$ and $T_w=2T_p$. Figure 1 gives the block diagram representation of APIOBPCS, where CONS is the sales rate, AVCONS, the forecast of sales rate, ORATE, the order rate, DWIP and EWIP, the desired and error in work in process levels respectively, AINV, DINV and EINV, the actual, desired and error in inventory levels respectively, and COMRATE, the completion rate or rate of goods receipts into inventory.

The difference equations used to simulate the Beer game (including APIOBPCS based ordering) are given in Appendix I. While initial demand at 4 units is considered the same as in Sterman
(1989) (refer equation 2.1a), different demand patterns are considered for subsequent periods (refer equation 2.1b). Equation 2.1c models the delay (lead time or LT) in communicating the order information with delays in material shipments and factory production captured in equations 2.2a and 2.2b respectively. Material shipment dynamics including creation of capacity constraints are represented through equations 2.3 to 2.6; shipment is set to zero when there is no inventory available, and alternatively, when it is available, ordered quantity plus accumulated backlog quantity is shipped subject to adequate shipping capacity being available (as discussed in the next sub-section). Inventory and backlogs are computed separately through equations 2.7 and 2.8 respectively. Sales forecast based on exponential smoothing is calculated through equation 2.9, while the desired, actual and error in inventory and WIP are calculated through equations 2.10 to 2.14. Finally, ordering based on APIOBPCS is captured through equation 2.15, where the parameter values used are as per John et al (1994) discussed earlier.

4.2 Capacity constraint due to unexpected increase in market demand

4.2.1 Analysis in the time domain

We first analyse capacity constraint from a step change in market demand in a system with a delay/lead time profile of 2,2,3 time units. The related simulation outputs are depicted in figure 2. The profiles are all normalised i.e. they have been mean centered and amplitude scaled to enable comparison.

Row 1 depicts the variable profiles for the case where no capacity constraints exist. The profiles are all oscillatory with delayed propagation of oscillations from downstream to upstream echelons. Such a behaviour is to be expected given the cascaded system structure, delays in information and material flows, and feedback loops, and which also conforms to that observed for the physical version of the system (Sterman, 1989) thereby verifying the simulation.

The profiles of variables corresponding to the Retailer being capacity constrained are shown in Row II. This involved restricting only the Retailer’s maximum shipping capacity to 5 units/time unit. Examination of the Retailer shipping profile shows it to be a step and which is then flat at the increased level for an extended period. While the step can be explained as arising from a
similar change in market demand, the subsequent flat profile represents sustained shipping at maximum capacity aimed at eliminating backlogs. The profile for Retailer orders (as seen in Row II, Column A) however, appears to be unaffected and similar to that for the unconstrained case. While intuitively surprising, such a behaviour is logical given that none of the Retailer order quantity determinants are affected by capacity constraint in Retailer shipping: while market demand is considered exogenous, the net inventory gap is unaffected with an increase in backlogs being negated by an equivalent increase in inventory from the inability to ship, and the pipeline gap remaining unchanged given that it is derived from unaffected previous orders. With Retailer orders that wholly drive the dynamics of upstream echelons being unaffected, it is reasonable to expect system variables in those echelons also to be unaffected (vis-à-vis the unconstrained case). Inspection of the order, shipping and net inventory profiles for the Wholesaler, Distributor and Factory echelons confirms this to be the case as seen in Row II.

Row III of figure 2 portrays the profiles of variables corresponding to the Wholesaler being capacity constrained. On observing these profiles the following is evident: a step jump in Retailer and Wholesaler shipping, Retailer net inventory steadily moving from a surplus to a backlog position which commences around the time of the step change in market demand, and Retailer orders showing bounded exponential growth i.e. an initial exponential growth which reaches a limiting level over time. These can all be rationally explained.

1. A step change in market demand causes a sudden increase in Retailer shipping and consequential increase in Retailer orders to the Wholesaler.
2. The Wholesaler copes by rapidly increasing shipping but given the limited shipping capacity, is only partially able to satisfy the Retailer’s orders.
3. The Retailer’s net inventory level therefore starts to reduce and move from a surplus to a backlog position causing a further increase in Retailer orders, whose profile therefore starts to show an exponential increase.
4. The Wholesaler’s inability to completely satisfy Retailer’s orders and the resulting accumulation of backlogs causes it to continue shipping at maximum capacity for an extended period of time which explains its shipping profile being flat after the initial step jump.
5. Accumulation of backlogs at the Retailer requires that it also ship at maximum capacity, which however is constrained by material receipts from the Wholesaler.
6. Retailer shipping therefore comes to acquire a similar flat profile as Wholesaler shipping after the initial step jump.
7. As regards other variables, orders placed by upstream echelons i.e. Distributor and Factory orders are seen to retain their oscillatory profiles which are not significantly different from their unconstrained versions despite the cascading affect of increased Retailer orders.

8. The same is true for Distributor and Factory shipping and net inventory, which is not surprising given the systemic linkages between orders and shipping and net inventory variables.

9. Finally, Wholesaler orders and net inventory do not appear to be significantly affected by Wholesaler shipping capacity constraint. The explanation for this is the same as that provided for an analogous situation in case of Retailer capacity constraint.

The variable profiles corresponding to capacity constraint at the Distributor are depicted in Row IV. Examination of the variable profiles shows them to be analogous to those for capacity constraint at the Wholesaler.

The above investigation was repeated with delay/lead time profile of 3,3,4 time units and a maximum shipping capacity of 6 units/time unit at the constrained echelon, although the relevant plots are not shown here to conserve space. A comparison of variable profiles for the 6 units/time unit maximum shipping capacity case with those for 5 units/time unit showed the difference between the two to be marginal, while those for delay/lead time profiles of 3,3,4 and 2,2,3 time units were seen to be lags/leads of each other, but with similar characteristics.

While in the above discussion, variable profiles corresponding to (shipping) capacity constraint at different distribution echelons could be rationalized, the next step is to do an inter-echelon comparison, in order to identify distinguishing characteristics/patterns that could be associated with the echelon location of the capacity constraint. Such a comparison of the plots in figure 2 reveals the shipping and net inventory profiles to be providing this discriminating pattern, which can be articulated as: capacity constraint at an echelon is characterized by

a) that echelon and all its downstream echelons having similar shipping profiles and which are dissimilar to the shipping profiles for the other echelons,

b) all downstream echelons having similar net inventory profiles which are dissimilar to the net inventory profiles for the other echelons.

Applying this to the case of capacity constraint at the Distributor echelon (Row IV) it can be seen that:
a) Shipping profiles for the Distributor and the echelons downstream, i.e. Wholesaler and Retailer, are similar but are dissimilar to the shipping profile for the Factory echelon.

b) Net inventory profiles for the echelons downstream, i.e. Wholesaler and Retailer, are similar but are dissimilar to the net inventory profiles for the other echelons, i.e. Distributor and Factory.

Order profiles were not seen to significantly discriminate the (different) echelon locations of capacity constraints and were therefore not considered.

4.2.2 Analysis in the frequency domain and signature formulation

The next important step is to explore if the above pattern could be automatically assessed so that capacity constraints at echelons could be detected speedily, consistently and at scale. We explore the use of clustering, which would mean representing the pattern as an equivalent cluster profiles based signature. Clustering however, works on the basis of mathematical precision, and therefore, the domain in which to represent variable data would need to be carefully chosen. In the time domain, there is a risk of similar but non-aligned profiles, that are typically encountered in supply chain contexts from lead times in processing (Thornhill and Naim, 2006; Shukla et al., 2012), and also seen in figure 2, being mis-clustered. Use of the spectra representation, or amplitudes of sinusoids derived by Fourier transformation of the time series, which is known to ignore lags/leads in profile relationships, is therefore explored. Figure 3 depicts the spectra and cluster profiles based on the spectra for shipping and net inventory variables, under capacity constraints at the Retailer, Wholesaler and Distributor. The spectra are scaled to the same maximum peak height to improve visualization. They are seen to stop at 0.5 on the frequency axis on account of the Nyquist sampling theorem (Chatfield, 2003), which requires the sampling frequency to be at least twice the maximum frequency contained in the data.

Examination of the spectra profiles depicted in the first two columns, i.e. Column’s A and B, in figure 3 shows all of them to have low frequency content, thereby accurately reflecting the trend created in the profiles from the step change in market demand. The effectiveness of the spectra in capturing lagged/leading profile similarities is also apparent. For example, for the case of capacity constraint at the Retailer, while in the time domain, the net inventory profiles for different echelons show a lagged similarity (refer Figure 2, Row II, Column C), they appear to be similar as such in the spectra representation (refer Figure 3, Row I, Column B), which is also
reflected in their being clustered together (refer Figure 3, Row I, Column D). A similar effectiveness of the spectra can be seen for capacity constraints at the Wholesaler and Distributor also. It is important to clarify that Rand Ref seen in figure 3 is a random time series that has been included to better characterize the cluster profiles.

Next, we assess the correspondence between the pattern discussed earlier as applied to capacity constraints at the Retailer, Wholesaler and Distributor and their respective cluster profiles seen in Columns C and D in figure 3; a good correspondence is seen between the two. For example, in the case of the Distributor being capacity constrained, Distributor, Wholesaler and Retailer shipping are seen to be clustered together and away from Factory shipping. Also, the net inventory profiles for the Wholesaler and Retailer are clustered together and away from those for the Distributor and Factory. These cluster profiles correspond exactly with the pattern for capacity constraint being at the Distributor discussed earlier. A similar correspondence is seen in case of capacity constraints at the Wholesaler and Retailer also. Hence, the pattern can be translated into an equivalent cluster profile based signature for locating the capacity constrained echelon. It is given in Table 2 below.

Looking at Table 2, the first three rows are relevant for a three distribution echelon supply chain like the Beer game. However, the last row shows the applicability of the signature for a supply chain with any number of distribution echelons; if for any echelon X in this supply chain, shipping variable for that echelon and shipping variables for all (its) downstream echelons are found to be clustered together and separately from others AND net inventory variables for all downstream echelons (of X) are clustered together and separately from others, then echelon X is indicated as being capacity constrained. Each echelon in the supply chain would need to be tested through the signature in this way; if at the end none is identified as being capacity constrained, then this would indicate the whole supply chain as being un (capacity) constrained. Therefore, detection of both whether a supply chain is capacity constrained i.e. any of its echelons is capacity constrained, and its echelon location is possible through the signature.

4.3 Capacity constraint due to a reduction in fulfillment capacity
The previous discussion involved capacity constraint creation from a step change in market demand and the distribution echelon’s inability to increase shipping capacity as required. We now consider a reduction in the distribution echelon’s shipping capacity as the cause of the capacity constraint, which could be caused by any of equipment/logistics/manpower related failures.

We first consider a system with a delay/lead time profile of 2,2,3 time units exposed to a constant market demand of 4 units/time unit. While its dynamics is discussed, the related plots are not presented due to a paucity of space. For the case with no capacity reduction, and therefore no capacity constraint creation at any echelon, order, shipping and net inventory profiles were all seen to maintain their initial steady state level as expected. On the other hand, for capacity constraints at the Retailer, Wholesaler and Distributor from reduction in their respective maximum shipping capacities to 2 units/time unit, the dynamics of the shipping and net inventory variables were seen to be analogous to those for capacity constraint from step change in demand in each case, and which can be similarly rationalised. The cluster profiles of the spectra of variables also showed complete correspondence with those given in Table 2 for the three echelons. We next repeated this analysis with Gaussian market demand. The corresponding time series profiles of variables and clustering of their spectra are given in figure 4.

On examining the variable profiles, their characteristic features for capacity constraints at different echelons appear to be similar to that for the constant market demand case discussed above, and capacity constraint from the step change in market demand case discussed earlier. The cluster profiles of variables show a similar correspondence, and which are as per the signature given in Table 2. Also, alternative delay/lead time profiles of 2,2,3 and 3,3,4 time units, and reduction in the maximum shipping capacity to 1 unit/time unit instead of 2 units/time unit were not seen to significantly affect the profile characteristics, with the signature in Table 2 still found to be valid.

Overall, it was possible to develop a signature for detecting the capacity constraint disturbance in a supply chain system. The signature, which is based on time series profiles of variables, specifically their spectra representations and clustering, is intuitive and amenable to automation thereby enabling speed, objectivity and scale in detection.

5 Discussion and Conclusions
In this study we demonstrated the feasibility of detecting capacity constraint at an echelon in a supply chain through a protocol based on system information, i.e. time series data of system variables. Such a protocol for capacity constrained systems is not seen in the previous literature and is practically relevant given the fluctuations in demand/capacity availability and resulting constraints that are encountered in supply chains. It applies to downstream supply chains, where the time sensitivity of demand and management complexity is greater, and which would therefore particularly benefit from the speed and automation in detection provided by the protocol. A large company/Retailer/3PL player that is managing many downstream supply chains for example could locate capacity constraints quickly, consistently and efficiently through use of the protocol. However, during application, the following would need to be considered:

**Availability of system information:** Supply chain wide system information is required to operationalise the protocol. However, this requirement is only for a few variables (shipment and inventory), and which should also be easily accessible for many supply chains (which share information for replenishment efficiency). More generally, the increasing thrust on information exchange and collaboration (Fawcett et al., 2007), further boosted by the discourse on big data and its benefits, where protocols of the kind discussed here are needed (LaValle et al., 2011) means that availability of information can be expected to be less of a hindrance.

**Contexts considered for detection:** While generally it would be beneficial to include all supply chains under a common protocol based detection approach to ensure management efficiency and consistency, a selective focus may be more appropriate in some cases. For example, newly established supply chains, supply chains with a history of capacity issues, and supply chains involving critical customers could be separately monitored. Similarly, from a product perspective, the protocol based assessment could be done at an aggregate level or only for certain key/critical products. Blackhurst et al. (2011) provide a good basis for deciding the criticality of supply chains and products from a disturbance management perspective.

**Mitigation options after detection:** Quick detection of capacity constraint at an echelon would only be useful if appropriate mitigation options could subsequently be quickly deployed. This would be possible only if mitigation options are clearly planned for and specified for each product/entity in the supply chain being assessed through the capacity constraint detection protocol. These options could be based on cost and customer service considerations and could include doing nothing (and accumulating backorders), using alternative shippers, and serving from alternative locations where capacity is available.
Technical aspects: The protocol is based on cluster profiles of spectra that are not difficult to interpret. However, there are important technical considerations such as frequency of data sampling, length of data to be used to develop the cluster profiles, and frequency of the protocol based assessment. The frequency of data sampling could be based on the transaction (volume) dynamics while a rolling window of appropriate data length that effectively balances sensitivity and false alarm considerations in detection could be used to develop the clusters profiles. The frequency of doing the protocol based assessment could be determined on the basis of the tradeoff between the increased effort/cost associated with more frequent assessments and the benefits from earlier detection.

While this study’s focus was on a specific disturbance, and only on developing a protocol for its (speedy) detection, the process used can be replicated for other disturbances. A company interested in detecting different disturbances in various supply chains under its management could follow an approach where:

a) Representative systems dynamics models of supply chains are first created and specific disturbances individually applied to investigate and understand their impact on the profiles of system variables,

b) The discriminating characteristics, as reflected in the profiles of specific system variables for those disturbances, are then developed into separate signatures/protocols for their quick detection,

c) These signatures/protocols are then made a part of a centralized disturbance management system, where based on relevant time series data of variables as inputs, the associated disturbances could be quickly detected. Specific mitigation options that have been planned for in advance for each disturbance are then quickly activated.

Such an approach would be most suited for recurring disturbances. This is because signatures/protocols presuppose repeating characteristics, and also the fact that the accumulated cost impact of these disturbances would be significant, and also quantifiable, which would enable investment in a disturbance management system to be justified.

This study, while contributing in terms of developing a protocol for detecting capacity constraints in supply chains, also has some weaknesses which create opportunities for further research. One limitation is the use of a sequential chain for the investigation. It would be useful
to consider different supply network configurations to assess the effectiveness of the protocol. Use of empirical data, though difficult to get for a capacity constrained system, could also be explored for validating the protocol. Another stream of research could be to try to develop signatures/protocols for some of the other recurring disturbances affecting supply chains.

References


<table>
<thead>
<tr>
<th>Domain</th>
<th>Disturbance detected</th>
<th>Detection Techniques</th>
<th>Typical information used for detection</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Banking and Insurance</strong></td>
<td>Consumer fraud</td>
<td>Anomaly and signature based</td>
<td>Type, frequency, time, amount and location of purchases (Credit card). Nature, amount and timing of claims and claims process used (Insurance)</td>
</tr>
<tr>
<td>(Bolton and Hand, 2002)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Ngai et al., 2011)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Telecom</strong></td>
<td>Consumer fraud</td>
<td>Anomaly and signature based</td>
<td>Geographical origin and destination of call, day of the week, time and duration of call</td>
</tr>
<tr>
<td>(Cortes and Pregibon, 2001)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Lopes et al., 2011)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Stock Trading</strong></td>
<td>Stock price manipulation and fraud</td>
<td>Anomaly and signature based</td>
<td>Related to stocks such as trading volume, trading venues, market capitalization, price, risk (Beta), financial news and filings, analyst opinions</td>
</tr>
<tr>
<td>(Diaz et al., 2011)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Kim and Sohn, 2012)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Manufacturing</strong></td>
<td>Equipment fault</td>
<td>Anomaly and signature based</td>
<td>Vibration data, lubricant chemistry, performance characteristics, temperature, acoustic emissions</td>
</tr>
<tr>
<td>(Raheja et al., 2006)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Jardine et al., 2006)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Project Management</strong></td>
<td>Project failure in terms of cost/time/scope</td>
<td>Anomaly based and Distribution</td>
<td>Scope, time schedule, planned and actual cost (Quantitative). Degree of openness/nature of communication, quality/tone of reports exchanged, time taken for decisions, extent of document and budget revisions and changes in personnel (Qualitative)</td>
</tr>
<tr>
<td>(Nikander and Eloranta, 2001)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Leu and Lin, 2008)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Healthcare</strong></td>
<td>Outbreak of disease</td>
<td>Primarily anomaly based</td>
<td>Extent of absenteeism from work, physician visits, laboratory tests, pharmacy sales, number of telephone calls to emergency services and nature of internet searches</td>
</tr>
<tr>
<td>(Unkel et al., 2012)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Sparks et al., 2011)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Computer Systems</strong></td>
<td>Unauthorized intrusions/system break-ins</td>
<td>Anomaly and signature based</td>
<td>Data trail left by users and system activity information like number of accesses, time of day, number of logins/failures, usage of memory, bytes sent</td>
</tr>
<tr>
<td>(Garcia-Teodoro et al., 2009)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Liao et al., 2013)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CONS: Consumption/sales rate  
AVCONS: Forecast of average consumption/sales rate  
ORATE: Order rate  
COMRATE: Production completion rate or rate at which orders are fulfilled and goods are received (for Distribution echelons)  
AINV: Actual inventory level  
DINV: Desired inventory level  
EINV: Error between desired and actual inventory level  
WIP: Work in process level or Orders in the pipeline (for Distribution echelons)  
DWIP: Desired work in process level  
EWIP: Error between desired and actual work in process level  
T_p: Delay or lead time (LT) in production or between placing orders and receiving goods (for Distribution echelons)  
T_i: Time to adjust inventory  
T_w: Time to adjust WIP  
T_a: Time to adjust demand

Figure 1. Block diagram for a single echelon APIOBPCS system
Figure 2: Normalised time series profiles of select variables for a simulated Beer game system under different capacity constraint settings. Beer game system configuration: Step change in market demand from 4 to 8 units/time unit; Delay/lead time profile of 2, 2,3 time units.
Figure 3: Spectra and their cluster profiles for shipping and net inventory variables for a simulated Beer game system under different capacity constraint settings. Beer game system configuration: Step change in market demand from 4 to 8 units/time unit; Delay/lead time profile of 2, 2, 3 time units.
Table 2: Signature for identifying capacity constrained distribution echelon in a supply chain

<table>
<thead>
<tr>
<th>Capacity* constrained echelon</th>
<th>Shipping cluster profile**</th>
<th>Net inventory cluster** profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retailer (R)</td>
<td>R shipping clustered separately from others</td>
<td>Net inventory for no echelon clustered separately or all clustered together</td>
</tr>
<tr>
<td>Wholesaler (W)</td>
<td>R and W shipping clustered together and separately from others</td>
<td>R net inventory clustered separately from others</td>
</tr>
<tr>
<td>Distributor (D)</td>
<td>R, W and D shipping clustered together and separately from others</td>
<td>R and W net inventory clustered together and separately from others</td>
</tr>
<tr>
<td>Echelon X</td>
<td>Echelon (X) shipping and shipping of all downstream echelons clustered together and separately from others</td>
<td>Net inventory of all downstream echelons (of X) clustered together and separately from others</td>
</tr>
</tbody>
</table>

*Shipping capacity ** Based on spectra of variables
Figure 4: Time series (normalised) of shipping and net inventory variables and cluster profiles of their spectra for a simulated Beer game system under different capacity constraint settings. Beer game system configuration: Market demand: Gaussian (Mean-4 units/time unit, Std dev-1 unit/time unit; Delay/lead time profile of 2, 2, 3 time units.

Appendix 1. Difference equations used to simulate the Beer game with capacity constrained distribution echelons

\( J = \) Supply chain echelon with \( J = 1, 2, 3, 4 \) representing the Retailer (R), Wholesaler (W), Distributor (D) and Factory (F) respectively

\[ \text{CONS}_J (J = 1 \text{ to } 4, \ t \leq 24) = 4 \quad \text{----- Initial condition; Rest of the equations are for } \ t > 24 \quad (2.1a) \]

\[ \text{CONS}_J (J = 1, \ t) = 4 \ (\text{constant demand}) \quad \text{or} \quad 8 \ (\text{step change in demand}) \quad \text{or} \quad 4 + \varepsilon_t \ (\text{Gaussian demand process}) \quad (2.1b) \]

\[ \text{CONS}_J (J = 2 \text{ to } 4, \ t) = \text{ORATE}_{J-1} (t - \text{LT}_{\text{order transmission}}) \quad (2.1c) \]

\[ \text{Shipment Received}_J (J = 1 \text{ to } 3, \ t) = \text{Shipment}_J+1 (t - \text{LT}_{\text{shipping}}) \quad (2.2a) \]

\[ \text{Shipment Received}_4 (J = 4, \ t) = \text{ORATE}_4 (t - \text{LT}_{\text{production}}) \quad (2.2b) \]

\[ \text{Material Available for Shipping}_J (J = 1 \text{ to } 4, \ t) = \text{AINV}_J (t-1) + \text{Shipment Received}_J (t) \quad (2.3) \]

\[ \text{Desired Shipping}_J (J = 1 \text{ to } 4, \ t) = \text{Backlog}_J (t-1) + \text{CONS}_J (t) \quad (2.4) \]

\[ \text{Maximum Shipping Capacity}_J (J=1 \text{ to } 3, \ t) = \text{MSC}_J \quad (2.5) \]

\[ \text{Shipping}_J (J = 1 \text{ to } 3, \ t) = \text{MIN} [\text{Desired Shipping}_J (t), \text{Material Available for Shipping}_J (t), \text{MSC}_J] \quad (2.6a) \]

\[ \text{Shipping}_4 (J = 4, \ t) = \text{MIN} [\text{Desired Shipping}_4 (t), \text{Material Available for Shipping}_4 (t)] \quad (2.6b) \]

\[ \text{AINV}_J (J = 1 \text{ to } 4, \ t) = \text{AINV}_J (t-1) + \text{Shipment Received}_J (t) - \text{Shipping}_J (t) \quad (2.7) \]

\[ \text{Backlog}_J (J = 1 \text{ to } 4, \ t) = \text{Backlog}_J (t-1) + \text{CONS}_J (t) - \text{Shipping}_J (t) \quad (2.8) \]

\[ \text{AVCONS}_J (J = 1 \text{ to } 4, \ t) = \text{AVCONS}_J (t-1) + \alpha [\text{CONS}_J (t) - \text{AVCONS}_J (t-1)], \text{where } \alpha = 1/(1+ T_a/\Delta t); \Delta t \text{ is simulation time increment set at 1} \quad (2.9) \]

\[ \text{DWIP}_J (J = 1 \text{ to } 4, \ t) = T_p \ast \text{AVCONS}_J (t) \text{ where } T_p \text{ is the lead time between placing an order and receiving the} \]

\[ \text{material for J supply chain echelon}; \text{T}_p (J = 1 \text{ to } 3) = \text{LT}_{\text{order transmission}} + \text{LT}_{\text{shipping}} - 1; \text{ T}_p (J = 4) = \text{LT}_{\text{production}} - 1 \quad (2.10) \]

\[ \text{WIP}_J (J = 1 \text{ to } 3, \ t) = \sum_{i=1}^{\text{LT}_{\text{order transmission}} - 1} \text{ORATE}_J (t-i) + \sum_{k=0}^{\text{LT}_{\text{shipping}} - 1} \text{Shipment}_J (t-k) + \text{Backlog}_J (t-1) \quad (2.11a) \]

\[ \text{WIP}_4 (J = 4, \ t) = \sum_{i=1}^{\text{LT}_{\text{order transmission}} - 1} \text{ORATE}_4 (t-i) \quad (2.11b) \]

\[ \text{EWIP}_J (J = 1 \text{ to } 4, \ t) = \text{DWIP}_J (t) - \text{WIP}_J (t) \quad (2.12) \]

\[ \text{DINV}_J (J = 1 \text{ to } 4, \ t) = 12 \quad (2.13) \]

\[ \text{EINV}_J (J = 1 \text{ to } 4, \ t) = \text{DINV}_J (t) - \text{AINV}_J (t) + \text{Backlog}_J (t) \quad (2.14) \]

\[ \text{ORATE}_J (J = 1 \text{ to } 4, \ t) = \text{MAX} [0, \text{AVCONS}_J (t) + (\text{EINV}_J (t)/T_i) + (\text{EWIP}_J (t)/T_w)] \quad (2.15) \]