The Forensics Aspects of Event Data Recorders

Jeremy S. Daily  
University of Tulsa  

Nathan Singleton  
University of Tulsa  

Elizabeth Downing  
Digital Forensics Professionals, Inc.  

Gavin W. Manes  
Digital Forensics Professionals, Inc.  

Follow this and additional works at: https://commons.erau.edu/jdfsl

Part of the Computer Engineering Commons, Computer Law Commons, Electrical and Computer Engineering Commons, Forensic Science and Technology Commons, and the Information Security Commons

Recommended Citation

DOI: https://doi.org/10.15394/jdfs1.2008.1044  
Available at: https://commons.erau.edu/jdfsl/vol3/iss3/2

This Article is brought to you for free and open access by the Journals at Scholarly Commons. It has been accepted for inclusion in Journal of Digital Forensics, Security and Law by an authorized administrator of Scholarly Commons. For more information, please contact commons@erau.edu.
The Forensics Aspects of Event Data Recorders

Jeremy S. Daily
University of Tulsa
600 S. College Ave.
Tulsa, OK 74104
918-631-3056
jeremy-daily@utulsa.edu

Nathan Singleton
University of Tulsa
600 S. College Ave.
Tulsa, OK 74104
918-631-3056
nathan-singleton@utulsa.edu

Elizabeth Downing
Digital Forensics Professionals, Inc.
401 S. Boston Ave. Ste. 1701
Tulsa, OK 74103
918-856-5337
Beth.downing@dfpinc.com

Gavin W. Manes*
Digital Forensics Professionals, Inc.
401 S. Boston Ave. Ste. 1701
Tulsa, OK 74103
918-856-5337
Gavin.manes@dfpinc.com

*to whom correspondence should be addressed

ABSTRACT

The proper generation and preservation of digital data from Event Data Recorders (EDRs) can provide invaluable evidence to automobile crash reconstruction investigations. However, data collected from the EDR can be difficult to use and authenticate, complicating the presentation of such information as evidence in legal proceedings. Indeed, current techniques for removing and preserving such data do not meet the court’s standards for electronic evidence. Experimentation with an EDR unit from a 2001 GMC Sierra pickup truck highlighted particular
issues with repeatability of results. Fortunately, advances in the digital forensics field and memory technology can be applied to EDR analysis in order to provide more complete and usable data. The presented issues should assist in the identification and development of a model for forensically sound collection and investigation techniques for EDRs.

Keywords: Event Data Recorder, Digital Forensics, Evidence Production, Civil Procedure, Crash Reconstruction

1. INTRODUCTION

Event Data Recorders (EDRs) can provide a wealth of data regarding automobile accidents, including the vehicle’s speed at the time of the accident, engine RPM at the time of accident, braking information, airbag deployment, and more. Increasingly, this information is being used to supplement physical evidence by accident reconstruction professionals. To date, most research concerning EDRs has focused on the use of data coming from the recorder and its relevance to physical events rather than the process of collecting data from it [3]. Given the attention to electronic evidence in the modern court system, this is a very important step of the process of gathering information for use in court. Therefore, independently assessing issues regarding data integrity, security, and reliability is needed to maintain the proper forensic qualities of the digital data stored on event data recorders.

Since there are no laws mandating the use of EDRs, their presence and capabilities vary widely between automobile manufacturers and models. Most EDRs are a part of a vehicle’s restraint system (i.e., the seatbelts or airbags) and help “decide” when the restraints are deployed. Note that an EDR is not necessary for an airbag system to operate, and some cars do not contain them. Regardless of whether the restraint systems were deployed, EDRs can record any of the following: (1) pre-crash vehicle performance data and system status; (2) accelerations during the crash; (3) safety restraint system use; and (4) driver control inputs. There are instances where a loss of power to the module can result in no data being available for a particular accident. As automotive and EDR technologies advance, additional data will be collected and new requirements will be mandated by regulatory agencies.

Once a crash has occurred, the data must be retrieved from the EDR module. All current methods of this data retrieval are proprietary: the Bosch Crash Data Retrieval Tool (CDR) is used on vehicles from GM, Ford, Chrysler, and certain partner companies, while Hexadecimal Translation Tools (HTT) are produced by the EDR manufacturers. For the purposes of this paper, only light vehicle EDRs are being discussed, since heavy trucks utilize entirely different systems and setups.

1.1 Modern Event Data Recorders

As of 2008, Bosch produces the only third party tool capable of obtaining and
interpreting data stored in EDRs. The three US automotive companies have licensing agreements in place with Bosch, and therefore have the most easily obtainable crash data. Data can be recovered on some GM vehicles as far back as 1994.

The Sensing and Diagnostic Module (which is GM’s air bag module) records crash related information by placing volatile memory within a data loop through which the information of interest passes. When an accident or other catastrophic event occurs, the computer automatically writes the last five seconds of data from volatile to solid-state memory, where it can be downloaded and investigated by proprietary cables and software. A good explanation of this process can be found in Chidester, et al [1].

Due to the increased inclusion of computers in automobiles for operational purposes, there is a large amount of untapped data that could be collected. Advances in memory technology also allow for greater storage capacities, which could result from an increased number of inputs or a larger time span of collected data. Furthermore, forensics professionals have identified areas in which digital systems controlling the vehicle could be infiltrated and compromised [9]. Detecting traces of such activity through the digital record could be a crucial component of an investigation.

Since many different types of EDR’s exist for light vehicles, concerns exist regarding standards for data recording, reporting, storage, and transfer. The Society of Automotive Engineers (SAE) has established the J1698 technical committee, “Vehicle Event Data Interface (VEDI)” to examine EDR standardization issues. Concurrently, the Institute for Electrical and Electronics Engineers (IEEE) joined with NHSTA to form working groups IEEE 1616 to discuss issues regarding tamper proof and crash proof electronics [7]. In addition to these voluntary standards, NHTSA has released rules regarding EDRs.

1.1 Event Data Recorder Regulations

The standards and definitions of EDRs are determined by the NHTSA’s Title 49 CFR Part 563, which was published as a Final Rule in early 2008. This rule requires that tools for accessing and retrieving the data be commercially available for model year 2011 vehicles. This rule also defines a minimal data set that must be recorded and the data elements to be stored.

Currently, there are no standards concerning the collection of information stored in other locations, such as the anti-lock brake system (ABS). The NHSTA has stated that it will consider other sources of data in vehicles once Part 563 has been finalized.

2. CRASH DATA RETRIEVAL (CDR) TOOL

The Bosch CDR is the only publicly available product capable of downloading data from an EDR. This system was developed in participation with GM, and has
licensing agreements with GM, Ford, and Chrysler. At present, the CDR is only capable of downloading data from select vehicles manufactured since 1994. Vehicles not included in this set can still be investigated through the examination of persistent memory, a common feature in Electronic Control Modules (ECMs). Many of these ECMs are used to control various systems in the car, such as antilock brakes, powertrain and engine management, airbag deployment, etc. Although not centralized, data from ECMs can be accessed and recovered for use in crash reconstruction.

The tools used in this process are all closed source, including the collection, interfacing, file formatting, and decoding procedures. Therefore, independent verification of the software, its applications, and decoding is difficult. No official statement has been made by Bosch in response to inquiries regarding some of the procedural questions that arise during use and experimentation with this tool. To demonstrate these complexities, the following section provides a brief introduction to the tool, its use, and where it seems data is interpreted. Issues arising from its use are discussed in Section 3.

Note that all of the following information represents an investigation performed on an EDR from a 2001 GMC Sierra 1500 pickup truck using the Bosch CDR 2.8 software. This vehicle had been involved in a head-on collision in 2003, and the following experiment was performed on the EDR which mimicked the process that would occur post-crash by an accident investigator. The data was retrieved from the EDR using the CDR Interface Module and all information was output to a .cdr file. This process was repeated several times, both in part and as a whole, and each step was analyzed with regard to the ability to repeat, verify, and authenticate the data. Note that vehicles with other EDRs or software versions may return different results.

### 2.1 Using the CDR Tool

There are several operations that must occur in the investigation of an EDR, beginning with its connection to a CDR Interface Module and a “collections” computer (see Figure 1). These connections are made through a combination of proprietary and standard cabling. Connection of the EDR can be accomplished either through the diagnostic ports under the dashboard (which is only possible if vehicle power can be restored), or through direct access. The latter typically requires disassembly of the car’s interior.
2.2 CDR and EDR Communication

Once connected, CDR software communicates with the serial communications port, checks for the presence of a CDR Interface Module and then begins a new case. Then, the investigator must enter case-specific information such as the Vehicle Identification Number (VIN), investigation date, etc.

After such information has been entered, the CDR software begins performing its functions, and data is exchanged between the CDR, EDR, and collections computer. The initial signal from the CDR software is the hexadecimal code 53 56 47 10, and the EDR responds with the code 88 59 91 17 00 00 77: the 91 17 sequence corresponds to the specific model and version of the EDR. This hexadecimal information dictates the cable to be used, specific commands needed, and setup requirements for the CDR Tool.

After the EDR model information is obtained, the CDR sends a dump command, which downloads the nonvolatile memory from the EDR into the CDR Interface Module. Completion of this process is indicated when the sequence D0 56 47 93 is sent to the CDR software. At this point, the information is stored in volatile memory on the CDR Interface Module. The CDR software then sends a command to ship approximately half of the downloaded data for processing: it is unknown why the data is downloaded in two separate stages. The following hexadecimal data is returned from the EDR:

```
EF D6 01 91 17 00 00 A7 18 41 53 30 33 34 30 4B
46 33 42 39 32 00 15 76 31 80 A3 A5 A4 F8 AC 00
03 A4 34 80 83 81 85 70 FF 00 FA FA FA FA FA
FA FA FA FA FA FA FF 02 00 00 FF FF FF
FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF
FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF
FF FF FF FF FF FF FF FF FF FF FF FF FF FF FF
FF FF FF 81
```
The beginning of file, EF D6 01, and the end of file, 81, are appended to the data in the CDR Interface Module. Following the appended header, the EDR type identifier code (91 17) is again displayed. After the first portion of the data is received, a request for the rest of the download is sent. The CDR Interface Module sends the second half of the information as follows:

```
EF B4 01 FF FF FF FF FF FF 80 00 00 FF 80 FE FF
BF FF FF FF FF FF FF FF FF FF FF 7C 04 03 01 01
02 00 00 00 00 00 00 00 FF FF FF FF FF 0A 10
00 61 70 70 6E 6C 6A 00 80 00 00 73 73 73 73 00
20 20 20 20 20 08 25 FE 00 00 00 04 00 FF FF
FF FF FF FF FF FF FF FF FF FF FF FF 98
```

Here, the sequence EF B4 01 is used for the beginning of file marker and 98 is the end of file marker, both of which are appended to the data by the CDR Interface Module.

The entire retrieval process, beginning with the CDR sending the EDR a dump command, is repeated twice more. Although the reason for the repetition is not known, this could represent an attempt to compare data due to discrepancies that occur between passes. Once the CDR software completes the retrieval process, it analyzes the EDR data and generates a report. This temporary file is automatically deleted unless saved prior to exiting the program.

### 2.3 CDR Analysis, Files, and Report

Once the EDR data is collected, the CDR software performs an analysis and generates a report in the form of a *.cdr or *.pdf file. The *.cdr file contains the data displayed in the actual report, formatting data, and error checking data in the form of a hash value and field size counts.

Three sources of data are contained within the *.cdr report file: user-entered data, computer-entered data (such as hash value, date and time, etc.), and EDR data.

User-entered data includes information such as VIN, investigator’s name, case number, comments, etc. With the exception of the comments field, user-entered fields contain a maximum of 64 characters. EDR-supplied data size depends on its storage capacity, and is indicated before the data appears by a size marker of two bytes.

The hex data contains up to three hash values: two for program verification and one used by the software to ensure data has not been altered. The former is the Reporting Program Verification Number and the Collecting Program Verification
Number, both of which are displayed in the CDR report. In this example, the two are identical since the same version of the software was used for collection and reporting. The latter is a hash value that appears toward the middle of the file itself and is not reported to the user.

The actual EDR data from a GMC Sierra 1500 (see below) appears later in the file and is preceded by a size field. Here, the hexadecimal data DE 00 computes to 222 bytes, which is the number of bytes stored on the EDR including the initial padding of six sets of zeros. The reason for the padding is unknown.

Two hash values are used to verify data which are created using a CRC 32 function. Based on experimentation, including attempts at data manipulation, it appears that the data is rehashed every time the CDR Tool is asked to open the document. The hashes are then compared, and if they are not the same, an error is expunged to the user and the program exits.

Although the .cdr file format appears to be simple, there is no public description of it and no method to review fields within the CDR software. Additionally, the current software version does not display any verification numbers, checksums, or hash values. Indeed, the hash value can only be found by looking at the native
format of the report with a hex editor. Such information has been recognized as necessary in properly executed digital forensics investigations when handling electronic evidence.

3. EDRS AS EVIDENCE

EDR data is typically only in cases where an automobile accident figures prominently, although sometimes this evidence can be included as a part of a larger investigation. This type of data is increasingly considered as a part of other types of investigations, since electronic devices such as EDRs are enjoying more prominence in court proceedings in recent history. Indeed, many EDRs can provide information relevant to other aspects of an investigation, such as GPS data, that can assist investigators outside of the realm of automobile accidents. However, current techniques for removing and preserving automobile data do not meet the standards for electronic evidence as outlined by the Federal Rules of Civil Procedure. The court system has generally recognized data from EDRs during trial.

There are a number of benchmark cases in which EDR data was ruled as admissible and reliable [4]. Many of these cases involved airbag product defect/negligence, including v. General Motors Corp where EDR data was admitted after a Frye hearing which was upheld by the appellate court [6]. Several of the airbag cases used EDR data to prove that non-deployment was proper, including Cansler v. General Motors Corporation [14], Batiste v. General Motors Corporation [15], and Sipes v. General Motors Corporation [16]. Several other important cases addressing the issue of EDR data reliability include criminal cases of manslaughter and homicide that occurred during accidents [4]. In most of these cases, including People v Muscarnera [11], Matos v State of Florida [12], and People v Hopkins [13], it was determined that a Frye hearing was not necessary and that EDR evidence is reliable as a matter of law and as it is generally accepted by the scientific community. Since EDRs have only recently been used in automobiles, their use as evidence in court cases has only occurred in the last ten to fifteen years.

Although most of the cases involving dismissal of EDR data have occurred due to process issues, i.e., the lack of a search warrant, the presence of discrepancies and lack of authentication for EDR data opens the door for acquittals or overturned verdicts. This GMC Sierra EDR experiment indicates unresolved issues with data authenticity, necessitating the use of “best evidence” practices until a better solution can be implemented.

As with any evidence handling, proper and well-documented chain of custody is a critical first step. A stopgap solution to many evidentiary issues with EDRs includes the corroboration of any electronic findings with physical evidence.

3.1 Missing and Uninterpreted Data

The information presented in the CDR report is both difficult to understand and
often inconsistent. First, the downloaded data is present in hexadecimal format, which proves to be of little use to the investigator. Additionally, the software’s conversion of the hexadecimal data into a human-readable format cannot be independently verified due to the program’s closed source nature. Also, the displayed hex data is incomplete.

Second, the memory register numbers are not sequential, and there are missing segments in the data. Registers are the physical locations in the memory chip where information is stored. The registers from the 2001 GMC Sierra 1500 contain 6 bytes of data, but investigation of other EDRs show that they have fewer. There is no reason given for this difference in the associated documentation, and there appear to be missing segments of these registers in the hex data that is output into the report. It is unclear whether the registers simply do not contain any information, which seems unlikely due to the lack of the FF sequence.

Additionally, the *.cdr report printout displays data not converted by the program, which is presumed to be proprietary manufacturer information (i.e., airbag deployment thresholds). When a sniffer was deployed to monitor data transfer from the CDR to the file, discrepancies were noted on each pass. The system’s method of choosing or inserting data is unknown, making it difficult to determine the effectiveness and repeatability of these results. Since repeatability of a particular process is a crucial component of any information being considered as evidence, this presents a problem.

3.2 Data Collection Standards

In comparing the hexadecimal data from the three data dump passes, discrepancies were found between each pass in the form of differing bytes. It was noted that multiple downloads performed in short order caused the number of discrepancies increased. The changed bytes were found in the first portion of the download from the CDR Interface Module.

The CDR software was run six different times with a “sniffer” in place in order to determine consistency of results. Each run contains three different passes, and the information sent to the CDR Interface Module was recorded for each pass of each run. The sniffer found discrepancies between the passes within a run. However, the information presented to the user in the final report was the same. This is likely due to the algorithm used by the program, which is difficult to determine due to the proprietary nature of these tools.

Surprisingly, the *.cdr file does not contain the actual three passes that are collected from the EDR. The implication is that the CDR software deletes the original EDR data once the “final version” of the report is complete.

All of these issues cast doubt on both the repeatability and the validity of data from any particular run of the software. Consistency is a key component of the Daubert standards for admissibility of scientific evidence in a court of law. As such, the aforementioned discrepancies are a cause for concern when being presented as
3.3 Multiple Locations for Data

EDRs are not the only source of valuable crash reconstruction data in automobiles. Indeed, vehicles may contain multiple Electronic Control Modules to control the different higher-level functions (i.e. ABS, Traction Control, and Power Train). Car manufacturers have taken both central and distributed approaches to storing such data. GM records all pre-crash information in the SDM, but Ford separates data into two different modules: the Restraint Control Module (RCM) and the Powertrain Control Module (PCM).

Ford’s RCM makes the airbag deployment decisions and executes necessary system diagnostics. In the event of an accident, the RCM will typically record acceleration data, cumulative velocity change (Delta V), seatbelt usage, and other restraint information: GM stores similar information in an EDR. Ford additionally records vehicle speed, brake and throttle usage, and other related information in a PCM, but only on post-2003 vehicles installed with Electronic Throttle Control (ETC). The primary function of the PCM is to run engine diagnostics while controlling fuel-air mixture and spark. In Ford vehicles, this data was only retrievable from the factory until the recent introduction of support in the Bosch CDR tool.

The Ford PCM continuously records data in a circular buffer capable of holding 25 seconds of data recorded in 200 ms intervals. When the 25 second limit is reached, the data is rewritten starting at the initial memory location. In the event of a crash, data is written for an additional 5 seconds, assuming power has not been interrupted.

Since the PCM has no accelerometer or air bag deployment logic, it relies on a bit called the “RDI_FLG” originating from the RCM. If the RDI_FLG bit is set to 1 then an airbag deployment has been commanded. If the RDI_FLG is 0, then the airbag status is not deployed or unknown. The data is not locked and can be overwritten upon the next application of power. If power loss occurs during a crash, all recording to the PCM will stop. The data is not locked and can be overwritten upon the next application of power.

Because of this, Ford warns investigators that PCM data can be lost if the vehicle ignition cylinder is turned to the on position, so investigators should never attempt to utilize the vehicle’s power system. This presents a challenge for digital forensics specialists since a very detailed accounting of the status of the PCM is required to generate acceptable evidence in court.

A complete examination and interpretation of the data contained in the Ford modules is available by sending the modules to Ford (at a cost). There, engineers retrieve the data and send back a report. Currently, authentication of the data is performed by evidence technicians, which includes matching the modules to the vehicle. Some PCM’s do not contain any vehicle identification information, which
may raise the scepter of doubt in a trial setting [10].

### 3.4 Unknown Methods

The final data produced in the report is passed through an algorithm prior to being displayed to the investigator. As stated at the front of the CDR report generated by the software:

> Once the crash data is downloaded, the CDR tool mathematically adjusts the recorded algorithm forward velocity data to generate an adjusted algorithm forward velocity change that may more closely approximate the forward velocity change ... The SDM Adjusted Algorithm Forward Velocity Change may not closely approximate what the sensing system experienced in all types of events.

There is no description of the algorithm or how it knows what is “more closely approximate” to what the vehicle experienced. The ability to determine speed at the time of a car accident is a key piece of data during crash reconstruction. The lack of information regarding the specific algorithm applied here prompts questions about the exactness of the determination made by the software, which could lead to questionable acceleration calculations.

### 3.5 Evidence Identifier

Initially it was presumed the CDR software used information from the VIN, World Manufacturer Identifier, Vehicle Attributes, and Model Year (MY) to determine what type and version of the EDR was being read.

However, it was discovered that any information could be entered as long as the World Manufacturer Identifier corresponded to a manufacturer supported by the CDR and followed the VIN formatting requirements (see Figure 2) [2]. This opens the door for incorrect information to be easily inserted into the fields that identify evidence.

![Figure 1 - VIN Breakdown for Vehicle Manufacturers Producing 500+ Cars per Year](image)

### 3.6 Clean Source Media

In the most egregious exception to evidentiary requirements, there are no indications that the CDR Interface Module memory is wiped between downloads. During experimentation, power was applied to the CDR Interface Module and a request was made for the data in memory before performing a download of an EDR. At initial power-up, it seems that the CDR Interface Module memory is set to unknown predefined data, which is available for download as authentic data.
4. FUTURE WORK

The aforementioned pitfalls of the CDR software can be addressed by applying longstanding digital forensics principals and methodologies to crash reconstruction. Although the collection of automobile data for the purposes of accident reconstruction is a relatively new field, many lessons can be taken from the evolution of the digital forensics process, as well as the recovery of black box data from airline accidents. It is relatively simple to apply the same process of creating and displaying a hash value for collected data and implementing the currently required chain of custody documentation to automobile data collection events.

The type of computers and amount of data available for automobiles varies widely when compared to the standard laptop or desktop computers, or for smaller digital devices such as cell phones and PDAs. Automobile records currently store only a minute amount of information, 1 kilobyte or less, which serves to simplify the length of the collection process. Many of the standard digital forensics tools could be used on automobiles, or less cumbersome programs could be implemented due to the small amount of data being analyzed.

Legislation has been proposed to mandate a common set of information recorded on all model year 2011 vehicles and later. In the event of such legislation, sound practices and procedures must be in place in order to most effectively transition to the widespread use of such information in legal proceedings.

The possibilities for future work in the field of accident reconstruction are significant. Performing these type of data-dump experiments for a variety of additional automobiles would likely lend both new questions and possibly some answers to data discrepancies. Identification of automotive systems beyond those identified by Bosch including EDRs and ECMs other than those used by GM or Ford would go a long way towards the ability to reconstruct a variety of accidents. Additionally, it is important to develop techniques to decode the data obtained from the different vehicle systems without the use of proprietary decoders.

The necessity to provide some kind of authentication for automobile data is critical for the use of this information in court. Many of the steps taken in this process require more significant data verification, which is one of the most important components of the future of automobile accident reconstruction.

5. CONCLUSION

Currently, digital evidence extraction from automobiles is not conducted with the same rigor as with more traditional digital forensics applications. This paper addresses some of these current shortcomings, specifically in the data collection technique. Lessons learned from the fields of airline crash investigations and digital forensics can provide models for the future of the EDR data handling in the accident reconstruction industry.
REFERENCES


[6] Bachman v. General Motors Corp, Electronic Citation: 2000 FED App. 0039P (6th Cir.)


[16] Sipes v. General Motors Corporation, 946 S.W. 2d 143 (Texas – App 1997).