A Project

entitled

combox

by

Siddharth Ravikumar

Submitted to the Graduate Faculty as partial fulfillment of the requirements for the Masters of Science Degree in Computer Science

Dr. Robert C. Green II, Committee Chair

Dr. XX, Committee Member

Dr. XX, Committee Member

Dr. Michael Ogawa, Dean
College of Graduate Studies

Bowling Green State University
May 2016

An Abstract of

combox

by

Siddharth Ravikumar

Submitted to the Graduate Faculty as partial fulfillment of the requirements for the Masters of Science Degree in Computer Science

Bowling Green State University
May 2016

File storage providers on the Internet have made it non-trivial for individuals to store personal files on the file storage provider's computers. After Mr. Snowden disclosed information about the National Security Agency' (NSA) surveillance programs that allowed the NSA to access information stored on file storage provider' computers, online file storage became a non-solution for storing personal files for everyone who detested the possibility of somebody else being able to access their personal files. In the past, there have been separate efforts to come with a solution to allow individuals to use storage space provided by file storage providers in a way that it made it impossible for file storage providers and to access the files. combox is one such effort. It allows an individual to store personal files in the "combox directory" on all her computers (running GNU/Linux or OS X) and the combox program takes the files, splits and encrypts them and spreads them across file storage providers' directories. Therefore, when an individual uses storage space provided by file storage providers through combox, each file storage provider gets only a part of the file in an encrypted form.

Dedicated to the `$EDITOR` I use to literally write everything.

Acknowledgments

Dr. Robert C. Green II who gave me an opportunity to work on combox.

Contents

List of Tables

List of Figures

List of Abbreviations

YAML ..... YAML Ain't Markup Language

Preface

42.

Chapter 1

Introduction

Chapter 2

Background

Chapter 3

Literature Review

Chapter 4

Architecture and Design

Chapter 5

Testing

> Testing shows the presence, not the absence of bugs.
>
> *Dijkstra*[1]

5.1 Unit testing

The `nose`[2] testing framework was used to write unit tests for the functions and classes part of the `combox.config`, `combox.crypto`, `combox.events`, `combox.file`, `combox.silo` `combox._version` modules. Unit tests were not written for `combox.cbox`, `combox.gui`, `combox.combox.log` modules.

Unit tests for combox become reality by pure serendipity. During the time, when I started working on combox, I was learning to use the `nose` library to unit test python code. Since, `combox` was being written in python, I started making it a norm to write unit tests for functions and classes in combox modules.

As mentioned before, unit tests were not written for some modules either because it would make no sense to write one (for the `combox.cbox` module, for instance, which basically uses functions and classes defined in other modules to run combox) or it was not clear how to write unit tests it (the `combox.gui` contains just the `ComboxConfigDialog` a graphical front-end which uses the configuration function

defined in the `combox.config` module to complete the combox configuration based on the user input).

It must be noted here that pure Test Driven Development (TDD) was not observed – most of the time the function/class was written before the its corresponding test was written.

5.1.1 Benefits

While writing unit tests definitely increased the time to write a particular feature, it enabled me to immediately check if a feature worked as it should for the given use case or given set of inputs.

With the benefit of hindsight, unit tests greatly helped in testing the compatibility of combox on OSX. Before the `v0.1.0` release, combox's node directory monitor always assumed that a file's first shard (`shard0`) is always available; while this assumption did not create any problems on GNU/Linux, on OS X, this assumption made the node directory monitor to behave erratically – this issue (bug #4[3] was immediately found when the unit tests were run for the first time on OS X. Another instance where unit tests helped was just before the `v0.2.0` release; major changes, including the introduction of file locks in the `ComboxDirMonitor`, were made to the `combox.events`. When the unit tests were run OS X, two tests failed, revealing a difference in behavior of watchdog[4] on GNU/Linux and OS X on file creation[5]; without unit tests, there is a high probability that this bug would never have been found by now.

5.1.2 Caveats

Unit tests are helpful in testing the correctness of a feature for `N` number of use cases but it does not necessarily mean the written feature correctly behaves for use

cases that the author of the feature did not consider or did not think about while writing the respective feature. As Dijkstra correctly observed:

Unit tests failed to reveal bugs #4, #5 #6 #7 #5 #10 #11[3]; these bugs were found when manually testing combox.

5.2 Manual testing

The unit tests for the `combox.events` module test the correctness of the `ComboxDirMonitor` and `NodeDirMonitor` independently; in order to comprehensively test the correctness of both `ComboxDirMonitor` and `NodeDirMonitor`, it was required to manually test combox running on more than one computer. As you'll see in the following subsections, several bugs were found and fixed while doing manual testing.

Three different types of setups were used to test combox. The first kind of setup has two GNU/Linux machines each using combox to sync files between each other with Dropbox and Google Drive being the nodes; the second kind of setup has a GNU/Linux machine and a OS X machine each using combox to sync files between each other with Dropbox and Google Drive being the nodes; the third kind of setup has a GNU/Linux machine and OS X machine each using combox to sync files between each other with Dropbox, Google Drive and a USB stick as nodes.

5.2.1 General setup and notes

- On the GNU/Linux machines, the official Dropbox client was used to sync the Dropbox node directory to Dropbox' servers. `rclone`[6] was used to sync the Google Drive node directory to Google Drive' servers;At the time of testing, Google Drive did not have client for GNU/Linux.

- On OS X, the official Dropbox client was used to sync the Dropbox node directory to Dropbox's servers; the official Google Drive client was used to sync the

Google Drive node directory to Google Driver' servers.

- Since combox is extremely event-driven, combox must be started before the Dropbox and Google Drive clients start syncing their respective directories (nodes).

5.2.2 Testing on two GNU/Linux machines

combox was run to two GNU/Linux machines and a file was alternatively created/modified/renamed/deleted on an of the GNU/Linux machine and it was verified if the respective file was also created/modified/renamed/deleted on the other GNU/Linux machine. One of the GNU/Linux machine (`lyra)` was a virtual machine running Debian GNU/Linux stable (version 8.x); the other GNU/Linux machine (`grus`) was a physical machine running Debian GNU/Linux testing. The node directories to scatter the files' shards were the Dropbox directory and Google Drive directory. The official Dropbox client was used to automatically sync files from the Dropbox directory to the Dropbox' server; `rclone`[6] was used to sync files from Google Drive directory to Google Drive' server.

5.2.2.1 Issues found

- Some editors, especially on POSIX complaint systems, create backup version of the file being edited. combox was detecting this backup file as a "new file" and it split it into shards, encrypted the shards and scattered the shards across the node directories. The right thing for combox to do was to ignore these backup files and do nothing about them. This issue was fixed on `2015-09-29`[3]. Now the `ComboxDirMonitor`, on a "file created" or "file modified" event, returns from the `on_created` or `on_modified` callback when it finds that the file is a backup/temporary file.

- Dropbox client maintains the `.dropbox.cache` directory under the root of the Dropbox directory.

  - When a file (shard) was created on another computer, the Dropbox client pulls the new file (shard) to this computer into `.dropbox.cache` as a temporary file and then moves the new file (shard) to its respective location with the appropriate name.
  - When a file (shard) was modified on another computer, the Dropbox client pulls the modified file (shard) to this computer into the `.dropbox.cache` as a temporary file; moves the old version of the file (shard) under the Dropbox directory into the `.dropbox.cache`; finally moves the updated copy of the file, stored as a temporary file, into the Dropbox directory to its respective location with the appropriate name.
  - When a file (shard) was deleted on another computer, the Dropbox client moves the delete file into the `.dropbox.cache` directory on this computer.

  All of the above behavior of the Dropbox client epically broke combox. Commits `3d714c5` to `6e1133f`[7] fixed combox by making it aware of Dropbox's client behavior.

5.2.2.2 Demo

Demo of combox being used on two GNU/Linux machines can be viewed at `https://ricketyspace.net/combox/combox-2-gnus.webm`.

`lyra` (virtual machine) and `grus` (bare-metal) are the two GNU/Linux machines being used for the demo.

Description of what happens in the demo follows:

- (lyra) install combox.

- (lyra) run combox (test mode).

- (lyra) create file `walden.pond` with content "It must be beautiful there".

- (lyra) sync Google Drive using `rclone`.

- (grus) sync Google Drive using `rclone`.

- (grus) git pull latest copy of combox.

- (grus) install combox

- (grus) run combox (testing mode).

- (grus) verify that `walden.pond` was create on this machine.

- (grus) append 'Peaceful too.' to `walden.pond`.

- (grus) sync Google Drive using `rclone`.

- (lyra) sync Google Drive using `rclone`.

- (lyra) verify that the latest copy of `walden.pond` is there in the combox directory; it should contain 'Peaceful too.' in the last line.

- (lyra) append "I've a dream" to `walden.pond`.

- (lyra) sync Google Drive using `rclone`.

- (grus) sync Google Drive using `rclone`.

- (grus) verify that the latest copy of `walden.pond` is there in the combox directory; it should contain "I've a dream" in the last line.

- (grus) remove `walden.pond` from combox directory.

- (grus) sync Google Drive using `rclone`.

- (lyra) sync Google Drive using `rclone`.

- (lyra) verify that `walden.pond` is removed from the combox directory.

- (grus) open dropbox and Google drive accounts from the web browser.

- (lyra) create file `manufacturing.consent.` with content "Chomsky stuff?".

- (lyra) sync Google Drive using `rclone`.

- (grus) sync Google Drive using `rclone`.

- (grus) verify that `manufacturing.consent` was created in the combox directory.

- (grus) verify that the shards of `manufacturing.consent` were created on Dropbox and Google Drive through the web browser.

5.2.3 Testing on a GNU/Linux and an OS X machine

combox was run on a GNU/Linux machine and an OS X machine and a file was alternatively created/modified/renamed/deleted on one of the machine and it was verified if the respective file was also created/modified/renamed/deleted on the other machine. The GNU/Linux machine was a virtual machine (`lyra`) running Debian GNU/Linux stable; the OS X machine was on Mavericks (10.9) during the initial stage of testing, later it was upgraded to Yosemite (10.10). The node directories to scatter files' shards were the Dropbox directory and the Google Drive directory. The official Dropbox client was used to automatically sync files from the Dropbox directory to the Dropbox' server on both the GNU/Linux machine and the OS X machine; the official Google Drive client was used to automatically sync files from the Google Drive directory to Google Drive' server on OS X and `rclone`[6] was used to sync files from the Google Drive directory to Google Drive's server on GNU/Linux.

5.2.3.1 Issues found

- When a file was modified on another computer, on this computer combox assumed that first shard (shard0) will be updated first and also counted on the existence of the first shard (shard0). It was observed that the order in which the shards were updated were unpredictable on this computer and if the first shard (shard0) was stored in the Dropbox directory, it will momentarily disappear before the most updated shard becomes available in the Dropbox directory; this broke combox. This issue was fixed on 2015-08-25[8]. This issue is not got to do with the nature of the setup but it is related to the Dropbox's behavior elaborated in section 5.2.2.1.

- The official Google Drive client when it pulls an updated version of the file from Google Drive' server, instead directly updating the respective file on the computer, it deletes the older version of the file and creates the latest version of the file at the respective location in the Google Drive directory; this behavior of the Google Drive confused and broke combox. This issue was fixed 2015-09-06 by making combox under the official Google Client's behavior[9].

- When a non-empty directory was move/renamed on another computer, the old directory was not getting properly deleted on this computer; this was happening because the files under the directory being renamed were not deleted when it was time for `NodeDirMonitor` to `rmdir` the old directory. This issue again is not specific to the nature of the setup but was found while testing combox on this setup. This issue was fixed on 2015-09-12[10].

- It was found that `combox.file.rm_path` function failed when it was given a non-existent path to remove; this issue was fixed on 2015-09-12[11].

5.2.3.2 Demo

5.2.4 Testing with a USB stick as a node

combox was run on a GNU/Linux machine and an OS X machine and a file was alternatively created/modified/deleted on one of the machine and it was verified if the repsective file was also create/modified/deleted on the other machine. The GNU/Linux machine was a physical machine (`grus`) running Debian GNU/Linux stable; The OS X machine was on Mavericks (10.9). The node directories to scatter files' shards were the Dropbox directory, Google Drive directory and the USB stick (`ZAPHOD`, FAT filesystem). The official Dropbox client was used to automatically sync files from Dropbox directory to Dropbox' server on both the GNU/Linux machine and OS X machine; the official Google Drive client was used to automati-

cally sync files from the Google Drive directory to Google Drive' server on OS X and `rclone`[6] was used to sync files from the Google Drive directory to Google Drive's server on GNU/Linux; the same USB stick (`ZAPHOD`) was used on bothe GNU/Linux and Dropbox to store the third shard (shard2) of a file.

5.2.4.1 Caveats

- When a removable USB disk is used as a node, combox must be turned off before ejecting/unmounting the USB disk; combox does not expect a node directory to disappear when it is running, if the USB disk is removed when combox is running, then combox goes to a undefined state.
- When a file modified on machine A is synced to machine B, combox must be turned on first before turning on Dropbox and Google Drive clients and the shard in the USB disk needs to be "touched" for combox to detect that the file was modified on the remote computer and update the file locally on this machine.
- File rename/move does not work. To make it work, core functionality of combox must be re-written.

5.2.4.2 Demo

5.3 Stress testing

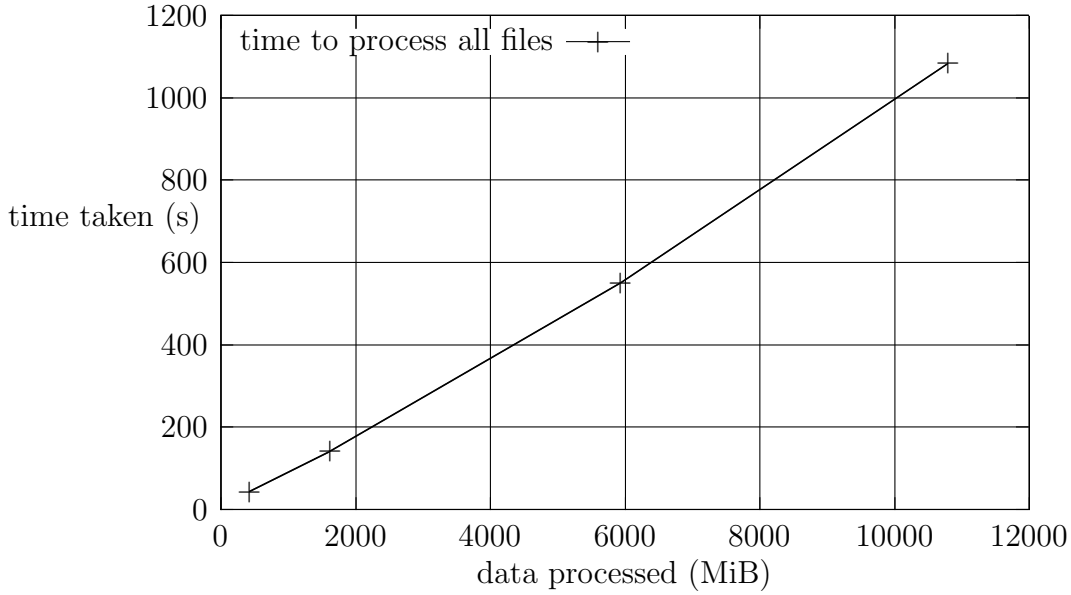

Figure 5-1: time to process all files

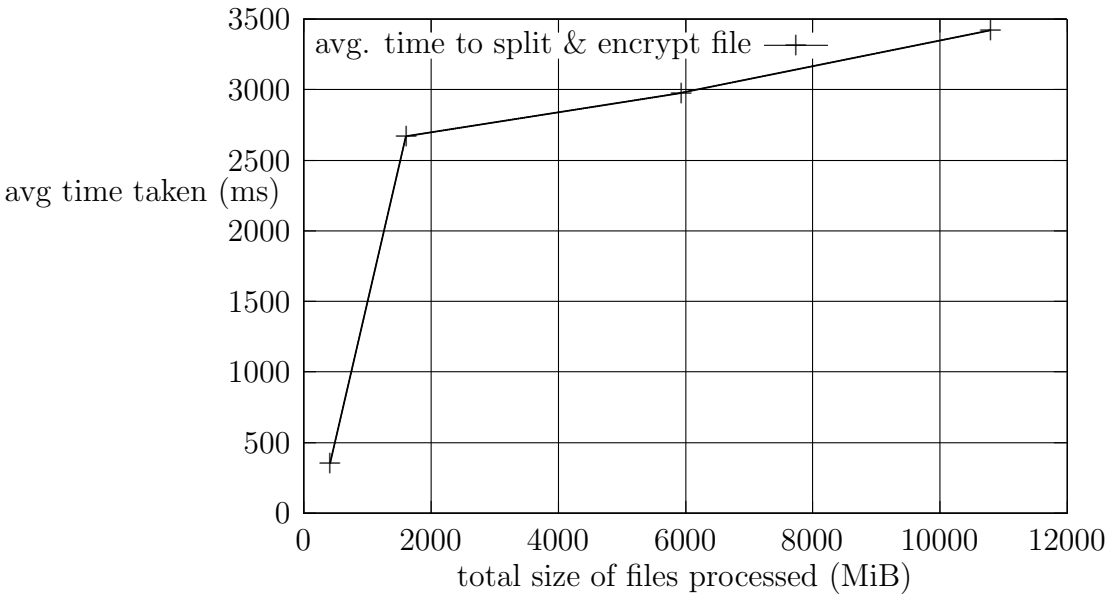

Figure 5-2: avg. time to split and encrypt

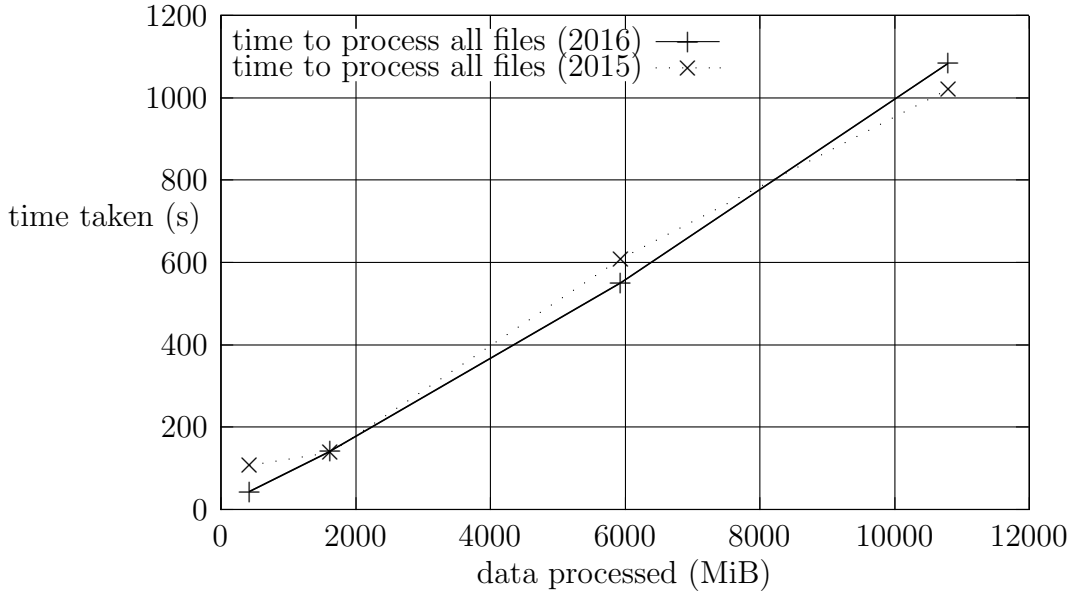

Figure 5-3: time to process all files - difference between 2015 and 2016

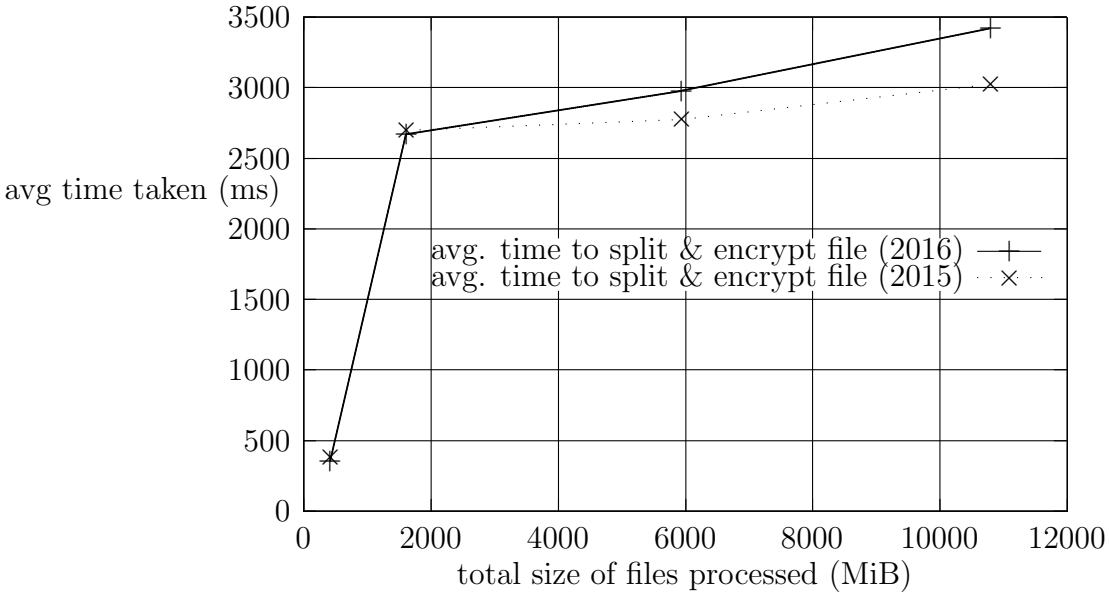

Figure 5-4: avg. time to split and encrypt - difference between 2015 and 2016

Chapter 6

Conclusion and Future Work

References

[1] J. Buxton and B. Randell, "Software engineering techniques," NATO Science Committee, Tech. Rep. p. 16, 1969. [Online]. Available: http://homepages.cs.ncl.ac.uk/brian.randell/NATO/nato1969.PDF

[2] "Nose - a nicer testing for python." [Online]. Available: https://nose.readthedocs.org/en/latest/

[3] "combox issue tracker (org-mode)." [Online]. Available: https://git.ricketyspace.net/combox/plain/TODO.org

[4] "Watchdog - python api library and shell utilities to monitor file system events." [Online]. Available: https://pythonhosted.org/watchdog/

[5] "combox - watchdog 'file create event' bug fix." [Online]. Available: https://git.ricketyspace.net/combox/commit/?id=8c86e7c28738c66c0e04ae7886b44dbcdfc6369

[6] "rclone - command line program to sync files and directories to and from google drive." [Online]. Available: http://rclone.org/

[7] "combox - git commits - dropbox client behavior fix." [Online]. Available: https://git.ricketyspace.net/combox/log/?qt=range&q=3d714c5..6e1133f

[8] "combox - git commit - shard modification fix." [Online]. Available: https://git.ricketyspace.net/combox/commit/?id=d5b52030348d40600b4c9256f76e5183a85fbb

[9] "combox - git commit - google client behavior fix." [Online]. Available: https://git.ricketyspace.net/combox/commit/?id=37385a90f90cb9d4dfd13d9d2e3cbcace8011e9

[10] “combox - git commit - bug six fix.” [Online]. Available: https://git.ricketyspace.net/combox/commit/?id=9d14db03da5d10d5ab0d7cc76b20e7b1ed552

[11] “combox - git commit - bug seven fix.” [Online]. Available: https://git.ricketyspace.net/combox/commit/?id=422238eb4904de14842221fa09a2b4028801afb