**Protection Identification of Corrupted Data Security**

**Project Members:**

**Project Consultant:**

|  |  |  |
| --- | --- | --- |
| **Name** | **Position** | **Email** |
| Mr. Ernesto Boydon | Professor | ernestob@apc.edu.ph |

**Project Adviser:**

|  |  |  |
| --- | --- | --- |
| **Name** | **Position** | **Email** |
| Mr. Justin Pineda | Professor | [justinp@apc.edu.ph](mailto:justinp@apc.edu.ph) |
| Mr. Paulo Oblepias | Professor | pauloo@apc.edu.ph |

**Project Team:**

|  |  |  |
| --- | --- | --- |
| **Name** | **Position** | **Email** |
| Kent Michael Miculob | Project Manager/Analyst | kpmiculob@student.apc.edu.ph |
| Reginald John Steven Alberca | Project Researcher | rralberca@student.apc.edu.ph |
| Maria Letty Laureta | Project Documenter/Designer | mmlaureta@student.apc.edu.ph |

1. **Introduction**
2. **Background of the Problem**

Nowadays, technology is inevitable. From personal matter all the way to school works to professional works, researchers are inputting information to computers and other gadgets, thus, safeguarding is vital and crucial. They need to protect it not just from the hackers, who wanted to steal information for their own benefit, but also from being corrupted that will steal all the hard works and left you nothing but despair. These scenarios wherein the data itself becomes corrupted are due to hardware and software failures. Computers don’t have the ability to prevent nor retrieve because to them, it appears as the same data.

There are applications such as File Repair which identifies the corrupted data in your hard drive and creates a new usable file where it extracts much of the recovered data as possible. There is also an application called BadCopy pro where it recovers deleted files, formatted drive, or data loss due to damage, media errors, bad sectors and other reasons for flash drives and save the recovered files into a directory you specify. Another is application that offers safeguard is Microsoft Document, The approach that the Microsoft document does is deferent from file repair & other recovery applications such as BadCopy pro, since the Microsoft Document provides backup for your document in case it might be corrupted if the computer suddenly turns off, and it is a good practice since prevention is greater than cure.

As we have read the information, we can conclude that there are a lot of application which fix and recover corrupted data, there are also application where they provide prevention mechanism to avoid having corrupted data. But the researches still thought that all of this application is not enough.

1. **Statement of the Problem**

Corrupted data had always been a problem in the computing industry, since most of the information are now stored in a mechanical device. A lot of factors can result to data corruption, such as software and hardware failures. There are identified tools or applications that can be used for in case of having corrupted data such as specified above. The researchers would like to know, how could they enhance the existing applications and how could they provide a better way in safeguarding your file from corruption.

1. **Objectives**
   1. General
      1. This application aims to provide a way for users to secure their data in the event of corruption.
   2. Specific
      1. To be able to provide an automatic backup on user’s work.
      2. To be able to search for corrupted data's and provide for its location.
      3. To be able to identify whether the corrupted data is fixable or not
      4. To be able to repair the fixable data.
2. **Significance**

A lot of applications that were made related to this project are mostly for corrupted data fixation and retrieval. The essence of this research is to provide a way on how to prevent having corrupted data by means of having an application that is capable of detecting and fixing corrupted data, preventing having corrupted data by providing a way to backup, and minimizing the existence of corrupted information in real-time. The issue of corrupted data already exists along with the advancements of mechanical technology. As said by Microsoft “No computer is failure proof, but preventive measurements and strategic planning can make a computer more failure resistant”. And this innovation offers more than that.

1. **Related Literature/ Related Studies**
2. **Causes of Data Corruption**

The common causes of corrupted data is power outages or other related to power failure, Inappropriate shutdown caused by the user, hardware problems or failures such as hard drive failure, bad sectors, and etc. another reason is software failures caused by bad programming, particularly if it results in either hard restarts or data that is saved incorrectly. Any of these causes can result in a corrupted hard drive directory. A corrupted hard drive directory can cause files to apparently "go missing" and lead to further data loss or corruption. (Wikipedia : Data Corruption, n.d.), (Smith, 2014)

1. **Data Corruption**

Some file systems, such as Btrfs, HAMMER, ReFS, and ZFS, use internal data and metadata check summing to detect silent data corruption. In addition, if a corruption is detected and the file system uses integrated RAID mechanisms that provide data redundancy, such file systems can also reconstruct corrupted data in a transparent way. This approach allows improved data integrity protection covering the entire data paths, which is usually known as end-to-end data protection, compared with other data integrity approaches that do not span different layers in the storage stack and allow data corruption to occur while the data passes boundaries between the different layers. (Wikipedia : Data Corruption, n.d.)

1. **Checksum**

A checksum is a count of the number of bits in a transmission unit that is included with the unit so that the receiver can check to see whether the same number of bits arrived. If the counts match, it's assumed that the complete transmission was received. Both TCP and UDP communication layers provide a checksum count and verification as one of their services. (Search Security, n.d.)

1. **File System**

A file system is a part of the operating system that determines how files are named, stored, and organized on a volume. A file system manages files and folders, and the information needed to locate and access these items by local and remote users. Microsoft Windows Server 2003 supports both the FAT and NTFS file systems. NTFS allows you to gain the maximum benefits for the needs of today’s enterprise business environments from Windows Server 2003, such as increased security, more robust and reliable performance, as well as a design for greater storage growth, features not found in FAT. (How to Repair Corrupted Windows System Files with the SFC and DISM Commands, n.d.)

1. **RAID Technology**

With RAID technology, data can be mirrored on one or more disks in the same array, so that if one disk fails, the data is preserved. RAID also offers the option of reading or writing to more than one disk at the same time in order to improve performance. In this arrangement, sequential data is broken into segments which are sent to the various disks in the array, speeding up throughput. A typical RAID array uses multiple disks that appear to be a single device so it can provide more storage capacity than a single disk. (Beal)

1. **Assembly Language**

In implementing the corrupted data retrieval we can use the Assembly language. It is a low leveled programming language for a computer, or other programmable device specific to a particular computer architecture meaning you can communicate through the device one on one. Each assembly language is specific to a particular computer architecture, in contrast to higher level of programming which are generally portable across multiple architectures, but require interpreting or compiling. Assembly language may also be called symbolic machine code. Assembly language also use mnemonic to show a list of each low leveled machine instruction or operation. Many operation takes one or more operands to be able to complete instruction. Assembly language is converted into executable machine code by a utility program referred to as an assembler like NASM, MASM, etc. (Assembly Language, n.d.) , (Assembly Programming Tutorial, n.d.)

1. **An Analysis of Data Corruption in the Storage Stack**

An important threat to reliable storage of data is silent data corruption. In order to develop suitable protection mechanisms against data corruption, it is essential to understand its characteristics. In this paper, we present the ﬁrst large-scale study of data corruption. We analyze corruption instances recorded in production storage systems containing a total of 1.53 million disk drives, over a period of 41 months. We study three classes of corruption: checksum mismatches, identity discrepancies, and parity inconsistencies. We focus on checksum mismatches since they occur the most.

We ﬁnd more than 400,000 instances of checksum mismatches over the 41-month period. We ﬁnd many interesting trends among these instances including: (i) near line disks (and their adapters) develop checksum mismatches an order of magnitude more often than enterprise class disk drives, (ii) checksum mismatches within the same disk are not independent events and they show high spatial and temporal locality, and (iii) checksum mismatches across different disks in the same storage system are not independent. We use our observations to derive lessons for corruption-proof system design. (Bairavasundaram, Arpaci-Dusseau, Arpaci-Dusseau, Goodson, & Schroeder, 2008)

1. **Method and apparatus for recovering data from damaged or corrupted file storage media**

An automated method and apparatus for identifying and copying lost files from a mass data storage device of a computer when file system information (as opposed to the actual data files) stored on the mass data storage device has been corrupted or destroyed. The mass data storage device is scanned on a sector-by-sector basis in order to attempt to identify sectors containing file system data structures and file attributes. Identification is made using data signature and/or pattern matching filters. The location of, and any valid information found in, such sectors is used then to derive information useful in locating files to be copied to another storage device. For example, in a FAT, NTFS or other cluster-oriented file system, if information on the number of sectors per cluster (SPB) is not available from a boot directory, it and a cluster base (the starting sector of cluster 0) are calculated using the physical location of the beginning sectors of the directories or folders. When a starting cluster is known from a directory entry, but not additional file allocation information, a cluster chain may be reconstructed utilizing one or more of several disclosed methods.

1. **Systems and methods for automated maintenance and repair of database and file systems**

The present invention relates generally to database and file system management and, more particularly, to automatic database and file system maintenance and repair to ensure data reliability. Various aspects of the present invention relate to responding and correcting data corruptions at a data page level for all data page types, as well as to recovery (including rebuild or restore operations) for various scenarios including, without limitation, index page corruptions (clustered and non-clustered), data page corruptions, and page corruptions in the log file.

1. **System and method for detecting correcting and discarding corrupted data packets in a cable data delivery system**

A method and apparatus for detecting errors and improving quality in real-time data transmissions is provided. In one embodiment, the packet header checksum field is turned off to allow uninterrupted transmission of data packet payloads. A checksum added to each independent data segment in the datagram payload permits each data packet to be examined separately, resulting in improved transmission quality.

1. **Validating Population Viability Analysis for Corrupted Data Sets**

Diffusion approximation (DA) methods provide a powerful tool for population viability analysis (PVA) using simple time series of population counts. These methods have a strong theoretical foundation based on stochastic age-structured models, but their application to data with high sampling error or age-structure cycles has been problematic. Recently, a new method was developed for estimating DA parameters from highly corrupted time series. We conducted an extensive cross-validation of this new method using 189 long-term time series of salmon counts with very high sampling error and non-stable age-structure fluctuations. Parameters were estimated from one segment of a time series, and a subsequent segment was used to evaluate the predictions regarding the risk of crossing population thresholds. We also tested the theoretical distributions of the estimated parameters. The distribution of parameter estimates is an essential aspect of a PVA because it allows one to calculate confidence levels for risk metrics. This study is the first data-based cross-validation of these theoretical distributions. Our cross-validation analyses found that, when parameterization methods designed for corrupted data sets are used, DA predictions are very robust even for problematic data. Estimates of the probability of crossing population thresholds were unbiased, and the estimated parameters closely followed the expected theoretical distributions.

1. Bibliography

Beal, V. (n.d.). RAID - redundant array of independent disks. Retrieved from Webopedia: http://www.webopedia.com/TERM/R/RAID.html

Wikipedia : Data Corruption. (n.d.). Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Data\_corruption

Assembly Language. (n.d.). Retrieved from Wikipedia: https://en.m.wikipedia.org/wiki/Assembly\_language

Assembly Programming Tutorial. (n.d.). Retrieved from tutorialspoint: http://www.tutorialspoint.com/assembly\_programming/

BadCopy Pro - Flash Drive Data Recovery Software. (n.d.). (Jufsoft) Retrieved from http://www.jufsoft.com/badcopy/flash\_drive\_recovery.asp

How to Repair Corrupted Windows System Files with the SFC and DISM Commands. (n.d.). Retrieved from http://www.howtogeek.com/222532/how-to-repair-corrupted-windows-system-files-with-the-sfc-and-dism-commands/

Keylogger. (n.d.). (webopedia) Retrieved from http://www.webopedia.com/TERM/K/keylogger.html

Smith, M. (2014, July 17). What Is Data Corruption And How To Prevent It. (makeuseof) Retrieved from http://www.makeuseof.com/tag/data-corruption-prevent/

File Repair. (n.d.). Retrieved from Softonic: http://file-repair.en.softonic.com/

Search Security. (n.d.). Retrieved from checkum: http://searchsecurity.techtarget.com/definition/checksum

Bairavasundaram, L. N., Arpaci-Dusseau, A. C., Arpaci-Dusseau, R., Goodson, G. R., & Schroeder, B. (2008). Retrieved from Digital Library: http://dl.acm.org/citation.cfm?id=1416947