TRIAGE – An Overview

Technology Resources Independent Archive Gathering Entity

Background

Due to the increasing number of sytsems being deployed within the local infrastructure, the Technology Resources Group (TRG) identified a need for a centralized backup solution that would provide automated archiving of user data in a remote location, without sacrificing security or data integrity. The TRIAGE system provides a solution to this gap.

Overview

The TRIAGE system is comprised of a small piece of custom code running on each target server, plus a dedicated server running additional custom code, deployed behind a firewall and connected to a multi-terabyte storage area network (SAN). The code running on each target server will create a local backup of all data on that box at a specified interval. This backup is placed in a secure folder to protect it from unauthorized local access. At scheduled intervals, the dedicated backup server will create an encrypted tunnel to each target server, and will copy over any new data it finds in the protected local backup directory. This data is then compressed and placed onto the SAN for long-term storage.

Scheduling and Frequency

Archives times and frequency are scheduled via the UNIX cron utility. The more mission critical a system is, the more often the backup should be performed. For example, the live Catalyst system (with student data) is backed up daily, whereas the CatalystDev server (where courses are constructed) is backed up weekly.

Shortcomings

The following list contains identified shortcomings with the current deployment of the TRIAGE system, along with potential mitigation solutions:

- Aging/limited SAN drive space. Due to the expense associated with procuring high-density SAN solutions, the currently deployed SAN is a repurposed Apple X-SAN from a previous project. This hardware is only 5.5 TB and is 3 years beyond it's expected usability period. Additionally, the limited space requires the TRG systems administrator to manually remove the cruft that accumulates over time, such as redundant or out-of-date backups. To mitigate the hardware must be replaced.
- Hit-by-a-bus. The TRIAGE system was constructed in-house by the systems administrator, and is therefore not a "supported" software solution. Should the SA become unavailable (retire/resign/be hit by a bus), someone else must take over the maintenance of the system. However, all code for this project was written using the industry standard Perl language, therefore mitigation of this threat is accomplished by simply hiring a replacement SA with knowledge of the Perl language.
- High duplication of data. By compressing the full data set with each backup, the amount of duplicate data associated with each backup is quite high. Mitigate of this issue will involve a single rotating compressed backup and a hard linking duplicate files.
- TRIAGE is not on a UPS. The TRIAGE system is not wired to an uninterruptable power supply. In the event of power loss, the TRIAGE system will hard fail. This may cause loss of data if it occurs during a write operation. To Mitigate, a UPS should be acquired and installed to support both the TRIAGE server and any associated SAN hardware.