FauxCorp anti-virus event, July 3rd through July 6th 2009

A Business Case for the Enterprise Major Incident Handling Plan

Braun Tacon – www.BraunsBlog.Com

The following story is a fictional dramatization of a real event that occurred on July 3rd 2009. A prominent anti-virus (AV) software provider delivers a definition update to their customers across the globe. This is a routine update that has zero consequences 99% of the time. In this instance, due to the use of out of support anti-virus engines by some customers combined with a lack of solid regression testing by the vendor, a perfect storm is created which causes considerable damage and disruption to both the AV vendor and their customers. Affected servers and clients are for all intents and purposes destroyed by this update; an unfortunate occurrence but one which serves extremely well to make the business case for a Major Incident Handling Plan as part of any Enterprise Risk Management Portfolio.

The FauxCorp is one of the leading wholesale distributors of automotive parts in the world. Headquartered in the United States with regional offices located in Europe, Southeast Asia and Australia they employ approximately 6,000 individuals of which almost 1/6th (900) work in the Sales Division. To provide the sales team the tools they need to best serve their customers, the FauxCorp has spent considerable time and money developing a highly customized Sales Ordering and Reporting Tool called “SORT”. This tool tightly integrates the FauxCorp sales and ordering process with the supply chain systems of their primary suppliers. This enables FauxCorp to keep their inventories low and deliveries swift. SORT provides the FauxCorp a very real competitive edge and is recognized company wide as a key strategic advantage.

The FauxCorp Sales Division is rolling out a significant upgrade to SORT via a software patch. The patch will be deployed to 57 servers supporting the entire sales division worldwide. The plan is to deploy the patch at 5:00 PM on Friday July 3rd using Microsoft’s Systems Management Server (SMS). July 3rd has been chosen because it is a US holiday and the majority of company employees will be off, making it an ideal time to schedule the required downtime. It is also critical to deliver the upgrade prior to Saturday July 4th as the upgraded SORT will be demonstrated at the Global Sales-All meeting being held in Colorado Springs, Colorado over the 4th of July weekend.

Application of the patch requires an automated reboot of the servers. This is not seen as an issue. The company regularly patches and reboots servers via SMS as part of their Windows patching process. Years have gone by without any significant incidents or disruption. The patch deployment process is mature, well tested, and proven.

Testing of the SORT upgrade has been in progress for over 60 days. The test plan is comprehensive and includes contingency plans for every foreseeable issue. The SORT software is built and maintained to exacting standards. All previous upgrades have gone extremely well. Testing in the development systems has consistently delivered an overall 99 % success rate and the rollback plan is 100% assured. Highly confident that their testing and deployment plan is solid, at 3:00 pm on Thursday July 2nd Senior Management gives the go-ahead to proceed with the upgrade as planned on Friday. The Change Request for the four hour outage is approved and published.

Friday July 3rd arrives and the upgrade team begins preparation for the patch. The patch is distributed to the SMS system. Testers worldwide are notified that post-patch testing will begin at approximately 5:00 pm US Mountain time. Barring any unforeseen problems testing should be complete by 8:00 pm and the upgrade team should be able to wrap up about 1 hour later to begin a well earned 2 and ½ day weekend. Besides…the majority of the IT Staff not assigned to this project will already be gone and enjoying the long-weekend. It is a good time to work at FauxCorp.

Approximately six hours before the planned patching of the SORT servers, FauxCorp receives an anti-virus definition update from their AV provider. This is not uncommon nor is it noted. Definition updates such as these happen weekly, daily, or sometimes even more than once in a day when fighting a major virus. In this case there is no virus to battle…the update is routine.

Rolling out a definition update to 480 servers and 5,500 clients worldwide is made possible by the use of automation. Many large companies distribute virus definitions to clients without testing because there is too much cost and risk to do otherwise. The AV provider does the testing and you deploy without delay.

At 11:01 am FauxCorp receives the July 3rd definition update. Their AV update system first sends a copy of the new definitions to every distribution server worldwide then the deployment of the definitions to individual servers and clients begins at approximately 11:35 am. By 1:00 pm over 90 % of the servers and 45% of the clients at FauxCorp have received the definition update.

The SORT system patching happens as scheduled with the SORT servers rebooting sometime between 5:10 and 5:20 pm. The upgrade team executes exactly to the plan but what should have been a huge win for the upgrade team turns into a company crisis which completely consumes the FauxCorp IT staff for days as they attempt to restore the critical SORT system. More than sixty-five hours pass before services are completely restored and normal operations resume. The restoration process is highly manual and significantly slows the recovery efforts. As a result there is substantial Brand Damage to FauxCorp as all order processing and shipments are delayed for at least forty-eight hours. Many large and important customers are impacted and FauxCorp's reputation is severely tarnished.

Root Cause Analysis – How could this happen?

There was an issue with the definition update. In an apparent failure of regression testing by the AV software vendor it is discovered that this particular definition file works without incident when installed to the current AV engine or one revision previous. Unfortunately for anyone using an AV Engine older than that it becomes an insidious cancer destroying it's host, the very object it is supposed to protect. Complicating the matter is the fact that the definitions are perfectly content installing into out of support software.

With this particular definition update, when coupled with an out of support AV Engine, critical operating system (OS) files are misidentified as being infected with a virus and are quarantined. This is not a problem as long as the system in question remains running, but if the device is rebooted the quarantine of the files will cause the OS to crash resulting in the infamous Windows "BSOD" (blue screen of death). At that point the only remedy is a full OS rebuild; a formidable task for any Enterprise.

Additionally, due to a failure in the FauxCorp AV software Release Management process, the majority of the SORT servers are running an out of support version of the AV software. Not every server at FauxCorp is affected by this chain of events but estimates are that 80 % of the SORT servers are destroyed and must be rebuilt. As a result of this Major Incident the SORT upgrade is a devastating failure instead of a spectacular success.

In analysis this failure was not caused by a lack of planning or testing by the SORT upgrade team. Nor was it singularly caused by the AV vendor's lack of regression testing or the fact that out of support AV software existed in the FauxCorp enterprise, although both are a contributing factor. The most significant contributor to this event was the absence of a Major Incident Handling Plan. The presence of a Plan would have helped insure that the AV false-positives were identified and correlated in a timely fashion. It would have provided contact information to insure that key management and subject matter experts were engaged early on, long before the event evolved into a crisis. Most importantly though, a Major Incident Handling Plan would have put the structure and processes in place that almost certainly would have surfaced the planned update allowing decision makers to cancel the SORT patch and reboot. This one factor, the rebooting of the SORT servers, was the most significant contributor to the FauxCorp anti-virus event and subsequent system downtime.

It was the lack of the Plan. No denying it.