INEFFICIENCIES IN IT PRODUCTION SUPPORT OPERATIONS: DETECTION AND IMPACT ANALYSIS

Complete Research

Puvvala, Abhinay, System Research Laboratory, Tata Research Development and Design Centre (TRDDC), Pune, India, abhinay.puvvala@tcs.com

Jha, Ashish Kumar, System Research Laboratory, TRDDC, Pune, India, ak.j@tcs.com

Rai, Veerendra Kumar, System Research Laboratory, TRDDC, Pune, India, veerendrak.rai@tcs.com

Vin, Harrick Mayank, System Research Laboratory, TRDDC, Pune, India, harrick.vin@tcs.com

Abstract

Complex systems carry inefficiencies within. While IT systems and automation have been used for long to remove inefficiencies in management, there has been relatively less effort to identify and remove the inefficiencies in the IT systems themselves and production support systems are no exception. In this paper, we use agent based approach to model a typical production support system and identify the inefficiencies present and corresponding impact of these inefficiencies. We focus on two particular instances of inefficiencies and impacts. First, the sub-optimal SLA compliance , wherein we detect the presence of bands where higher SLAs can be provided by the firms for little or no extra cost and hence the lower SLAs become, in essence, an inefficient point of operation. Second, the inefficient resource handling. Determining the number of resources (full time equivalents) to be added or removed from an ongoing production support service engagement is a complex problem and sub-optimal determination of the same impacts the firm's performance.

Keywords: Agent based modelling, IT Production Support, FTE Reduction, Inefficiency Reduction

1 Introduction

All complex systems have inefficiencies to some extent (McDermott, 1988). This holds true more for software systems which are expected to evolve and become more efficient over time (Dromey, 1995). The presence of inefficiencies is not just a characteristic of the software and its development process, but also of the software maintenance and management process. Maintenance of software encompasses handling incidents, which is "*an event that is not part of standard operation of a service and causes or may cause an interruption to or reduction in quality of that service*" (Kapella, 2003). Such maintenance contracts are highly prevalent in the domain of software services and often called as production support or production management engagements. The importance of efficient maintenance of software can be gauged from the fact that average maintenance cost of software over its complete life cycle can be as much as 90% of the total cost (Wiederhold, 2006; Singh et al., 2014). Various recent research efforts in this direction have concluded that not only are the maintenance costs increasing, but also estimation of maintenance costs are getting more difficult (Chawla et al., 2014; Kumari & Pushkar, 2013).

Parasuraman and Riley (1997) state that removal of inefficiencies is one of the primary reasons for automation or use of software. Execution of an IT production support engagement has remained managerial in nature. The managerial nature of execution and human interventions make the system prone to a higher amount of inefficiencies. While some of these inefficiencies may be a result of unavoidable management policies due to real world constraints as explained later in the paper, many others are present due to lack of analytical tools and techniques aimed at identifying and removing such inefficiencies from the system. With rising cost of production support services due to rising wages of skilled labour the need for detection and removal of such inefficiencies acquires paramount importance.

This paper presents a method for assessing the inefficiencies present in an engagement and the possible benefits that could be derived by its elimination. We have observed following two major instances of inefficiencies in a typical production support engagement.

- *Inefficiency in Effort-FTE reduction:* The number of FTEs is not reduced as a result of reduction in effort involved in providing production support.
- Efficiencies involved in non-identification of SLA bands wherein higher SLA (more stringent SLAs) can be attained at lower or no additional cost. This inefficiency disallows engagement managers to exploit optimized SLA bands which are extant in the engagement as a result of overall dynamics of engagement. These bands have not been created through conscious policy intervention.

Identification of such inefficiencies can help engagement managers derive higher value from the engagement for all stakeholders of the engagement. . The values include better quality of service at comparatively lower cost. Determination of overall impact of these inefficiencies would also help firms better plan the engagement contracts and resource allocation.

To achieve these goals, we use an Agent Based Modelling technique to model the engagement and study the impact of the inefficiencies described. Our work suggests that substantial amount of benefit can be derived from more careful planning and execution of projects to eliminate these inefficiencies. The agent based model as explained in subsequent sections helps assess the impact of inefficiencies present in the engagement and in certain cases suggest optimal and efficient operating point.

This paper is organized in sections. The next section summarizes the extant literature and establishes the current work in the framework of relevant literature. The research method employed in this paper and its literature has been explained in section 3. Section 4 elaborates on the model developed and employed for this study, elucidating the properties and behaviours of various agents. Section 5 presents the results and provides a discussion and analysis of the results and its inferences for the academia and practitioners. We conclude the paper in section 6.

2 Literature Review

ITIL (IT Infrastructure Library), the premier body for defining the best practices framework for IT services management refers to service management as “*a set of specialized organizational capabilities for providing value to customers in the form of services*” (Cannon et al., 2007, pp.273). Production support and Project service are the other terms which refer to the IT services management engagements in different contexts. Aspects of a production support engagement include ensuring the reliability and quality of the IT service being serviced. This might include the IT infrastructure maintenance or software maintenance. Cannon et al. (2007) states that a typical production support engagement should attempt to ensure maximum uptime as well as quality for serviced software during the tenure of the engagement. Analysis of IT services management from the service provider’s perspective is a relatively less researched area in the outsourcing and service management literature. The important aspect of identifying and determining the impact of inefficiencies in such a system draws from literature of economics, IT services and automation for theory and Agent Based Modelling for implementation.

The need for automation in varied sectors of management decision making has its roots in the efforts to remove or minimize the inefficiencies present in the system (Venkatraman, 1994; Bohn, 1998). The efforts throughout have been to figure out through automation or otherwise methods and techniques to remove inefficiencies in any system (Hammer, 1990). Most of the inefficiencies in any system are a result of human intervention. Human intervention either knowingly or due to the bounded rationality fails to capture the overabundance of information available to make a reliable and optimized decision and the decisions made are more often than not sub-optimal decisions (Venkatraman, 1994; DeHon and Wawrzynek, 1999). Even in more recent times in related fields there has been continuous effort to identify and remove inefficiencies in the system. Most of these efforts centre around using automated decision support systems which can capture and evaluate available options optimally (Rubino and Vitolla, 2014).

In IS research also identification of inefficiency and resultant automation to remove the same has been widely researched. Laudon & Laudon (2012) have specifically attributed the advent of advanced information systems management methods to identification and removal of inefficiency causing processes from several systems. These includes systems ranging from supply chain management in a lean way (Chen, Cheng & Huang, 2013) to an efficient energy management system (Corbett, 2013). Information systems theory has placed considerable focus on the theorization of the causes of such inefficiencies at a generic level and how to tackle the same. Schultze & Leidner's (2002) work on knowledge management (KM) in its exploration of the KM in IS research finds the importance of such methods and tools in removal of systemic inefficiencies. IS strategy of firms has revolved around the generalist presumption of presence of inefficiencies and its removal from the complex systems (Salmela & Spil, 2002; Rao, Mansingh & Osei-Bryson, 2012). Our study in this paper tries to tackle an interesting plot in this discourse taking cue from one of the most important IS based industries (IT production support services) of our times.

Even beyond the generalist presumption of presence of inefficiencies in most system and its removal with use of automation, the specific field of IT services has enough cases of presence of inefficiencies to be tackled. Wu (2006) found that one of the most important implications for increased use of IT services and software in a firm is to detect and eradicate the inherent inefficiencies in not only administrative segments of the firm, but in the IT services sector itself. Making the decision making processes optimized and alert to smallest changes in context of the business is one of the most important aspects of making firm lean (Waterhouse, 2008). Most of these inefficiencies in determining the optimal point result from human inability to comprehend the complex web of information which is required to reach optimal decision point. The dynamic environmental or contextual information makes this decision making process complex which can be then resolved using autonomic systems (Wang et al., 2004). Although, these inefficiencies are present in most complex systems, its presence and mitigation in systems as complex as IT production support has not been sufficiently studied. IT and automation tools have been used to detect and mitigate inefficiencies in other areas but not in the service of managing IT itself which causes severe losses to the industry.

One of the most important aspects of an IT production support engagement is SLA compliance. It includes factors such as the duration of response and resolution time of incidents and the minimum percentage of incidents to be resolved within pre-specified SLAs. SLAs have been getting more and more stringent over past decade and a half and this has put the IT service providers under tremendous pressure (Lezama et al., 2013; Liu et al., 2001). Hence, determination of the highest optimal SLA that can be serviced easily at minimal additional costs can add tremendous value to the IT service providers (Bi, 2007). Most of the times in the industrial situations, multiple vendors cater to different aspects of IT services for a single client firm and in such complex situations knowing the highest serviceable SLA at lowest cost is a huge challenge for vendors (Bhattacharya, 2006). With the complete ecology of the IT services management being under tremendous scrutiny and analysis (Reiter et al., 2013), it is of paramount importance, both for academia and industry, to be able to detect and mitigate such inefficiencies leading to sub-optimal SLA compliance.

Another important aspect of inefficiencies in IT production support engagement is the number and type of resources to be employed. Overstaffing runs the risk of resource underutilization and understaffing would lead to failure in SLA compliance. Both of these issues are manifestations of inefficiencies in the in IT production support system. Such inefficiencies would have serious impact on financial performance of the firm and reflects poorly on its organizational capabilities (Aral and Weill, 2007). The number of resources to be deployed in any IT services engagement is directly impacted by various contextual parameters like the maturity of the engagement, skills required, incident volume etc. Automation and sudden changes in contract terms might force the required number of resources to be re-optimized (Shang and Seddon, 2002). The number of additional resources to be hired or resources to be reduced is a tricky exercise and the contextual parameters play an important part in arriving at optimal solution. The number and type (in terms of skills) of resources required in any production support engagement is not a direct or proportional figure of the quantum of work (Kien et al., 2013), but is in fact dependent on multiple factors as explained in sections below.

The extant literature surveyed, indicates that inefficiencies in IT production support engagement is a logical product of the complexity of the system, and its continued presence is detrimental to firms' performance and its capabilities.

3 Research Method: Agent Based Modelling

To enable analysis of an engagement from a systems perspective by analysing not only the engagement alone but also its contextual parameters, we use Agent Based Modelling (ABM) as a research method in this work. Agent based methods are considered as one of the premier tools to simulate and study the interactions involving human systems (Bonabeau, 2002; Davidson, 2002). Production support system is one such system. Authors could identify various agents in the incidents and the resources who individually interact with each other under defined processes which are a result of the policies. The inefficiencies as stated above are results of such complex interactions and agent based method provide a reliable method to study them (Axelrod and Tesfatsion, 2006; Mataric, 1993).

ABM approach studies interaction of agents in a given environmental context under defined rules. These interactions are analysed by simulation of the agents' behaviour. It is a relatively new and emerging method in social sciences, which can be applied to a problem by defining a set of agents with related attributes, behaviours and fitness function, the simulation environment and the overall performance-measuring objectives of the environment. Agents and interactions between them are two most important things in Agent Based Modelling (Axelrod and Tesfatsion, 2006).

A typical ABM model consists of an agent having certain attributes, rules/actions, goals and decisions to make. These defined agents are generally governed by a fitness function. The aim of creating a fitness function is that it allows multiple agents of similar nature to have different attributes by creating differences in parameters of fitness function. This heterogeneity thus created is an essential component of ABM and helps mimic the real world more closely than other methods. These countless interactions lead to 'emergence' of new behaviour which had not been programmed into the behaviour of the individual agents (Waldrop, 1992). Agent based modelling has already been extensively used in economics (Agent Based Computational Economics (ACE)). Zaffar *et al*. (2008) used it to identify the impact of Variability of Open Source Software (OSS) support costs, length of upgrade cycle and interoperability costs on OSS diffusion.

ABM is an important tool which has been widely used for theory development and is considered an effective and reliable method for robust theory development. Newell and Simon (1972) established that if a particular instance of Agent based model A gives result R, then the sufficiency of theorem "R if A" has been established for deterministic models.

In view of these advantages offered by the method, we are using ABM for the current work to model the production support engagement and see the implications and impact of the inefficiencies present in the system. The results derived by this method are fairly generalizable to any similar context.

4 Research Model

The research model in this paper represents a typical production support engagement between a typical large IT service provider and any global client. In this engagement model we have considered two technical towers. An engagement model, consisting of two technical towers each having employees of three skill levels operating in 3 shifts, represents a typical engagement model without any loss of generality.

The various facets of the engagement model, its entities and terms are explained below:

- Technical Towers: A technical tower represents a method of work organization usually employed in IT production support where issues with roots in different technical domains are grouped together and labelled a technical tower. As an example a technical tower may be "Dot net", "Java", and "VM" etc. An engagement may have multiple towers, our research model comprises of two towers.

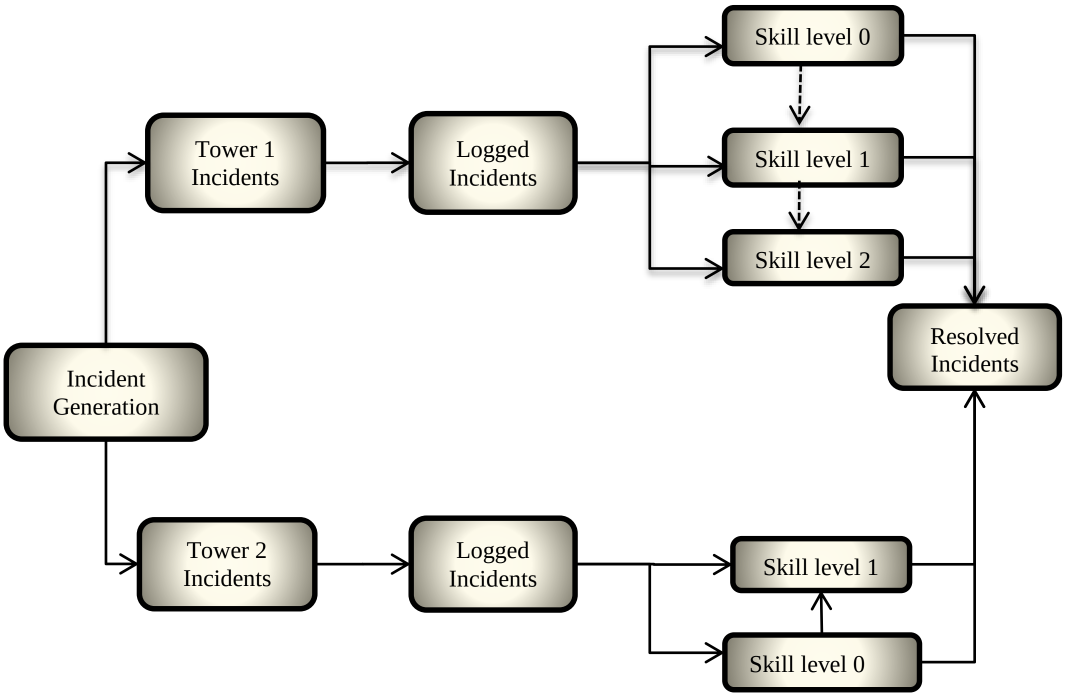

Figure 1. Research Model of the IT services engagement.

- Incident: - As mentioned earlier International Network Services' Incident and Problem Management Framework (Kapella, 2003) defines incident as "*an event that is not part of standard operation of a service and causes or may cause an interruption to or reduction in quality of that service*". These incidents must be resolved at the earliest to ensure production support stability. An incident may belong to different categories depending on the classification style followed. As explained above, in the current model the incidents are characterized according to the technical towers to which they belong.

    a. Tower 1 Incidents: - These are the incidents which belong to classification of tower 1.

    b. Tower 2 incidents: - These are the incidents which belong to classification of tower 2.

- Resources: - The most important resources in any IT production support engagement are the employees who work on incidents to resolve them. These are the resources which have been modelled for this research. Resources belong to different technical towers according to the skill type they possess. In any given tower a resources may have different levels of expertise in that skill.

The resources may have three levels of skills where skill level 0 is the most elementary and skill level 2 is most advanced knowledge of that skill. Since the service to the client has to be provided continuously for 24 X 7, there are 3 sets of resources working in 3 shifts of 8 hrs each.

Using the above entities as defined we have created an Agent Based Model with following agents, interactions and the attributes.

Agents

- Resources: - Resources have been treated as an agent in the model with following attributes as described in table 1 below.

Attribute	Description
Resource ID	Unique identifier for each Resource
Resource Tower	The tower to which the resource belongs specifying the skillset which the resource possesses in terms of Tower 1 or Tower 2
Resource Skill Level	Skill level of the resource i.e. Skill level 0, Skill level 1 or Skill level 2
Effort Required	The effort in terms of time required to resolve/respond to the incident on which the resource is currently working
Incident ID	The Incident which the resource is working on
Shift of resource	The shift of the resource
Resource Cost	Cost of this particular resource

Table 1. Attributes of agent "resource" and their definition

- Incidents: - Incidents are the central agent in the model. Table 2 describes all the attributes and the description for the incident agent.

Attribute	Description
Incident ID	Unique identifier for each Incident
Priority	Priority of the Incident. It can range be one of Critical, High, Medium and Low
Incident Type	Incident's Technical Tower i.e. Tower 1 or Tower 2
Effort Required	It is the maximum effort required to resolve the Incident
SLA Time	Time within which the Incident needs to be resolved to fulfil SLA obligation
Generated Time	The time at which the Incident was generated by the client
Logged Time	Time when the Incident was logged for resolution
Assigned Time	Time at which the Incident was assigned to the resource working on the Incident
Resolved Time	Time at which the Incident was fully resolved
Resolution Time	Time difference between Incident generation and Incident resolution
Resolved ID	ID of the resource who worked on the Incident
SLA compliance	Whether the resolution time was lower than SLA time or not

Table 2. Attributes of agent "incidents" and their definition

Policies for Interaction of Agents

One of the most important features of a typical Agent Based Model is its ability to simulate and produce the outcome under given policies governing the interaction environment. In case of our research model the policies determine the behaviour of the agents as explained:

- Generation of incidents: - Based on our analysis of various engagements data we found that incidents are generated according to a Poisson distribution of a given mean which varies from one engagement to other. The incidents are distributed between the various priority levels and also between different technical towers.
- Logging of incidents: An incident once generated at client's end is first logged at the service provider's end. The logging process consists of following steps:
  - Identification of the type of incident based on its tower.
  - Identification of priority level of the incident based on the severity and impact of the issue caused due to it.
  - Acknowledging the receipt of incident and setting up a maximum time for resolution adhering to the SLA.
  - Assigning the incident for resolution to a specific resource based on resolution policies.
- Resolution of Incident – Incidents can be allotted to the resources based on various policies employed by the engagement manager. A typical policy for handling incidents employed in this model is explained below:
  - Assigning "low priority" and "medium priority" ticket to skill level 0 employee first
  - Assigning "high priority" ticket to skill level 1 employee directly
  - Assigning "Critical priority" ticket directly to skill level 2 employee
  - Some of the problems which cannot be resolved at skill level 0 and skill level 1 are escalated to skill level 2 employees for resolution.

5 Analysis and Results

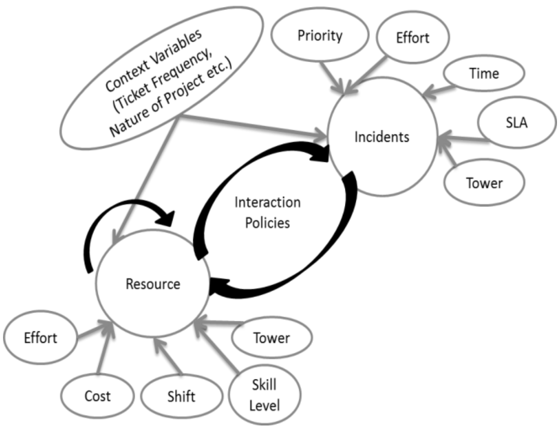

Figure 2 Visual representation of the agents' interaction

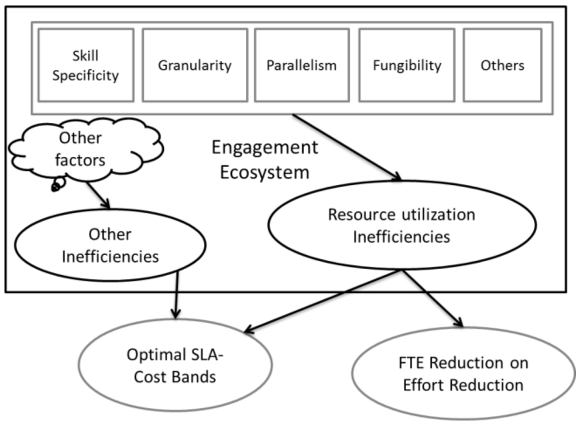

Figure 3 Engagement black box and the inefficiencies

As mentioned in section 1 the aim of the current work is to identify the inefficiencies in a typical production support engagement and estimate the impact of the same. For that purpose we simulated the specified agent based model for a period of about a month. Figure 2 shows the agent based model's interaction schema where the circles are the agents with their attributes in ovals. The thick black arrows show the interaction which is defined by the policies and contextual factors. Resources interact with themselves hence self-interacting loop on resources. In the following sections we discuss the twin inefficiencies and their impacts.

Figure 3 shows the ecosystem analysed in the current work. Multiple factors like skill specificity, granularities etc. are responsible for inefficiencies in resource utilization. This and other types of inefficiencies have considerable impact on the engagement. We focus on the impact of inefficiencies due to resource utilization constraints in our current work, treating the engagement as a black box.

5.1 Presence of SLA-cost bands

Determination of the optimal SLA for an engagement is very crucial for both the client as well as the vendor. For the client, the general objective is to achieve the highest SLA compliance possible at minimum cost and for the vendor it is to ensure that the SLA compliance levels are consistently met at minimum costs. As discussed in previous sections, determination of optimal SLA for an IT service engagement is a difficult, but critical task.

The production support engagement as a system has systemic issues which lead to service levels LAs being sub-optimal. The service levels chosen for the engagement may be suboptimal either due to the fact that higher service levels could be offered at similar or slightly higher cost or that slight decrease in service levels might lead to huge savings for the parties involved.

One of the most important sources of such inefficiency is inefficient utilization of resources. The following example explains a typical scenario and its implication.

Let's suppose that the current SLA requirement is 90% in an engagement which forces the vendor to deploy 5 resources (employees/ FTE) and resource utilization of is about 85%. For this engagement, if the requirement for SLA increases to 92%, vendor would have to hire one more resource raising the costs, but since one resource can do much more work, and the current resource distribution is sub-optimal, resource utilization goes down to 65%. Vendor can, in this case, easily increase the SLA compliance to 94% without any significant additional cost as the resources currently are underutilized and can use the same resources to offer higher levels of service.

The above example is not exhaustive list of the sources of inefficiencies or its impact on determination of optimal SLA. However, in this work we are attempting to determine the impact of such systemic inefficiencies. We would not attempt to state the exhaustive list of all sources of such inefficiencies.

The contextual parameters of the system, which determine the environment in which the engagement is executed has important bearing on the outcome of any such analysis. Following table shows the contextual parameters for the current analysis.

- Incident Generation: Incidents per hour were drawn from a random Poisson distribution of mean 40
- Escalation of tickets from Skill level 0 to 1 and from skill level 1 to 2 : 10%
- Distribution of incidents: Tower 1 and Tower 2, 50% each. According to priority, the distribution was Critical- 15%, High- 25%, Medium- 40% and Low- 20%.
- SLA Time duration as per priority

Priority	Resolution SLA time duration (in Mins.)
Critical	40
High	60

Medium	90
Low	150

Table 3. The resolution time for incidents as per the SLA

The following table contains the cost per hour of the resources deployed across towers.

Technical Tower	Resource Type	Case 1: Cost	Case 2: Cost
Technical Tower 1	Skill level 1	15	20
	Skill level 2	45	60
	Skill level 3	75	150
Technical Tower 2	Skill level 1	30	20
	Skill level 2	60	80

Table 4. The cost of the resources in the two cases as simulated

To evaluate the impact of systemic inefficiencies on SLA and related costs, we test the model with different distributions of employees in each shift and skill level. The aim was to determine the realizable SLA levels and associated costs for any given distribution of employees. To make the results more generalizable and rigorous, we simulated the model for 1000 times with different employee distributions.

The analysis was performed with two different set of costs of resources for two extreme cases. Figure 4 and figure 5 show the output of the two cases. Table 4 shows the two sets of costs of resources used for the simulation. The graphs in figure 4 and 5 represent the costs and utilization of resources at various SLA compliance levels.

For the same volume of incidents (as defined in the contextual parameters) being generated, different resource configuration at different skill levels might mean different amount of incidents being resolved successfully. Our simulation covers a wide spectrum of cost and utilization for SLA compliances ranging from 38% to 98%. Although for all practical purposes the SLA compliance is somewhere between 90-98% only, the complete spectrum has been modelled to better understand the dynamics of the engagement.

It is intuitively apparent that for any engagement higher SLA compliance can be achieved with higher amount of resources to resolve the incidents in pre-specified times. However, increased number of resources would also cost more. The results as shown in both figure 4 and 5echo these results at a broader scale. But at a more granular view the impact of inefficiencies as explained with the example above starts getting visible. There are certain bands in between the overall increasing curve of cost of resources where the graph for small duration remains stagnant or becomes decreasing.

As an example in figure 4, point B shows a position where the cost is approx. 395,000 and point C shows a position where the cost is again approx. 395,000 but the SLA compliance that can be achieved at almost similar cost levels varies from 87% at point 1 to 92% at point 2. Also another example which is visible in figure 4 is the difference between A and B. Point A shows a position where the SLA compliance level of 86% is met at cost of approx. 345,000 whereas SLA level of 87% requires cost of 395,000. This shows the sub-optimality of operating at position like 87% SLA compliance with the cost structure and contextual parameters as assumed for model showing figure 4.

These bands where increasing SLA compliances have little or no additional cost implications are an important feature of all such engagements. It is crucial to identify such bands and, to the extent possible, keep the engagement at the maximum (rightmost end) of such a band. Operating at any other position within the band is sub-optimal

The band identified between B and C in figure 4 is one of the bands. Multiple bands, may exist in a single project, some large and some small. Another such band has been identified in figure 4 between D and E. Some might be relevant like the identified band while some other might be not so relevant if the band is somewhere in the range of 40-60% as identified by band D-E.

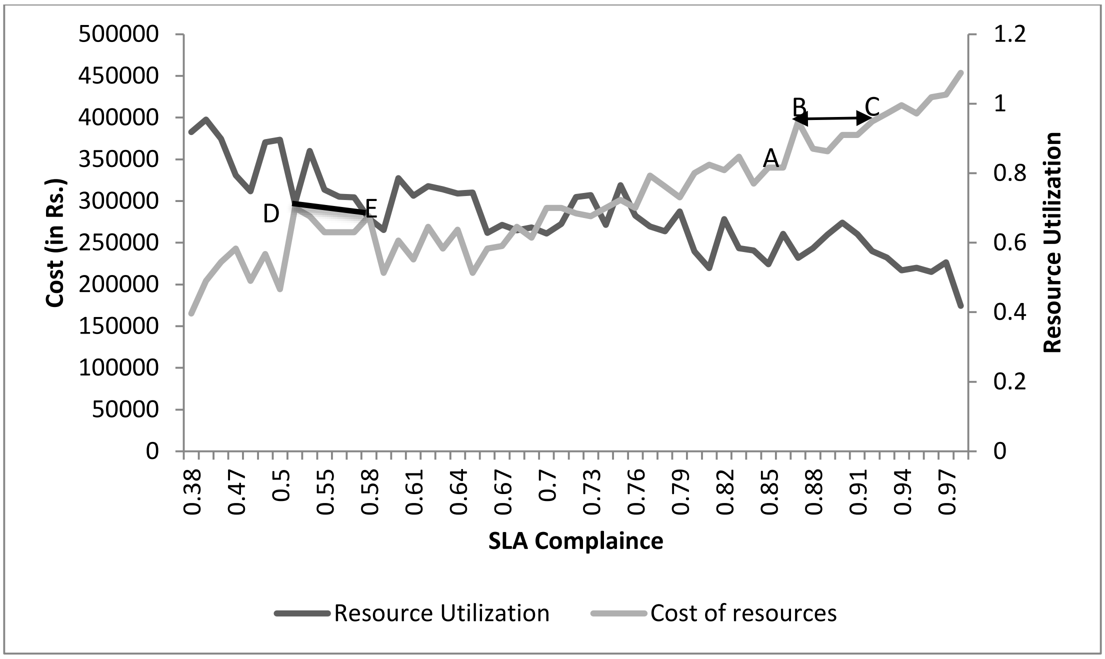

Figure 4. Resource utilization and cost of resources plotted against SLA compliances for case 1 model

Another aspect of such SLA-cost bands is these bands are context dependent. The band identified in figure 4 is not a universal fact. For an engagement with different resource costs and ticket distributions, different band may exist. E.g. Figure 5 shows the presence of such a band between the limits of 89% to 94% of SLA compliance at cost below 606,900. Similarly, there exists an ‘A’ in figure 5 also which indicates the cost of meeting 85% SLA compliance as 516,000. This shows the high investment required to reach just 4% higher SLA compliance or in other words it also means that slightly lower SLA compliances might lead to huge savings for the firms.

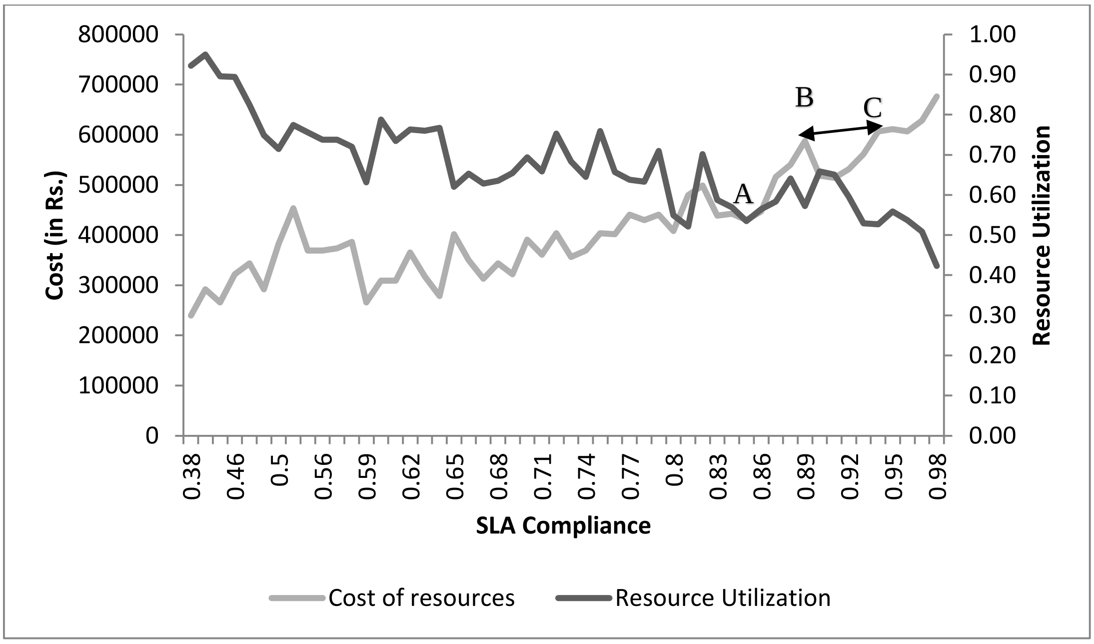

Figure 5. Resource utilization and cost of resources plotted against SLA compliances for case 2 model

These bands are representative of the impact that systemic inefficiencies may lead to in a system. Where common wisdom suggests higher costs for higher SLA compliances, these issues enable higher SLA compliance at almost similar cost levels. The identification of such inefficiencies can allow the vendor (IT service provider) to provide high levels of service at little or no extra cost. This would be beneficial for both the clients as well as the vendors.

5.2 Determination of optimal FTE reduction

The other inefficiencies discussed in this paper is determination of resource (full time equivalent) reduction. In any typical engagement reduction in required efforts due to various reasons like automation, reduction of scope of the project etc., results in reduction of number of resources required. However, the reduction of the resources is not linear i.e. reduction of 8 man hours (total number of hours a resource is supposed to work) of work each day does not result in reduction of 1 Full Time Equivalent (FTE). Some of the important reasons for this are:

- **Fungibility:** Fungibility here means FTEs (resources) are free to move across levels of support to take up the tasks across levels. A non-fungible structure leads to formation of silos where resource in a silo cannot come to another silo where the workload is higher at a given instant and more resources are required. A fungible structure in production service engagements allows more equitable distribution of work amongst resources leading to higher resource utilization and lower waiting time for issues to be resolved.
- **Task Granularity:** Granular tasks can be easily divided amongst employees and lower employee count can achieve higher utilization. However, if the tasks are large and non-granular and they cannot be distributed and only one resource can work on that task at a time while others would be idle. This would lead to lower resource utilization. However, unmindful and abrupt removal of resources from the engagement could be risky. Production support environment is volatile and it is quite possible that incident volume goes up suddenly thereby creating a situation wherein there are not enough resources to handle the incidents and production support stability suffers due to inefficient handling of incidents. Such large tasks are more likely to be found in problem management and change management aspect of production support rather than in incident management.
- **Skill Specificity**: Reduction in effort requirement might not lead to reduction in requirement of resources as each resource will have a specific skill which might not be done away with. Every production support service engagement needs a set of skills and the set of resources deployed in the engagement must cover those skills. So, must be a minimal number of resources which cover all the skills required for the engagement. However, all resources are not available all the time. They take leave and are engaged in training and therefore net availability is much less than perceived availability.
- **Parallelism**: Inability to do tasks in parallel is another systemic inefficiency that might be present which leads to lower than expected employee reduction. This problem can be described as a situation where the tasks are structured so that one task can be done only after the other finishes. This situation leads to waste of resources' time where the free resource has to wait for some other task and this leads to inefficiencies.

The factors mentioned above are some of the major factors and not an exhaustive list as we have concerned ourselves mainly to determination of impact of such inefficiencies. The impact of such inefficiencies is again context dependent. The amount of Full Time Employees (FTEs) that can be reduced for any given reduction in effort is dependent on the way the engagement is structured and its underlying factors including the three factors mentioned above.

To detect the true impact of such inefficiencies and the optimal reduction that is possible in an engagement we have used the ABM explained in previous sections. The agent based method can be utilized to detect the maximum inefficiency that can be removed and provides the maximum number of employees that can be reduced for a given engagement and engagement context. The context variables for this simulation are same as given in section 5.1 above. To gather sufficient insights, we ran the simulation for 1 month and after 15 days reduced the volume of tickets, to signify effort reduction.

Pre Reduction Mean -40/hour Post Reduction Mean -20/hour
(Poisson Distributed)
Ticket Volume -50%

0 15 days 30 days

Shift	Resource Type	Pre Reduction	Post Reduction
1	Skill 0, Tower 1	3	5
	Skill 1, Tower 1	1	1
	Skill 2, Tower 1	4	2
	Skill 0, Tower 2	2	1
	Skill 1, Tower 2	3	2
2	Skill 0, Tower 1	3	4
	Skill 1, Tower 1	1	1
	Skill 2, Tower 1	3	2
	Skill 0, Tower 2	2	1
	Skill 1, Tower 2	5	2
3	Skill 0, Tower 1	3	3
	Skill 1, Tower 1	1	1
	Skill 2, Tower 1	3	2
	Skill 0, Tower 2	2	1
	Skill 1, Tower 2	4	2

Table 5. Results of analysis of reduction of effort required on employee required

		Pre Reduction	Post Reduction
Effort (person-hours/day)		243	120(123↓)
FTE Reduction	**Expected**	123/8 ~ 15 FTEs	
	Actual	10 FTEs	
Average Resource Utilization		78.59%	61.64%
SLA Compliance		95.24%	95.21%
Costs		307200	194400 (36.71%)

Table 6. Comparison of pre and post reduction scenarios in case of effort reduction

Table 5 and 6 show the simulation results of this engagement. *The reduction of incidents by 50% should typically lead to reduction of employees by half which is not the case.* The most efficient allocation of employees is given in the last column of table 5. Table 6 analyses the impact of the reduction and the possible inefficiency present in the system.

As evident in analysis shown in table 6, 50% reduction in effort should lead to reduction of 15 resources on a theoretical basis but actual simulation results show that it should be a reduction of only 10 resources. Reduction of 15 resources in this situation would make the engagement difficult to execute at the agreed SLA levels.

6 Conclusion

The analysis presented in this section using ABM strengthens the initial hypothesis that there are considerable inefficiencies present in a system as complex as IT production support. The paper extends the literature of the field by using a simulation based study to analyse the impact of inefficiencies that are present in the system. One of the benefits of recognizing the presence of such inefficiencies is that projects can accomplish optimal SLA compliance levels at minimum costs in a given SLA band. This would be beneficial to the entire engagement ecosystem including vendor and client. Our research also highlights the context dependent nature of the results. Change in the context of the engagement can considerably change the optimal points, SLA bands etc.

The work presented in this paper can be taken further by identifying all possible inefficiencies in IT production support engagements and analysing the root causes of these inefficiencies. Further studies also need to isolate and identify the impact of these causes individually. Determination of exact extent of inefficiencies which can be attributed to various factors described above like fungibility, granularity etc. would greatly enhance the current research.

References

Aral, S. and Weill, P. (2007). " IT assets, organizational capabilities, and firm performance: How resource allocations and organizational differences explain performance variation". *Organization Science*, 18(5), 763-780.

Axelrod, R., Tesfatsion, L. (2006). Appendix AA Guide for Newcomers to Agent-Based Modelling in the Social Sciences. Handbook of computational economics, 2, 1647-1659.

Barquet, A. L., Tchernykh, A. and Yahyapour, R. (2013). "Performance Evaluation of Infrastructure as Service Clouds with SLA Constraints". *Computación y Sistemas*, 17(3), 401-411.

Bhattacharyya, N. (2006). "Operational Level Agreements and Cross Vendor Cooperation". *Enterprise IT perspectives*, 21.

Bi, L. (2007). "Managing the risks of IT outsourcing". *Journal of Corporate Accounting & Finance*, 18(5), 65-69.

Bohn, R. E. (1998). "Measuring and managing technological knowledge". *The Economic Impact of knowledge*, Butterworth-Heinemann, Boston, 295-314.

Bonabeau, Eric. (2002) "Agent-based modelling: Methods and techniques for simulating human systems." *Proceedings of the National Academy of Sciences of the United States of America 99.*: 7280-7287.

Cannon, D., Wheeldon, D. and Sharon, T. (2007). *ITIL.[4]. Service operation*. TSO (The Stationery Office).

Chawla, R., Ahlawat, D., and Kumar, M. (2014). "Software Development Effort Estimation Techniques: A Review." *International Journal of Electronics Communication and Computer Engineering. 5(*5).,

Chen, J. C., Cheng, C. H., and Huang, P. B. (2013). Supply chain management with lean production and RFID application: A case study. *Expert systems with applications, 40*(9), 3389-3397.

Corbett, J. (2013). Using information systems to improve energy efficiency: Do smart meters make a difference?. *Information Systems Frontiers, 15*(5), 747-760.

Davidsson, P. (2002). "Agent based social simulation: A computer science view". *Journal of artificial societies and social simulation*, 5(1).

DeHon, A., and Wawrzynek, J. (1999). "Reconfigurable computing: what, why, and implications for design automation". In: *Proceedings of the 36th annual ACM/IEEE Design Automation Conference* (pp. 610-615). ACM.

Dromey, R. G. (1995). "A model for software product quality." *Software Engineering, IEEE Transactions on*, 21(2), 146-162.

Hammer, M. (1990). "Reengineering work: don't automate, obliterate". *Harvard business review*, 68(4), 104-112.

Kapella, V. 2003. *A framework for incident and problem management*. International Network Services.

Kien, S. S., Soh, C., and Markus, M. L. (2013). "A New IT Organizational form for Multinational Enterprises". In" *Proceedings Pacific Asia Conference of Information Systems 2013*

Kumari, S. and Pushkar, S. (2013) "Performance Analysis of the Software Cost Estimation Methods: A Review*", International Journal of Advanced Research in Computer Science and Software Engineering.* 3(7). (pp. 229-238).

Laudon, K. C., and Laudon, J. P. (2012). *Management information systems: managing the digital firm* (Vol. 12). Pearson.

Liu, Z., Squillante, M. S. and Wolf, J. L. (2001). "On maximizing service-level-agreement profits". *In: Proceedings of the 3rd ACM conference on Electronic Commerce* (pp. 213-223). ACM.

Mataric, M. J. (1993). "Designing emergent behaviors: From local interactions to collective intelligence". In: *Proceedings of the Second International Conference on Simulation of Adaptive Behavior* (pp. 432-441).

McDermott, J. (1988). "Preliminary steps toward a taxonomy of problem-solving methods". In: *Automating knowledge acquisition for expert systems* (pp. 225-256). Springer US.

Newell, A. and Simon, A. H. (1972). *Human problem solving*. NJ: Englewood Cliffs.

Parasuraman, R. and Riley, V. (1997). "Humans and automation: Use, misuse, disuse, abuse". *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 39(2), 230-253.

Rao, L., Mansingh, G., and Osei-Bryson, K. M. (2012). Building ontology based knowledge maps to assist business process re-engineering. *Decision Support Systems*, *52*(3), 577-589.

Reiter, M., Fettke, P., and Loos, P. (2013). "Towards a reference model for ecological it service management". In: *Proceedings of International Conference of Information Systems (ICIS)*

Rubino, M. and Vitolla, F. (2014). "Corporate governance and the information system. How a framework for IT governance supports ERM". *Corporate Governance*, 14(3), 3-3.

Salmela, H., and Spil, T. A. (2002). Dynamic and emergent information systems strategy formulation and implementation. *International Journal of Information Management*, *22*(6), pp. 441-460.

Sattarova, F. Y., Sattarov Shavkat, Y. and Furkhat, T. (2009). "Model of business service management with security approach". *International Journal of Advanced Science and Technology*, (9), 39-50.

Schultze, U., and Leidner, D. E. (2002). Studying knowledge management in information systems research: discourses and theoretical assumptions. *MIS quarterly*, (pp. 213-242).

Shang, S., and Seddon, P. B. (2002). "Assessing and managing the benefits of enterprise systems: the business manager's perspective". *Information Systems Journal*, 12(4), 271-299.

Singh, B. K., Punhani, A., Singh, G. P., and Misra, A. K. (2014). Effect of Variations in Measurement Process for Software Development Efforts. In *Multimedia, Computer Graphics and Broadcasting (MulGraB), 2014 6th International Conference on* (pp. 3-8). IEEE.

Venkatraman, N. (1994). "IT-enabled business transformation: from automation to business scope redefinition". *Sloan management review*, 35, 73-73.

Waldrop, M. M. (1992). *Complexity: The Emerging Science at the Edge of Order and Chaos*. New York, NY: Touchstone.

Wang, J., Cao, J., Leckie, J. O., and Zhang, S. (2004). “Managing e-government IT infrastructure: an approach combining autonomic computing and awareness based collaboration”. In: *Computer and Information Technology, 2004. CIT'04. The Fourth International Conference on* (pp. 998-1003). IEEE.

Waterhouse, P. (2008). “Improving IT Economics: Thinking ‘Lean’”. *White paper*, CA.

Wiederhold, G. (2006). “What is your software worth?” *Communications of the ACM*, (49:9), September, pp. 65-75.

Wu, D. (2006). “Detecting information technology impact on firm performance using DEA and decision tree”. *International journal of information technology and management*, 5(2), 162-174.

Zaffar, M. A., Kumar, R. L. and Zhao, K. (2008). “Diffusion Dynamics of Open-Source Software in the Presence of Upgrades: An Agent-Based Computational Economics (ACE) Approach”, In: *Twenty Ninth International Conference on Information Systems (ICIS 2008)* Paris: Paper.55.