IDEA GROUP PUBLISHING

701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com

ITP4896

A Case of Serial Handoff: Production Problems with a Strategic Mission Critical System

James White
DePaul University, School of Computer Science, Telecommunications, and Information Systems, 243 South Wabash Ave., Chicago IL, 60604, USA, jwhite@cit.depaul.edu

Linda V. Knight
DePaul University, School of Computer Science, Telecommunications, and Information Systems, 243 South Wabash Ave., Chicago IL, 60604, USA, lknight@cti.depaul.edu

ABSTRACT

This case describes a calamitous production failure, as it occurred at a major financial services firm. As a consequence of this failure, all North American distribution of trading information for a major London-based financial services firm via its U.S. partner's network was disrupted for approximately 10 hours. In addition to the financial effect on the firm and its customers, the handling of this event by the operations personnel in both organizations strained the newly founded partnership and uncovered several serious procedural and organizational deficiencies. This paper consists of a description of the critical incident followed by analysis and conclusions concerning the causes of this specific production error, and then by a consideration of the broader issues involved in studying the causes of production problems in strategic or mission critical systems. Conclusions are drawn concerning the applicability of this research to practitioners and future researchers.

METHODOLOGY

Data for this study was collected using the Critical Incident Technique (Flanagan, 1954). The Critical Incident Technique is a way to obtain a record of specific behaviors from those in the best position to make the necessary observations and evaluations, and was important to this study because it emphasized direct observation of specific situations versus generalization based on opinions, hunches, and estimates. The Critical Incident Technique also is well suited when the purpose of the research is to increase knowledge of a real-world phenomenon about which relatively little has been documented (Bitner et al., 1990). Analysis of the data was conducted within the framework of the National Transportation Safety Board (NTSB) accident investigation model (White, 2003). The NTSB model describes accidents in terms of direct events (chain of events/mechanisms) and contributing factors (conditions) that may arise from systemic factors (constraints).

BACKGROUND

Company Alpha (a pseudonym) is a leading financial services firm headquartered in New York City and servicing a mix of international and domestic customers. Company Beta (a pseudonym) is a leading financial services company located in London, U.K. serving U.K., European, Asian, and North American customers. Former competitors, Company Alpha and Company Beta formed an alliance about 9 months ago to reduce operating costs and improve service to respective international clients. A key element in this collaboration was distribution of each other's price information within local geographic regions. Company Alpha, for example, receives price information directly from Company Beta and redistributes the Beta information throughout North America through its existing network of wholesale distributors. This effectively enhances Beta's market distribution in North American without significantly increasing its distribution costs. A similar benefit is accrued by Company Alpha through expanded distribution of price information via Beta's existing distribution channels.

CRITICAL INCIDENT

In the case description that follows, certain details have been expunged and actor and organization names have been changed to preserve anonymity of organizations and individuals.

On Saturday, March 4, a failure occurred that prevented distribution of all information for Company Beta, via Company Alpha's information distribution network for the entire trading session beginning Sunday night, March 4, through 6:30 a.m. Monday morning, March 5. All dates and times are shown as local time - New York City.

On Saturday, March 4, Mr. Frank McNulty, a consultant in Company Alpha's Information Technology Department, implemented changes to the price information system designed to relieve sporadic distribution slowdowns that occurred on Friday. The consultant correctly implemented the changes but unintentionally changed the configuration of the *ports* that control information distribution. Prior to February 11, information for Company Alpha and Company Beta distributed via the information network was consolidated and distributed through one port (port A, for example). On February 11, the system was changed to distribute Company Alpha information through port-A, and Company Beta information through a separate port (port B, for example). On Saturday, March 4, the consultant unknowingly changed the port configurations back to the pre-February 11 condition: Company Alpha information and Company Beta information were both distributed via port-A and no information was distributed via port-B (Table 1). Information vendors were expecting only Alpha information through port A and only Beta information through port B. Common practice among vendors is to discard unexpected data.

The first evidence of missing Beta information was reported at 8:00 p.m. when Beta Market Supervision received a call from a 3rd party information vendor who said it was not receiving any Beta information. Beta Market Supervision called Beta Operations who said there was no problem with the Beta system and advised Beta Market Supervision to call Company Delta (a pseudonym). Company Delta is a New York City-based service bureau that processes data for both Alpha and Beta. Around this time, a Beta spokesperson reported to the press that no Beta information was being distributed due to a Company Delta problem. Later, a Beta spokesperson said that the problem was related to Alpha information distribution, not Delta.

At 9:00 p.m., the Alpha Technical Operations Control Center received calls from two vendors, who said that they were not receiving Beta information. Mr. John Andrews checked his system displays, which showed that Beta information was being sent to vendors, and transferred

Table 1: Details on data being passed through ports

Table 1: Details on data being passed through ports		
Port Configuration		
Period	**Port A**	**Port B**
Pre-February 11	Alpha information and Beta information	Not used
Post-February 11	Alpha information	Beta information
March 4 – March 5, as a result of the consultant's unintended change	Alpha information and Beta information	Not used

these calls to Beta Market Supervision. The Beta information that the Alpha Technical Operations Control Center saw on the system display was being distributed over port-A. The information vendors were expecting Beta information to be delivered over port-B.

At 11:45 p.m., Mr. Mark Schultz from the Beta Technical Help Desk called Company Delta and said there was no Beta information on the Vendor-R (a pseudonym) system. Delta investigated and said they saw no problems. Delta called the Alpha computer room who forwarded the call to Mr. Mike Ferrero in Alpha's Vendor Network Operations, the group responsible for the information network. Information Network Operations investigated, determined that the system was operating correctly, and transferred the call to the Alpha Technical Operations Control Center, who again checked the information display (unknowingly connected to the wrong port) and reported that information was definitely being sent. A variation of this cycle occurred at 1:00 a.m., when a vendor called the Alpha Application Support Group. The Application Support Group is not staffed at this time of night and their telephone is automatically forwarded to the Alpha Technical Operations Control Center. Mr. Terry Clark, in Technical Operations, took the call, saw that Beta quotations were being sent and forwarded the call to Alpha Vendor Network Operations. Information Vendor Network Operations reported that only heartbeat messages were being sent. Despite this apparent contradiction, neither group escalated this problem.

At 3:00 a.m. (mid-morning in the UK), Beta Market Supervision called Alpha software developer Ms. Joan Hudson's office. She was not at her desk. It is unclear why Beta Market Supervision called Ms. Hudson, but it is likely that she worked with Beta Market Supervision sometime during Beta system development and implementation and was remembered as the expert in this area.

At 5:00 a.m., Mr. Rich Prentice from the Beta Technical Help Desk called Delta and said that they were not receiving Beta information on the Vendor-R system. Delta called the Vendor-R technical help desk and was advised that there was a Vendor-R problem and that no information would be available until 9:00 a.m. It is unknown why Vendor-R said this.

At 6:10 a.m., Ms. Karen Lavone of Alpha Market Supervision arrived and called the Alpha Application Support Group. Ms. Jean Kowalski, in the Application Support Group, saw that Beta information was being sent to the SUN Microsystems system, which receives information from Beta and redistributes it over the Alpha information vendor network. Quotations were not being sent out through the mainframe computer system that connects to the information vendor systems. The Applications Support Group notified the application programmer who corrected the problem at about 6:15 a.m.

ANALYSIS: SURFACE CAUSES

Investigation by the organization involved identified the cause of this failure as a configuration error resulting from a program change. This evaluation was accurate, but incomplete. As the description above illustrates, other factors contributed to the length and severity of the problem.

One factor contributing to this failure was lax and inconsistent change control. Alpha's Information Systems department had a formal methodology for managing the systems development process from project request and initiation through system testing and implementation. Use of the methodology varied depending on the complexity and risk-exposure of the effort. Typically, large, complex, and high-exposure projects made greater use of the methodology than smaller projects. Very small projects may not have used the methodology at all. Instead, they relied on meetings and less formal communication to set objectives and coordinate activities. The remedial system tuning change to balance the information distribution load across the mainframe computer's multiple CPUs, implemented on Saturday, March 5, used an abbreviated and less formal version of the change management methodology. The participants met Friday afternoon and identified the scope of the change as small and the timeframe as production-critical. On Saturday, as planned, the intended changes were, in fact, made correctly, but the developer made an unintended change that prevented the distribution of Beta information. The change on Saturday was not tested. In the Alpha environment, sometimes it was not possible to verify the accuracy of changes by testing. This was particularly true where load balancing of multiple CPUs was involved. Common practice in this situation had been to have a second person independently verify the work of the first and to monitor the first production event after the change. This was not done. If a person familiar with the changes that were made Saturday had monitored the performance of the system at the time of market opening, this problem likely would have been identified more quickly and corrected. This is evidenced by the speed with which the problem was corrected Monday morning when Market Supervision's Ms. Karen Lavone called Ms. Jean Kowalski in the Application Support Group. Market Supervision called Application Support at 6:10 a.m. and the problem was corrected and information distribution restored at approximately 6:15 a.m.

Lax change control of the February 11, change also contributed to duration and severity of the problem. Prior to the change, Alpha and Beta information was distributed through one port. After the change, Alpha information was distributed via one port and Beta information was distributed via a separate port. The project to split the combined distribution ports into separate ports concentrated on external coordination with the information vendors. Technical Operations Control Center staff were not involved and monitoring tools and operating procedures were not modified to reflect this change.

A second contributing factor was the lack of adequate monitoring tools. Despite investigation by no fewer than 11 individuals and groups in 4 organizations, only the Application Support Group had the correct tools, properly configured, to identify the source of the problem. In a failure report to senior management, the IT department recommended that technical support areas should have the ability to "see what the customer sees."

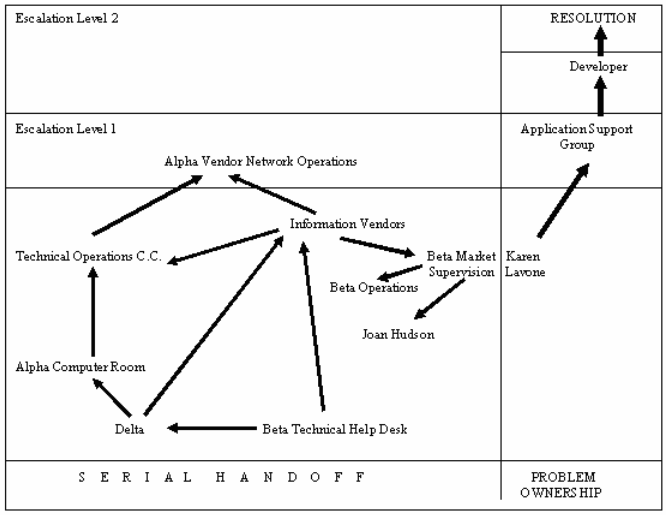
Thus, the initial cause of the failure described here was a software error that was introduced with a seemingly simple modification. Lack of adequate change control and monitoring tools, however, were significant contributing factors because the failure could have been avoided if satisfactory change control procedures were in place, and the severity and duration of the failure could have been minimized if monitoring tools were adequate. Clearly, the value in this study lies in moving beyond these clear-cut causes, and recognizing the role of underlying systemic factors in exacerbating and extending the problem.

ANALYSIS: UNDERLYING ORGANIZATIONAL FACTORS

Alpha and Beta problem handling practices may have been reasonably effective at solving well-defined problems that can be addressed within one particular area such as the Technical Operations Control Center, Market Supervision, Information Vendor Network Operations, or the Application Support Group. Problem handling was less effective when problems were not clearly defined or could not be resolved within a single area.

Analysis of this failure data suggests that three elements are necessary to handle problems in a complex system environment such as Alpha and Beta: problem ownership, problem collaboration; and, problem escalation (White, 2003). This process is characterized by decisive action taken to resolve the problem including collaborating and working interactively and in a parallel problem solving mode with others and, when needed, escalating the problem to a person or group with greater

Figure 1: Problem ownership and communication/coordination



decision making authority and/or access to additional resources. Clear problem ownership is the key ingredient on which collaboration and escalation depend. Rather than problem ownership, the problem handling behavior in this failure was dominated by serial handoff, characterized by notification of another area without continued involvement in problem resolution.

The elapsed time between first problem reports and ownership of the problem and escalation was 10 hours and 10 minutes. As Figure 1, Problem ownership and communication/coordination, illustrates, the serial handoff approach perpetuates problems rather than solving them (White, 2003). The elapsed time between Karen Lavone's ownership of the problem and escalation to the Application Support Group and problem resolution was 5 minutes.

A clear opportunity for problem ownership and escalation occurred when Alpha Vendor Network Operations received a call from a vendor and, investigating, found that only a system heartbeat message was being sent. Vendor Network Operations called the Technical Operations Control Center who found that quotations were being sent. These findings are mutually exclusive. The system cannot be sending both only a heartbeat message and information. Neither Vendor Network Operations nor the Technical Operations Control Center took ownership of this problem or escalated this apparently contradictory situation. It is worth noting that the ability to escalate a problem is directly related to the presence of someone to whom to escalate. Although Alpha operates 24 x 7, there is no clear management onsite presence on weekends and 2nd and 3rd shift. Escalation would have involved contacting managers at home, which operations personnel are reluctant to do.

In a rich technical environment, support groups are not homogenous. Clearly, support groups must have a level of specialization, such as networking, systems administration, and specific application expertise, but effective troubleshooting relies on collaboration among specialized groups. For the most part, the Alpha and Beta approach consisted of investigating the potential causes of a problem within one area before beginning to investigate other areas. This proved ineffective because it increased the elapsed time from problem recognition to problem resolution by investigating potential causes serially, and was predicated on two faulty assumptions: other areas had no relevant information about the problem and, the problem had a single cause. A more effective problem management approach might have followed a parallel problem-solving model. In parallel problem solving, multiple search paths are defined and followed as long as they continue to yield relevant information. All information is shared among the participants who function as a problem solving team versus technical specialists working in isolation. Parallel problem solving relies on two capabilities: a problem facilitator and coordinator who may also fulfill the role of problem owner, and communication technology that supports trouble-ticket-type documentation and a telephone conference line or e-meeting-type system that supports interactive communication (White, 2003).

RELATING THE CASE TO THE LITERATURE

This case illustrates the complexities involved in responding effectively to major system production problems. Despite the reliance of most organizations on key production systems, there is little prior research on the reasons for major system errors during the operating phase of the system development life cycle, and much of what does exist assumes there is a single failure cause. To make matters worse, research on the causes of major operating problems with mission critical systems is contradictory. Ballou (1992), Rocco et al. (1997), Scott (1999), Adams et al. (2001), and Mac Neela (2002) each identified hardware and software failure categories, but the five studies do not correlate. If guided by the work of Rocco et al. (1997), researchers and practitioners might conclude that they should give priority to application software, network, and hardware issues. Scott (1999) and Mac Neela (2002), however, suggest that hardware failures account for only one-fifth of total failures and that research and practitioner emphasis should be given to application design and planning and operator errors, while Adams et al. (2001) report that the primary sources of unplanned outages are technical and that human error represents only 15 percent of failures.

CONCLUSIONS

While the causes of the error described here are certainly not generalizable, the case does help explain the lack of consistency in prior research into operating problems, by suggesting that major production problems may have many causes, some clear-cut, and some relating to underlying organizational imperfections. Future research into the causes of operating problems should explore the problem in its full context and not oversimplify or over emphasize the initial cause. Similarly, practitioners, in the aftermath of a major production problem, should take care not to seize too easily upon eliminating a simple cause. Other organizational deficiencies that contributed to the problem may still be lurking in the background, waiting for another opportunity to strike.

REFERENCES

Adams, T., Rocco, E., Igou, B., Silliman, R., and Mac Neela, A. (2001). Sustainable Infrastructures: How IT Services Can Address the Realities of Unplanned Downtime. *GartnerGroup,* May 15, 2001.

Ballou, M. C. Survey Pegs Computer Downtime Costs at $4 Billion. *Computerworld,* August 10, 1992, 53-56.

Bitner, M. J., Booms, B. H. and Tetreault, M. S. 1990, The Service Encounter: Diagnosing Favorable and Unfavorable Incidents. *Journal of Marketing,* January 1990, 54(1), 71-84.

Flanagan, J. C. (1954). The Critical Incident Technique. *Psychological Bulletin,* July 1954, 5, 327-358.

Mac Neela, A. (2002). IT Availability: Views From European End Users. *GartnerGroup,* July 26, 2002.

Rocco, E., Sweeny, T., and Gately, A. (1997). Mission-Critical Services Part One: The End-User View. *GartnerGroup,* October 1997.

Scott, D. (1999). Making Smart Investments to Reduce Unplanned Downtime. *GartnerGroup,* March 16, 1999.

White, J. D. (2003). Why do bad things happen to good systems? Unpublished dissertation. DePaul University School of Computer Science, Telecommunications, and Information Systems.