INFORMATION THEORY IN USB FLASH MEMORY DEVICE
ANALYSIS

1JUANCHO D. ESPINELI, 2JASMIN NIQUIDULA
Technological Institute of the Philippines 363 P. Casal St. Quiapo Manila
E-mail: juancho.espineli@lpu.edu.ph, jasniguidula@yahoo.com

Abstract- As the prolific use of USB flash memory drives increases, the call to validate issues regarding non-conformity of these devices also rises, moving away from such factors like as operating systems, hardware features, dependability and power utilization. This paper presents a different form of analysis on a subset of USB flash memory drives. While using various analysis tools live trace will be performed to capture error related to the issue. More over root cause analysis will also be conducted to find what really causes the error then find its relation to information theory particularly the “noise on the channel” [6].

I. INTRODUCTION

The extensive market for USB flash drives which is projected to exceed 555 million units by the year 2020 had shown the increasing importance of the technology in consumer applications [16]. Perhaps because of its small size and continuous increase in memory capacity we could easily conclude that this generation is now being swayed to move away from the conventional use of storage devices. At present flash drives enable it’s user to rapidly transfer data from one device to another without having a need of a computer network. It will also allow the user to run their software directly from the USB flash drive without compelling its user on a machine installation. With the superiority of USB flash drives as compared to the traditional random access memory not to mention it’s continuously enjoyed popularity, comes the need to further analyze its performance or non-performance. Since USB flash memory drives needed to use of USB device drives in order to work and with much sluggish USB counterpart as compared to the extremely high-speed interface normal RAM enjoys in order to the CPU, it will entail considerable latencies on communication channel and its transmission operations. Although there were previous researches in the analysis of USB flash drives, none of them focused solely on transmission and applied the concept of information theory in the technical sense. Information theory as it is used today returns back to the work of Claude Shannon and was initiated as a means to study and find solutions to the problems of communication or transmission of signals over channels [3]. In this paper testing will be performed to determine prior information that contributes to the decreasing performance that a subset of USB flash memory device may encounter. This will be in accordance to the theory of communication channel developed by Claude Shannon wherein with X will be input source and Y is the output. By monitoring the output Y we could anticipate a positive amount of information regarding the input source X. Although in particular communication, uncertainty about the input is increased [3]. In order to conduct a comprehensive comparison of USB flash memory devices, Advance USB port monitor will be used together with Microsoft Message Analyzer tool to display the packets sent, decode the descriptors, detect errors drivers and measure USB flash memory device performance. The main objective of this research is to explore reasons why USB flash drives were considered unrecognized which means it was not detected by the USB port of machine where the USB flash drive was inserted. Normally, when USB flash drive doesn’t work anymore there are two categories it might fall into. Either it has broken stems, and connectors, meaning the kind of recovery is applicable if you accidentally knock, bend, snap, jam, or otherwise made a physical damage to the drive, or any of the connectors and dead drive if it is unrecognized, undetected, underdone or must be formatted, inaccessible, which means it’s either the storage chips are carefully removed from the PCB board and placed in a special reader which are normally done by the manufacturer or experts [15]. This study will be different in the sense that it will look into second category of USB flash memory devices that are unrecognized and tagged to be unknown but physically undamaged but instead of the USB being brought back to the manufacturer it will be analyzed using different tools to investigate the possible cause. For a device to be tagged as unknown or unrecognized, it could fall to several events before the USB hub driver to enumerate a USB device as an unknown device. To know such this project will perform event tracing on USB flash memory devices using the Microsoft Message Analyzer tool. This will in turn help in the trouble shooting process in parallel with the investigation to be conducted on the issue of inconsistency encountered by a USB flash drive. While performing a root-cause analysis on this specific problem another goal is to relate the results of the experimentation to Shannon’s information theory specifically on communication which includes the “noise” on the channel. This will be completed by
II. RELATED WORKS

The USB which stands for Universal Serial Bus is already a customary in business for attaching peripheral devices to different types of computer [10]. While USB flash memory drives are tiny, movable electronic devices with memory like a piece of plastic almost like a size of a matchbox. It has customary USB interface which allows you to plug the device into most computers ports [5]. USB Flash memory devices has gained a lot of recognition in the previous decade, could be because of its capacity to store huge amounts of data, extremely fast access to data, stable memory, portable size, power minimal devices and economically priced [11]. As a result, USB is supported by a huge number of systems manufacturers today. It includes class codes terminologies for several device classes particularly device HUB, audio, storage device, printing devices, communications device, content security and imaging devices [13]. Upon using the class code, the capability of the device that was inserted in is identified and an suitable device driver is executed. Similar class codes permits the device driver software to hold up devices from various firms which conform with the code.

Current researches studied the USB flash memory drive by focusing more on the hardware features rather than the system including the software behind the device. One study proposed to create a USB flash memory drive similar to a pen. Advantage of which is that it work even without using a computer but challenge on that project is to how they will insert the OS to the pen drive itself so that it can still show the contents of the USB pen drive [2]. Though, the prime objective of the research which is to create a pen like USB flash memory drive model which can work even without using a computer system were achieved however it still needs to focus on the cost of the design. Another study was made this time disregarding the type of storage media during data analysis. It showed how a complete memory copy of USB flash memories can be created and what are the processes needed to transform the data extracted into a set up that can be read by common tools on forensic media analysis [11]. The aforementioned studies provide reasonable motivation for the testing to be pursued in this paper. While this paper will focus more on the USB flash memory device communication system and processes, it will also discuss in parallel the information theory behind the system of USB flash memory device during its non performing state. Most scientists agree that information theory began in 1948 with Shannon’s famous writing. In that paper, he provided answers to the following questions on what is “information” and how to measure it? [12]. It was during 1948 when C. Shannon wrote paper on “Mathematical Theory of Communication” on the journal of Bell Systems. There it was shown how to quantify information with precision, while the information media are still conversely united. Waves on radio, signals on the telephone, images and text or the entire mode of communication, could be encoded in bits. The said paper actually showed “The digital age’s structure” [1]. Another theory that proceeded is Shannon’s concept of communication over noisy channels as highlighted in figure 1-1. In that communication channel, possible inputs could be a sound of voice, email data, and contents of USB media, the channel would be the transmission media either wired or wireless, home network, cloud, and storage media though the main purpose is to replicate data input at the intended path including the noises while transmitting) [6]. For the purpose of this paper, the analysis and separation of the source and media decoder and encoder as previously described should not deter the performance of similar communication scenarios and that will be demonstrated using specialized tool such as the Microsoft Message Analyzer. Said tool will allow the concentration on each component USB flash memory drive independently of the other components while fully understanding the information theory behind the device.

III. METHODOLOGY

The methodology to be followed is also illustrated figure 1-1 and that includes Selection of subset, Diagnostic tool set-up, Execution of Test and Evaluation of results.
3.1 Selection of Subset
Subset of USB 2.0 flash drives were selected on this project. These subsets are known to be defective. The USB flash memory devices that were chosen are 2 Generic 2GB, 1 Nano Imation 1GB, 1 Nano Imation 2GB, 1 Jetflash 1GB, 1 Transcend 1GB, 1 Transcend 16GB, 1 Sandisk Cruzer 4GB, 1 Toshiba 2 GB, 1 CD-R king 2GB flash drives. These flash drives when plugged in into various computer system that runs on different platforms. These devices were considered to be erroneous if not unrecognized or malfunctioning. For this project Windows 7 operating system starter kit will be used during testing. Figure 1-2 below shows one of actual Generic USB 2.0 that was not recognized by the OS. Table 1-1 also enumerates the set of devices that were used during this project along with its manufacturer, capacity information and standard.

<table>
<thead>
<tr>
<th>Manufacturer</th>
<th>Capacity GB</th>
<th>USB Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generic</td>
<td>2</td>
<td>2.0</td>
</tr>
<tr>
<td>Generic</td>
<td>2</td>
<td>2.0</td>
</tr>
<tr>
<td>Nano Imation</td>
<td>1</td>
<td>2.0</td>
</tr>
<tr>
<td>Nano Imation</td>
<td>2</td>
<td>2.0</td>
</tr>
<tr>
<td>Jetflash</td>
<td>1</td>
<td>2.0</td>
</tr>
<tr>
<td>Transcend</td>
<td>16</td>
<td>2.0</td>
</tr>
<tr>
<td>Transcend</td>
<td>2</td>
<td>2.0</td>
</tr>
<tr>
<td>Toshiba</td>
<td>2</td>
<td>2.0</td>
</tr>
<tr>
<td>Sandisk Cruzer</td>
<td>4</td>
<td>2.0</td>
</tr>
<tr>
<td>CD-R king</td>
<td>2</td>
<td>2.0</td>
</tr>
</tbody>
</table>

Table 1-1 Flash Drive Specification

3.2 Set-up of Diagnostic Tool
Microsoft Message Analyzer (MMA) was setup in order to capture traffic, reduce communication channel noise and to expose at top-level some of the issues that occur at lower level not to mention the information that is not visible at first but is critical for quick analysis. Microsoft Message Analyzer was used because tool can capture the lower-layer messages which in return allowing focus on higher layer data of interest. The said tool also allows display of individual message summaries as well as high-level overview of trace statistics and trend. MMA would also confine specific data that we want to view through filtering. Advance USB port monitor will also be used to capture the data transfer rate of the flash drives during its testing.

3.3 Execution of Test
Once the Microsoft Message Analyzer is already set-up chosen USB flash memory devices will now be analyzed for traffic? Unlike other USB analyzer tools which captures previously saved trace here USB devices will be analyzed using a live trace. During this trace desired ports will be selected depending on USB standard of the selected USB subset. As mentioned, USB 2.0 will be used for all selected flash drives. Upon execution the trace messages will be shown on the main panel while details pane shows the detailed field information. Raw bytes can also be seen from the message data pane. USB subset will be tested to identify the root cause of each device enumeration failure. This will also determine why each of the USB subset was not recognized when attached on the USB port. To get preliminary information device property manager was opened to view the properties of device. It was certain that device enumeration failed from all selected USB flash memory device. Normally, upon hearing of a USB device, we think of it as a USB secondary device, but a USB device could denote a transceiver device used at the computer system ports. The customary standard then may refer to USB utility such as the capability Zip Drive, Printer, Scanner, Internet modem or other secondary devices such as a USB flash memory storage device [4]. Majority of its functions will have a series of buffers. Normal is eight bytes long. Each of them will have an endpoint. Endpoints can be coined as data sink or some sort of a source of communication among USB’s. Since busses are technically host centric, endpoints should be expected at the final end of the transmission channel of a USB utility [4]. If one of those fails an error should automatically be displayed by the operating system. As with the enumeration failure above indicates, it says that the device did not the proper identification string during the said enumeration and Windows could not see an INF that matches the device to for it to install a driver. Those were the reasons to further analyze the event in the log. Several series of events also pertains to the USB’s controller hosts, devices and hubs that were attached to the system during the trace. It can be categorized simply as summary events. Notice that the initial events of course will not describe activity of the driver. Since summary events would normally record the devices state from the beginning of a logging session as well as other events that will represents the scenarios on the bus plus the communication between client drivers with the system or state changes inside.

Since the goal of this study is to see specifically how the device enumeration failed and how the communication theory of Shannon affects the entire system. As the theory of communication by C. Shannon portrayed, it reveals that any type

Proceedings of Academics World 63rd International Conference, Manila, Philippines, 28th-29th April 2017
communication system that creates transmission can be divided into several parts and can be taken separately as individual models. Thus, the possibility of complete separation from design inputs from the design of the transmission channel is there. Shannon also contributed that his model had numerous applications apart from communication. Since all of these can be applied to computers, telephone design exchanges and to any other fields [1].

Currently, all communication systems are technically created from this model and this will become visible while using both tools mentioned. Thus, in using MMA we are able to make a trace and capture the errors relevant to the issue. Looking at the latest device numbers it clearly identified the most recently attached device. By expanding the group, it showed the device enumeration groupings for each device. It also showed how much the system knew for each device. In order for the tool to greatly reduced the noise on the wire as to coincide with Shannon’s theory regarding the “noise on the channel” MMA enables truncated parsing mode, which implement a limited parsing set to deal with data files that contain truncated messages, for example, a .cap file. This also advance the performance and creates a distinct analysis, new perspective by returning headers only for message types that are limited to a pared-down parser set, as described in supporting message truncation [8].

Earlier messages do not speak clearly why the devices were not recognized. In using MMA filtering tool it can diagnose only the errors types. As shown in figure 1-3 device report we can see that USB device encountered a USB Stall Events these indicates why the last attempt for the device enumeration failed.

IV. RESULTS AND DISCUSSION

Table 1-2 List of Tested USB Flash Memory Device including obtained results.

<table>
<thead>
<tr>
<th>Manufacturer</th>
<th>GB</th>
<th>USB Descriptor / Event</th>
<th>Data transfer rate (Mbits/sec)</th>
<th>Protocol Overtime (Count)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generic</td>
<td>2</td>
<td>Stall Packet</td>
<td>none</td>
<td>750</td>
</tr>
<tr>
<td>Generic</td>
<td>2</td>
<td>Stall Packet</td>
<td>none</td>
<td>735</td>
</tr>
<tr>
<td>Nano Injection</td>
<td>1</td>
<td>Stall Packet</td>
<td>none</td>
<td>705</td>
</tr>
<tr>
<td>Nano Injection</td>
<td>2</td>
<td>No response</td>
<td>none</td>
<td>0</td>
</tr>
<tr>
<td>Jetflash</td>
<td>1</td>
<td>Stall Packet</td>
<td>none</td>
<td>690</td>
</tr>
<tr>
<td>Transcend</td>
<td>1</td>
<td>No response</td>
<td>none</td>
<td>0</td>
</tr>
<tr>
<td>Transcend</td>
<td>16</td>
<td>Stall Packet</td>
<td>none</td>
<td>765</td>
</tr>
<tr>
<td>Toshiba</td>
<td>2</td>
<td>Stall Packet</td>
<td>none</td>
<td>725</td>
</tr>
<tr>
<td>Sandisk Cruszer</td>
<td>4</td>
<td>No response</td>
<td>none</td>
<td>0</td>
</tr>
<tr>
<td>CD-R king</td>
<td>2</td>
<td>Stall Packet</td>
<td>none</td>
<td>715</td>
</tr>
</tbody>
</table>

Table 1-2 USB Testing Results

Three parameters were selected to investigate the factors that could imply the USB flash drives non-recognition. These are USB Descriptor, USB Data Transfer Rate (M/bit per second) and Protocol overtime count. Since USB descriptor of a USB device represents the entire device. USB devices can only have one device descriptor. It specifies some basic, yet important information about the device such as the supported USB version, maximum packet size, vendor and product IDs and the number of possible configurations the device can have [4]. Data Transfer Rate this shows exactly how fast can the connection transfer data? The higher the number, the greater the amount of files you can transfer per minute [17]. USB Protocol (overtime count) unlike other similar serial interfaces where the format of data being sent is not defined, USB’s are made up of several layers of protocols. Though, higher level layers are more important since most USB controller I.C.s will take care of the lower layer, thus making it almost invisible to the end designer [4]. The fact that USB is a host centric bus. The host initiates all transactions. The first packet, also called a token is generated by the host to describe what is to follow and whether the data transaction will be a read or write and what the device’s address and designated endpoint is. The next packet is generally a data packet carrying the payload and is followed by an handshaking packet, reporting if the data or token was received successfully, or if the endpoint is stalled or not available to accept data.
Almost all of the defective USB flash drive registered the same result which is a stall packet on USB descriptor event parameter. A stall which is a type of packet indicating the endpoint has had an error while a NAK packet indicates to the host that the endpoint is working, but temporarily has no data to send [4]. Since most of the USB flash drives that has been tested registered a stall packet this signifies failure to communicate in terms of supplying the right information needed by the host to recognize the device. In the Data Transfer Rate parameter all of the USB flash drives had registered zero or no transfer rate for obvious reason that devices were considered to have defects prior to the start of test. Also notice the protocol overtime count which signifies lesser amount of protocol count generated from the USB generic flash drive as represented by USB spec on the bar chart. As compared to a normal USB peripheral device defective flash drives have shown lesser protocol count as depicted in figure 1-4 below.

![Figure 1-5 USB Device showing less protocol count](image1)

To exemplify how a typical USB flash drive performs in terms of data transfer rate, Figure 1-6 shows a sample Performance report of a SanDisk Cruzer Blade - USB Mass Storage Device for reference.

![Figure 1-6 USB Device with normal transfer rate](image2)

Notice in figure 1-6 the consistency of data transfer as represented by the small dots at the bottom of the graph. These signified normal data transfer rate from a functioning USB flash drive. According to a Microsoft report a device might report dissimilar information in its device descriptor when devices are connected to the host computer in a high speed capacity than when it’s connected in a full speed capacity. A device also must not alter the information contained in the device descriptor during the lifetime of its connection, including during power state changes [17]. This information also justified somehow the results of the test which relates to the theory of Shannon regarding his claim on noise on the channel. Since according to this report from Microsoft all devices must have constant speed and this is not always possible because all devices are also subject to wear and tear and by this happening power surges are also not inevitable.

**CONCLUSION**

USB devices provide several advantages either on storage along with its reliable data transmission. Though, USB flash memory drives are usually utilized as a type of media storage, their benefits also became numerous thereby making necessary to count not just their performance but rather understand also other unseemly incidents that may hamper its normal function. In order to determine dissimilar factors which may have an effect on USB flash memory devices, this paper tested several non-working flash drives and analyze them to determine its rather unusual performance. As the USB flash drives encountered problems underneath root cause analysis were performed on the issue using analyzing tools. During the process it has been observed that a stall message which also indicates a stall packet on the USB’s event descriptor parameter on the tested USB’s such as: 2 Generic, 1 Nano Imation, 1 Jetflash, 1 Transcend, 1Toshiba and 1 CD-R king subset justified the failure that caused the non-recognition of USB flash drives. Aside from these stall packet encountered among the majority of the subset which comprises seventy percent, results also concluded that data transfer rate must be present in order to verify if there is indeed a data transmission on a USB device. It also became helpful to check the amount of protocol transfers because packet transfers are generally data packets carrying the payload which supports a handshaking packet that creates the reporting if indeed data or token was received successfully. Protocol could also be used as a bench mark for comparing a functioning USB flash drive to a defective one. During this project there were also attempts to prove if indeed C. Shannon’s “noise on the channel” theory is factual. Upon exploration and through other report gathering means it became conclusive that communication can really suffer if not supported by appropriate speed and power. These things seemed to happen since all devices are subject to its own wear and tear that is why inconsistencies in electrical power can somehow affect the USB’s device performance. Nevertheless, further analysis can still be made in the future to support this claim.

**REFERENCES**

Information Theory in USB Flash Memory Device Analysis


***