IT Application Downtime, Executive Visibility and Disaster Tolerant Computing

Michael A. HARPER
Critical Infrastructure Protection Center, SPAWAR Systems Center Charleston
Department of the Navy
North Charleston, South Carolina 29419, U.S.A.

and

Chad M. LAWLER
Engineering Management, Information, and Systems Department, Southern Methodist University
Dallas, Texas 75275, U.S.A.

and

Mitchell A. THORNTON
Department of Computer Science and Engineering, Southern Methodist University
Dallas, Texas 75275, U.S.A.

ABSTRACT

This paper examines the relationship among disaster tolerant systems, *Information Technology* (IT) application operation and availability, and the executive level management visibility necessary for system operational success. The current state of disaster tolerant application systems is explored including an investigation into the reliability and survivability requirements necessary to achieve disaster tolerant system operation through a simplified network architecture analysis. Specific focus is directed towards the risk of IT application downtime attributable to the increasing dependence placed on critical data-driven applications operating in distributed and unbounded networks. A method for disaster tolerance is proposed which mitigates unplanned downtime through a disciplined approach of computer and people based processes implementing specific documentation procedures. In addition, the importance of executive visibility into the system wide impact of downtime and the resultant effects on the total cost of ownership (TCO) of these critical systems is addressed.

Keywords: Disaster Tolerance, Disaster Tolerant Computing, Survivability, Application Downtime

1. INTRODUCTION

The proliferation of geographically distributed, interconnected, and complex networks throughout both the Government and the private sector has increased the vulnerability for cascading failures with widespread consequences. Secure and reliable operation of these systems is fundamental to the economy, national security, and the quality of life of a nation. However, avoiding failures in complex IT application systems is a challenge due to their large-scale, nonlinear, and time dependent behavior where mathematical models describing such systems are typically vague or non-existent. Critical sectors of our society are becoming increasingly dependent upon highly distributed information systems that operate in unbounded networks, such as the Internet. As these sectors continue to grow and expand in a distributed nature, the importance that they be able to resist and circumvent disasters similarly increases.

Disaster Tolerance is a superset of fault tolerance methods. Disasters, which may be the result of a force of nature or a terrorist event, have cascading effects on the interdependent sectors of an infrastructure. Models for disaster tolerance can differ from those for fault tolerance since they assume that failures can occur due to massive numbers of individual faults rather than a single point of failure. Specifically, the system model can be described as multiple individual system faults that occur nearly simultaneously or close together in time as a series of related events. A naïve way to provide disaster tolerance in a system is to utilize redundancy with redundant components geo-located in different areas; however, this approach has two serious consequences:

1. Communication between the redundant systems becomes a critical link and redundant communication channels may also be required.
2. Some systems are so large that it is impractical to replicate them (for example, the United States electric power grid).

Organizations that have adopted this redundancy based approach must rely on disaster recovery techniques to protect their critical systems. There is a distinct difference between disaster tolerance and disaster recovery. Disaster recovery is the ability to *resume* operations after a disaster, whereas disaster tolerance is the ability to *continue operations uninterrupted* despite a disaster. Developing a system that adapts to preserve essential services involves identifying the mission-critical applications and the availability requirements to provide necessary support for operational success [6].

Continuing advancements in both processing power and storage capacity provides for significant expansion for research in disaster prevention and mitigation technologies. There exists a plethora of literature discussing disaster recovery technologies; however, only a limited subset of this literature is available that focuses on disaster tolerance. Three specific technologies are identified that address the issue of disaster tolerance: the first of these technologies named Myriad is an alternative to data site mirroring for achieving disaster tolerance in large, geographically-distributed storage systems. This methodology is implemented through a protocol permitting cross-site checksums which are updated in such a way that data recovery is always possible [3]. The second technology is an optoelectronic technique which leverages *Dense Wave Division Multiplexing* (DWDM). This approach employs multiple wavelengths to transmit signals over a single optical fiber allowing system-to-system communication and database replication. Currently, the maximum distance a signal can travel without degradation or decrease in reliability is limited to 100km [9]. The final technique is the implementation of high availability clustering technologies. This technology includes multiple nodes in a cluster that allow simultaneous access to data in a shared file system. Therefore the view of the file system is effectively the same from any node in the cluster which provides the potential for disaster tolerant communication [9].

Executive visibility, with regard to information systems, is defined as the ability of executive management to understand the business aspects of an information system or application. Executive visibility should include awareness of service level agreements (SLA) including: compliance, status, stability, and availability of an application. In addition, executive visibility includes the requirement that key decision makers understand the financial costs of downtime as well as the value of uptime of the application.

2. IMPACT OF APPLICATION DOWNTIME

The Internet is an example of an unbounded environment with many client-server networked applications [6]. Users of the Internet exist within a network-centric environment operating in many different administrative domains. Many business-to-business Web-based e-commerce applications depend on conventions within a specific industry segment for interoperability. In the military setting, network-centric interoperability has been adopted through initiatives such as the Department of the Navy's ForceNet architecture which integrates warriors, sensors, networks, command and control, platforms, and weapons into a networked and distributed combat force. The ability to acquire and maintain information superiority demands intense requirements of availability and disaster tolerance.

The capability to deliver essential services in a constant and continuous manner must be sustained even if a significant portion of the system is incapacitated. This capability should not be dependent on the survival of a specific information resource, node or communication link. In a wartime environment, *essential services* might be those required to maintain technical superiority and *essential properties* may include integrity, confidentiality, and a level of performance sufficient to deliver results in less than one decision cycle of the enemy. Similarly, in the public sector, an IT application system maintaining financial information has the requirement to maintain integrity, confidentiality, and availability of essential information and financial services, even if particular nodes or communication links are incapacitated because of a debilitating event [6].

The problem associated with application downtime centers mainly on the economic aspects that result in daily inefficiencies within businesses. The potential risk of massive system or organizational outages leaves complex organizations with the serious risk of complete functional failure. With operations being net-centric, the risks associated with downtime illustrate the need of prevention and/or mitigation against the contributing factors. Emerging technologies in Information Engineering and in particular Disaster Tolerant Computing may hold solutions for medium and large sized business as well as government organizations.

The process of IT solution development and implementation for companies building custom applications, as well as for vendors who are providing applications and services, should inherently deliver solutions with an appropriate level of disaster tolerance built into the architecture. Unfortunately, this practice is often not present, particularly in business scenarios where competition for resources is intense and is thus often ignored for managing the risk of outages and failures.

Building redundancy and disaster tolerant designs into the initial architecture itself is not a new a concept. However, establishing a proven process that incorporates disaster tolerant technologies early in the IT solution design is different from the current concept of disaster recovery. This approach would alter the way the IT solution design process has historically been done but offers the potential for significant benefit in terms of disaster tolerance.

Providing architectural features for disaster tolerance and redundancy into business and organizational IT solutions and applications is a step in a different and potentially financially beneficial direction. However, a disaster tolerant infrastructure and application alone will not solve all of the challenges surrounding the issue of application downtime. In addressing this issue, consideration should be given to technology as well as to the business or organizational strategy and the people and processes that affect the availability of a system. Furthermore, executive visibility into the availability of the application, as well as the resulting insight such visibility would provide would help technology managers better understand the people, strategy, processes and technologies involved in keeping an application available.

3. APPLICATION DOWNTIME AND THE RISK OF IT DISASTERS

IT applications are composed of complex systems and subsystems that are vastly interrelated. The larger an organization, the greater the complexity the systems the organization depends on become. This complexity increases the potential for multiple failures at various system levels or throughout an entire organization, both of which posse serious risks containing enormous consequences for an organization.

IT infrastructures face varying risks of interruption. Although attention is given to contingencies for natural disasters such as hurricanes, tornados, floods, and earthquakes, an IT disaster may be any event that prevents a business from accessing necessary data and systems to conduct normal business operations. In the past, it may have been acceptable to assign a very low probability to the risk of a major disaster occurrence. However, with the rising potential for terrorist activity, this assumption is no longer the case.

A study conducted by KPMG portrays the changing nature of system and application interruptions. According to data collected from 1998 to 2000, natural disasters increasingly comprise a smaller portion of the total causes of IT interruptions. Manmade disasters including both human and IT-related failures continue to represent an increasing portion of the total causes of IT interruptions [15].

Additionally, data from Gartner suggests that almost 80 percent of application downtime is due to people or process related issues caused by application and operation error [16].

4. EXECUTIVE VISIBILITY AND THE COSTS OF DOWNTIME

In observing operational outages of a firm, as well as evaluating the potential of a large-scale system or organizational failure, executive managers often may not have adequate information regarding the actual financial costs of such downtime and outages. Consequently, the value of uptime is also often not understood. Instead, a lack of visibility into the practical impact of such outages on business processes, customer service, product/service delivery and revenue generation tends to be more common. Organizations are affected daily by system outages, with technology outages increasingly becoming as costly and financially detrimental as utility outages. Business application downtime is inevitably the result of information technology outages, which in turn negatively affect a firm's ability to conduct day to day business.

A system that could model and integrate information detailing SLA compliance, cost of downtime, value of uptime, as well as stability and availability statistics could assist in providing greater executive visibility to management staff. This information, in turn would allow management greater insight in making decisions regarding IT applications, infrastructure and business continuance planning. Executive dashboards may assist in delivering this information to management, but calculating it accurately presents additional challenges, as formulating such calculations must be customized to individuals', organizations, processes and applications.

Eagle Rock Alliance released the results of the "2001 Cost of Downtime" conducted as a joint effort between Contingency Planning Research, and Contingency Planning & Management Magazine. A subset of their findings are as follows: 46% said each hour of downtime would cost their companies up to $50K; 28% said each hour would cost between $51K and $250K; 18% said each hour would cost between $251K and $1M; 8% said it would cost their companies more than $1M per hour [15].

All too often, the financial cost of implementing redundant applications or hot/warm failover sites prevents management from implementing these technologies. Executive management equipped with accurate information regarding the financial ramifications of application downtime would be able to more readily engage in the cost benefit analysis of implementing an IT infrastructure that is disaster tolerant. With appropriate executive visibility, management would have supporting information to budget for the costs of implementing technology and applications that are able to survive traumatic disruptions

5. DISASTER TOLERANT IT APPLICATION ARCHITECTURE

The small percentage of firms who have the foresight and resources available to consider the risk and costs of mitigating against application downtime and major outages commonly invest in disaster recovery plans and in alternate 'hot', 'warm' or 'cold' failover sites. Unfortunately, such efforts are often done after an IT solution has been designed and implemented, not before, where it could have the most beneficial effect on architecture and appropriate implementation. Such efforts are often unsuccessful in reaching the goal of providing organizational continuance because they attempt to force an application or technology solution to function in a manner in which it was not designed to function. In actuality, a large portion of organizational investment in disaster recovery is literally wasted in the failed recovery processes itself, reducing the value of this investment, as it does not produce the desired result: IT infrastructure, applications and functionality that are disaster tolerant.

Methodologies addressing disaster or business recovery technologies typically involve local available systems that have the capability of being reinstalled or rebuilt at separate geographic locations. These systems usually require local backups that are stored at offsite vault facilities. In the event of a disaster, the system may be restored offsite, utilizing documented installation processes and backup data. Challenges related to complexity and integration with other systems often renders this approach deficient in the event of a real disaster. Additionally, this basic form of recovery requires significant amounts of downtime while the system in being rebuilt.

An alternative technology approach to disaster or organizational recovery is one that utilizes a multiple node, geographically separate servers and shared storage systems, such as a storage area network (SAN). Nodes in such systems may serve as warm failovers or active members, depending on the nature of the application. Traditionally, backend databases and database servers are critical components that are often duplicated in this manner. Such systems may implement local storage for each database server member, with replication between nodes, or may utilize stretch cluster and SAN technologies to leverage shared storage across limited geographic distances. Challenges arise in these scenarios with locked tables, database writes, synchronous or asynchronous replication and data consistency. An n-tier scenario for disaster tolerance, established as a client-server architecture in which the user interface, functional process logic, business rules and data storage/access are developed and maintained as independent modules on separate platforms, will likely combine various forms of redundancy, leveraging the benefits of different technologies to address disadvantages.

Replication and backup technologies are converging to make complex and expensive disaster recovery management more manageable. Replication applications and appliances that allow subsets of data to be replicated instead of entire storage system replication, and virtual servers, which allow for virtual machine instances to run on geographically separate hardware, are technologies that provide lower cost options for Windows and UNIX environments. This is of particular interest for firms that may not be able to afford the cost of complete multiple sites datacenter and server replication infrastructure or service providers.

Continuous Data Protection (CDP) is a time-addressable form of backup that records file transactions and allows for rapid data restoration and system recovery through the use of synchronization points and preservation of changes to files. Snapshot technologies, which allow for the capturing of images of the state data on disks, also provide a means of rapid recovery to points in time where the snapshots were taken.

Virtual machines can provide hardware-agnostic business continuity functionality, as well as make more efficient use of existing servers in scenarios where duplicate hardware may not be affordable. Virtual machines can provide clustering of physical servers to virtual servers, allowing multiple physical primary servers to failover to virtual servers, many of which could run on one physical server. If separated geographically with the appropriate data replication, this strategy may provide a cost effective means of providing geographical disaster tolerance.

Technology alone will not resolve all of the challenges surrounding organizational continuity. Organization workflow and process automation are emerging areas of business continuity that assist IT managers in defining and documenting the people, processes and technology steps required to recover systems and to execute these steps in the event of a disaster. In addition to Service Level Agreements for organizations, recovery management service levels such as *Recovery Point Objective (RPO)*, defined as the amount of data loss that is acceptable, if any, and *Recovery Time Objective* (RTO), the amount of downtime that is acceptable, if any, are becoming more critical in defining specific recovery metrics [8].

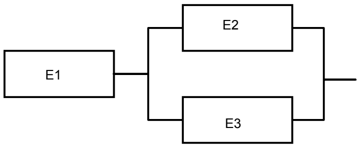
A critical step in designing a viable disaster tolerant IT application is to begin with the idea of disaster tolerance in mind. Disaster recovery and business continuity technologies and plans are often conceived after an application has been designed and implemented, adding into the existing infrastructure disaster recovery functionality features that were not designed into the application itself. As a result, the applications and technologies that they are built upon are then supposed to function in a manner in which they were not designed. In turn, it is often the case that results in failure of the planned disaster recovery functionalities occur. Disaster tolerant applications should be designed from their initial stages with replication, failover, multiple site architecture and other redundant technologies be built into the design itself.

6. RELIABILITY / AVAILABILTY

The following analysis assumes that independent local storage systems at geographically separate sites protecting an alternate site's data with a redundancy scheme other than mirroring are employed. The specific models introduced are not intended as a realistic representation of designed systems; instead it is understood that in specific domain applications, each level of increased model sophistication will enable new structures crucial to the robustness and predictability of the system are used. The goal of this initial research is to take the first step toward more complicated structures in the context of familiar models to illustrate how even a small amount of design preparedness can lead to significant changes in the nature of an interconnected system. This research will expose new directions for the study of complexity and cascading failure in the systems described previously.

7. METHODS

Analysis is performed on a simplified, geographically distributed architecture that models a secure data transfer network undergoing a failover (data site loss). The system under analysis is structured as a series-parallel configuration as follows:



For the purpose of this research, an analytical approach has been used to determine a Time to Failover model for a simplified series-parallel reliability architecture with the random variable *t* as the time to failover. A simulation was performed that modeled the time to failover to an alternate data site in response to a disaster and the loss of a data site. The simulation was run using the EMC^2 Legato RepliStor. Statistical analyses revealed that each element (E_i) of the system follows an exponential distribution of time to failure $T_i \sim \mathcal{E}(\Theta_i)$ for i = 1, 2, … n

Based on these results the following analysis of the system was determined:

The Reliability of the system, R(t)
The instantaneous failure rate, h(t)
The cumulative failure rate, H(t)
The Mean Time to Failover (MTTF)

The following assumptions were made for the system model:

Perfect failure sensing and switching
Zero failure rate during standby
Independent elements
Element time to failure is exponential with parameter λ

The system reliability for the configuration is as follows:

$$R_s(t) = e^{-(\lambda_1+\lambda_2)t} + e^{-(\lambda_1+\lambda_3)t} - e^{-(\lambda_1+\lambda_2+\lambda_3)t} \quad (1)$$

Where system mean time to failure:

$$\text{MTTF} = \frac{1}{\lambda_s} = \Theta_s \quad (2)$$

The MTTF of the series parallel configuration is determined through the relationship:

$$\text{MTTF} = \int_0^\infty R(t)dt = \frac{1}{\lambda_1+\lambda_2} + \frac{1}{\lambda_1+\lambda_3} - \frac{1}{(\lambda_1+\lambda_2+\lambda_3)} \quad (3)$$

The instantaneous failure rate, *h(t)*, is calculated through the relationship of the failure density function and reliability the function:

h(t) = $\frac{f(t)}{R(t)}$ where f(t) = $-\frac{d}{dt}R(t)$, which yields:

$$h(t) = \frac{(\lambda_1+\lambda_2)e^{-(\lambda_1+\lambda_2)t} + (\lambda_1+\lambda_3)e^{-(\lambda_1+\lambda_3)t}}{e^{-(\lambda_1+\lambda_2)t} + e^{-(\lambda_1+\lambda_3)t} + e^{-(\lambda_1+\lambda_2+\lambda_3)t}} - \frac{(\lambda_1+\lambda_2+\lambda_3)e^{-(\lambda_1+\lambda_2+\lambda_3)t}}{e^{-(\lambda_1+\lambda_2)t} + e^{-(\lambda_1+\lambda_3)t} + e^{-(\lambda_1+\lambda_2+\lambda_3)t}} \quad (4)$$

This allows us to derive the Cumulative Failure Rate

$$\text{H(t)} = \frac{1}{t}\int_0^t h(t)dt$$

$$\text{H(t)} = \frac{1}{t}\int_0^\infty \left[\frac{(\lambda_1+\lambda_2)e^{-(\lambda 1+\lambda 2)t} + (\lambda_1+\lambda_3)e^{-(\lambda_1+\lambda_3)t}}{e^{-(\lambda_1+\lambda_2)t} + e^{-(\lambda_1+\lambda_3)t} + e^{-(\lambda_1+\lambda_2+\lambda_3)t}} - \frac{(\lambda_1+\lambda_2+\lambda_3)e^{-(\lambda_1+\lambda_3)t}}{e^{-(\lambda_1+\lambda_2)t} + e^{-(\lambda_1+\lambda_3)t} + e^{-(\lambda_1+\lambda_2+\lambda_3)t}} \right] dt \quad (5)$$

The objective of the analysis is to determine a Mean Time to Failover (MTTF). This value can be estimated using a (1- α) $\bullet$95% Lower Confidence Interval (Θ_L, ∞). This is based on the condition that testing has been discontinued after a fixed amount of total time T_c has elapsed.

$$\Theta_L = \frac{2T_c}{\chi^2_{\frac{\alpha}{2}, 2r}} \quad \text{where} \quad (6)$$

$\chi^2_{p,df}$ is the value of x~ χ^2_{df} such that P(X> χ^2_{df}) = p. Five simulations of time to failover (in seconds) were tested. The results follow:

22 27 33 47 73

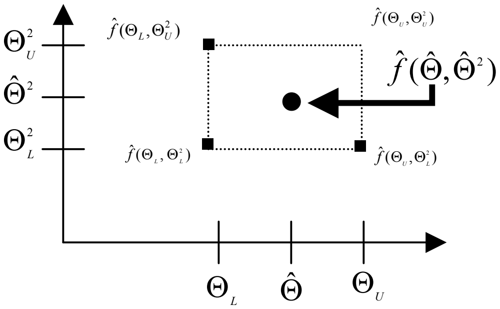
The MTTF is estimated by the point estimate

$$\hat{\Theta} = \hat{f}(\hat{\Theta},\hat{\Theta}^2)\ \frac{\sum_{i=1}^{5} x_i}{5} = \frac{95}{5} = 19 \text{ seconds.}$$

A 95% lower confidence interval on the mean failover is set. This provides a measure of the possible variation. Appropriate chi-square test values are

$$\chi^2_{0.05,10} = 18.31$$

Therefore, a 95% lower confidence limit on Θ the MTTF is 10.38 seconds for the series-parallel system configuration. Using our point estimate, $\hat{f}(\hat{\Theta},\hat{\Theta}^2)$, on the MTTF and varying the confidence interval, we determine that $\hat{f}(\Theta_L, \Theta_U^2)$ and $\hat{f}(\Theta_U, \Theta_U^2)$ are the worst case scenarios as these points have the greatest variability. In a similar manner, $\hat{f}(\Theta_L,\Theta_L^2)$ and $\hat{f}(\Theta_U,\Theta_L^2)$ have the least variability and is therefore more desirable for an organization to work to achieve an MTTF in this area for increased predictability.


Θ_U^2
$\hat{\Theta}^2$
Θ_L^2
$\hat{f}(\Theta_L,\Theta_U^2)$
$\hat{f}(\Theta_U,\Theta_U^2)$
$\hat{f}(\hat{\Theta},\hat{\Theta}^2)$
$\hat{f}(\Theta_L,\Theta_L^2)$
$\hat{f}(\Theta_U,\Theta_L^2)$
Θ_L
$\hat{\Theta}$
Θ_U

This risk avoidance posture affords protection from the costs and consequences of unpredictable downtime and gives the enterprise the ability to analyze, predict, and rationally accept risk, as warranted by an application's availability requirements.

8. CONCLUSIONS

Events such as 9/11, the North East electrical grid failure, and the deluge of hurricanes hitting Florida emphasize the need not only to continue disaster recovery plans, but to develop disaster tolerant systems. In the event of a disaster, success hinges on the ability to restore, replace, or re-create.

Increased awareness and application of the areas discussed in this paper will provide executive management significant benefit through increased visibility into the business aspects of information systems and applications regarding the value of uptime, the costs of downtime and the associated aspects involved in implementing disaster tolerant IT architectures. A sufficient level of visibility will provide executive level management with the information necessary to make appropriate decisions regarding architecture, implementation, maintenance and support of IT applications for technology infrastructures capable of surviving and adapting to disasters.

9. REFERENCES

[1] G. A. Alvarez, W. A. Burkhard, and F. Cristian. "Tolerating multiple failures in RAID architectures with optimal storage and uniform declustering." *Proceedings of the 24th Annual International Symposium on Computer Architecture*, pp. 62–72, Denver, CO, June 1997. IEEE Computer Society Press.

[2] M. Aminl, "National Infrastructures as Complex Interactive Networks", *Electric Power Research Institute. Automation, Control, and Complexity: An Integrated Approach*, Samad & Weyrauch (Eds.), John Wiley and Sons, pp. 263-286, 2000.

[3] F. Chang, M. Ji, S.T Leung, J. MackCormick, S. Perl, L. Zhang, "Myriad: Cost-effective Disaster Tolerance," Proceedings of the FAST 2002 Conference on File and Storage Technologies. USENIX Association. Monterey, California. January 28-30, 2002.

[4] D. Fruend, "Disaster tolerant Unix: removing the last single point of failure," Illuminata, Inc, 2002. Accessed at http://h71000.www7.hp.com/openvms/whitepapers/Illuminata.pdf

[5] J. Gray, "The Revolution in Database Architecture," Microsoft Research Technical Report (MSR-TR-2004-31). Microsoft Research, Microsoft Corporation, 2004.

[6] H. F. Lipson and D. A. Fisher, "Survivability-A New Technical and Business Perspective on Security," *Proceedings of the New Security Paradigms Workshop*, September 21-24, Association for Computing Machinery, 1999.

[7] R. J. Ellison, D. A. Fisher, R.C. Linger, H. F. Lipson, T. A. Longstaff, N. R. Mead, "Survivability: Protecting Your Critical Systems," *IEEE Internet Computing*, November/December 1999

[8] K. Parris, "Disaster Tolerant Cluster Technology and Implementation", HP World 2003 Solutions and Technology Conference and Expo, 2003.

[9] "Improving system availability with storage area networks," Barocade Communications Systems, Incorporated, 2001. (http://www.dlt.com/storage/WhitePapers/Brocade/HA_WP_02.pdf)

[10] "HP Extended Cluster for RAC-100 kilometer separation becomes a reality," A White Paper, Hewlett-Packard Development Company, L.P, 2004.

[11] The Hidden Cost of Downtime," A White Paper, SmartSignal, Inc., 2002.

[12] Naval Transformation Roadmap. "Power and Access…From the Sea; Sea Strike, Sea Shield Sea Basing." Accessed at http://www.onr.navy.mil/ctto/docs/naval_transform_roadmap.pdf

[13] "Integrating Availability and Disaster Tolerance" 1999; Strategic Research Corporation.

[14] "Designing Disaster Tolerant High Availability Clusters" Hewlett-Packard Development Company, L.P, 2004. December 2004. Accessed at http://docs.hp.com/en/B7660-90016/B7660-90016.pdf

[15] Contingency Planning & Management/KPMG Business Continuity Planning Survey," cited in Andy Hagg, "BCP on the Rise," Contingency Planning and Management, January 2001

[16] Lanowitz, Theresa, Gartner, Inc. 2001