COMPUTERS & SECURITY XXX (2015) 1–17

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cose

Computers & Security

Network investigation methodology for BitTorrent Sync: A Peer-to-Peer based file synchronisation service

Mark Scanlon[*], *Jason Farina*, *M-Tahar Kechadi*

School of Computer Science and Informatics, University College Dublin, Belfield, Dublin 4, Ireland

ARTICLE INFO

Article history:
Received 1 February 2015
Received in revised form
2 May 2015
Accepted 13 May 2015
Available online xxx

Keywords:
BitTorrent Sync
Distributed storage
Peer-to-Peer
Network traffic analysis
Remote evidence acquisition

ABSTRACT

High availability is no longer just a business continuity concern. Users are increasingly dependant on devices that consume and produce data in ever increasing volumes. A popular solution is to have a central repository which each device accesses after centrally managed authentication. This model of use is facilitated by cloud based file synchronisation services such as Dropbox, OneDrive, Google Drive and Apple iCloud. Cloud architecture allows the provisioning of storage space with "always-on" access. Recent concerns over unauthorised access to third party systems and large scale exposure of private data have made an alternative solution desirable. These events have caused users to assess their own security practices and the level of trust placed in third party storage services. One option is BitTorrent Sync, a cloudless synchronisation utility provides data availability and redundancy. This utility replicates files stored in shares to remote peers with access controlled by keys and permissions. While lacking the economies brought about by scale, complete control over data access has made this a popular solution. The ability to replicate data without oversight introduces risk of abuse by users as well as difficulties for forensic investigators. This paper suggests a methodology for investigation and analysis of the protocol to assist in the control of data flow across security perimeters.

1. Introduction

Applications such as Evernote and Dropbox leverage the decreasing cost of hard disk storage seen in infrastructure as a service providers, e.g., Amazon S3, to provide data storage on the cloud to home users and businesses alike. The main advantage of services such as Dropbox, Google Drive, Microsoft Skydive (now OneDrive) and Apple iCloud to the end user is that their data is stored in a virtual extension of their local machine with no direct user interaction required after installation. It is also backed up by a fully distributed data-centre architecture that would be completely outside the financial reach of the average consumer. Their data is available anywhere with Internet access and is usually machine agnostic so the same data can be accessed on multiple devices without any need to re-format partitions or wasting space by creating multiple copies of the same file for each device. Some services such as Dropbox, also have offline client applications that allow for synchronisation of data to a local folder for offline access.

* *Corresponding author.*
E-mail addresses: mark.scanlon@ucd.ie (M. Scanlon), jason.farina@ucdconnect.ie (J. Farina), tahar.kechadi@ucd.ie (M.-T. Kechadi).
http://dx.doi.org/10.1016/j.cose.2015.05.003
0167-4048/

As Internet accessibility continues to become more commonplace and allows for increasingly faster access, it is not unexpected that many utilities that are intended for general use will aid in the perpetration of some variety of cybercrime. One attribute that is highly desirable by those contemplating illegal activities is the notion of anonymity and data security – especially the ability to keep data secure transfer secure from inspection while in transit. BitTorrent Sync (also referred to as BTSync, BitSync and bsync) is a file replication utility that would seem to serve exactly this function for the user. Designed to be server agnostic, the protocol is built on already popular and widespread technologies that would not seem out of place in any network activity log.

Each of the aforementioned consumer focused services can be categorised as cloud synchronisation services. This means that while the data is synchronised between user machines, a copy of the data is also stored remotely in the cloud. In recent headline news, much of this data is easily available to governmental agencies without the need of a warrant or just cause. BTSync provides the same synchronisation functionality (without the cloud storage aspect) and provides a similar level of data availability. The service has numerous desirable attributes for any Internet user (BitTorrent Inc, 2013a):

- Compatibility and availability – clients are built for most common desktop and mobile operating systems, e.g., Windows, Mac OS, Linux, BSD, Android and iOS.
- Synchronisation options – users can choose whether to sync their content over a local network or over the Internet to remote machines with no requirement for scripting or schedule management making this an accessible technology compared to existing options such as RSYNC.
- No limitations or cost – most cloud synchronisation services provide a free tier offering a small amount of storage and subsequently charge when the user outgrows the available space. BTSync eliminates these limitations and costs. The only limitation to the volume of storage and speed of the service is down to the limitations of the synchronised users machines.
- Automated backup – like most competing products, once the initial install and configuration is complete, the data contained within specified folders is automatically synchronised between machines.
- Decentralised technology – all data transmission and synchronisation takes place solely in a Peer-to-Peer (P2P) fashion, based on the BitTorrent file sharing protocol.
- Encrypted data transmission – while synchronising data between computers, the data is encrypted using RSA encryption. Under the BTSync API, developers can also enable remote file storage encryption (BitTorrent Inc, 2013b). – this could result in users storing their data on untrusted remote locations for the purposes of data redundancy and secure off site backup.
- Proprietary technology – the precise protocol and operation of the technology is not documented by the developer. There is debate over whether security through obscurity or peer code evaluation, i.e., open source, is better. Some enterprise security policies prohibit the use of open source applications as a result of the source code being open to inspection by those looking for flaws in the implementation. From the point of view of the consumer, BitTorrent Inc. have stated that they will not give access to traffic to any LEA without due process and the bespoke protocol makes casual eavesdropping or crawling less likely.

As a result of these attributes, BTSync has grown to become a popular alternative to cloud based synchronisation services. Less than a year after its release, the active user base had grown to over one million by November 2013, doubling to two million by December 2013 (BitTorrent Inc, 2013c), and to over ten million users by August 2014 (BitTorrent Inc, 2014). Due to this rapid growth and popularity the service will undoubtedly be of interest to both law enforcement officers and digital forensics investigators in future investigations. Like many other file distribution technologies, this interest may be centred around recovery of the data itself, proof of the modification of data or evidence of data distribution and enumeration of the recipients.

While BTSync is based on the same technology as BitTorrent for the transfer of files, the intention of the application is quite different. This results in a change of users' behaviours, as well as a necessary change in the assumptions an investigator should make. BitTorrent is designed to be a one-to-many data dissemination utility. The uploader usually does not care about the identity of the downloader and a single seeder can deliver data to a large number of unique peers over the life of the torrent file. Data integrity and transfer speed take precedence over privacy of data in transit.

BTSync on the other hand, is designed to be a secure data replication protocol for making a faithful replica of a data set on a remote machine. Data integrity is still highly prised but data privacy is now the top priority and speed-through-dispersion is sacrificed as a result. The files can only be read by users specifically given access to the repository. The advertisement of data availability is completely scalable by the owner with options ranging from restricting access to known IP addresses through to registration with a centralised tracker. Given the nature of the application, users are much more likely to know the operator of the remote site (this does not apply to secrets advertised online though that could be a point of commonality that would not necessarily have existed for pure BitTorrent clients).

1.1. *Aim and contribution of this work*

The aim of this work is to provide a reference for digital investigators discovering the use of BitTorrent Sync in an active investigation. However, it is hoped that the analysis presented may be of use to security personnel looking to detect and control the use of this protocol within their perimeter.

To accommodate these goals this work presents an analysis of the protocol and its network interaction. Activities undertaken to perform a synchronisation are presented and described at the packet level in order to facilitate both post mortem traffic analysis and to enable the development of feature based detection rules and deep packet inspection for Network Intrusion Detection Systems (NIDS) or firewall appliances.

Please cite this article in press as: Scanlon M, et al., Network investigation methodology for BitTorrent Sync: A Peer-to-Peer based file synchronisation service, Computers & Security (2015), http://dx.doi.org/10.1016/j.cose.2015.05.003