Pregledni naučni rad *UDC 004.738.5:343.3/.7*

DIGITAL DARK AGES AS A MAJOR CYBER SECURITY THREAT

Aleksandar Mihajlovic [1], Ivica Stankovic[2] , Radomir Mihajlovic[3]
[1] Mathematical Institute of the Serbian Academy of Sciences and Arts, Belgrade Serbia,
e-mail: mihajlovic@mi.sanu.ac.rs
[2] University „Union – Nikola Tesla“, Faculty for Business Studies and Law, Belgrade, Serbia,
e-mail: ivica.stankovic@fpsp.edu.rs
[3] NYIT, New York, USA, e-mail: rmihajlo@nyit.edu

Summary: *Within this paper the notion of the "digital dark ages" is presented in terms of cyber security. In addition to elaborating on all issues relevant to this newly identified problem, certain new views and solutions not found in technical literature or popular publications are proposed. This paper begins with a unique approach to the concept of information and data codes taxonomy. The taxonomy is correlated with the problems of losing codes that are similar if not equivalent to the problem of denial of data attacks. Along with the risks of losing data due to data code loss, a loss of code processing software engines is also elaborated on. The paper considers the loss of processing engines as being analogous to denial of service attacks. The proposed solutions are based on observations made, requiring protection and preservation of both data codes and data code processing engines.*

Keywords: *information, document, codes, information black holes, digital dark ages*

1. INTRODUCTION

Many authors [1] and public speakers [2] [3] are sending alert messages about the threats of the so called digital black holes, *information black holes* and the *digital dark ages*. Recent news bombastically declares the emergence of the *information black hole* threat. The 21st century is destined to become an "*information black hole*" to future historians because no organized effort has been made to store digital records in a format that could be read and understood many years into the future.

The 20th century has started as the century of electricity and has ended as the century of electronics and data processing. As the 20th century progressed, the hard copy printed media (such as paper or celluloid film) have been slowly complemented and then slowly replaced by electronic data storage media. Initially, electronic data was stored in analog form using magnetic tape as the primary media. However, even after less than a decade of being stored, data recovered from the magnetic media showed tremendous loss of fidelity. For instance, old audio or video magnetic tape recordings were impossible to render, play back, with the initial original recording quality. The first alarm flags were raised [4].

In response to the rapid aging of analog electronic data storage media, digital methods of storage were hailed as the solution of all problems related to the losses of electronically archived content. However, just after a few decades into the digital age, a new issue emerged, an issue of losing data formats or data codes as well as the problem of losing computing hardware devices and software capable of interpreting data codes. The first problem, the problme of losing data due to defective media or loss of format documentation is referred to as the Denial of Data or DoD problem. The second problem is referred to as the problem of losing processing hardware and/or the software engine's capability of rendering data, is referred to as the Denial of Service or DoS problem. The DoS problem is a well-known cyber security problem created by malicious attacks on certain services or computing and communication infrastructure. T In DoS attacks service may be denied due to:

- The reduced processing capacity or disabled service providing engine,
- The total destruction of the engine, or
- The complete loss of the engine.

The first two DoS problems are caused mainly by the malicious cyber attackers while the last one is caused by the negligence of the engine owner. The last problem creeps slowly and transparently, emerging as a problem often too late to defend against.

2. INFORMATION, DATA AND CODES

In these discussions, *information* and *data* are strictly distinguished, where data is perceived to be a carrier of *information*. To be more specific *information* is defined as follows:

Definition 1: Information is the representation of the semantic content of data.
This definition implies presence of a human user, mentally capable of semantically decoding and interpreting data and extrapolating some meaning or conclusion based upon it, i.e., capable of using data to make a decision and act. To add to the term of "actionable intelligence" originally introduced and promoted as a double talk term by the US 66th secretary of state, Condoleezza Rice, it may be reformulated that *information* is "actionable intelligence." Actionable intelligence as "*information* that can be acted upon" is in a direct inverse relationship with *Definition 1*.. In fact, there is no *information* that is not actionable. *Information* related decision and consequent action are assumed to cover a wide spectrum; from the very simple to the very complex. Some trivial decision and action may be just a simple acceptance that interpreted data are understood, followed by the decision not to proceed further with data interpretation.
Depending upon immediate purpose or application, data may be encoded in many different ways, i.e., a number of different codes could be used. Upon careful examination of numerous code driven applications,a unique data relevant code taxonomy is proposed. It is proposed that all data related codes be sorted according to the list of the general code types. In other words any given code could be:

1. Original user data type and digital computing matching code, known as data code, (e.g., English letter or glyph conversion to binary symbols relies on the ASCII character code, and integer numeric value to binary conversion relies on the signed-integer code).
2. Minimal storage capacity or minimal transmission time bound code, in which case the so called source or compression data codes are present.

3. Efficient storage retrieval related code in which case the index data structure codes such as those used by data base management system engines are present, (Examples of such codes are ISAM or B+ Tree algorithm based data organization codes).
4. Communication media accommodating code in which case line or modulation codes need to be used, (Popular codes of this sort are NRZ, Bi-Phase or AMI codes).
5. Storage or transmission error bound codes when, error detection and/or correction channel codes are used, (Popular codes of this sort are parity, CRC or Read Solomon codes).
6. Original-data and data source identity protection codes that are also known as encryption codes, (Examples of such codes are one-way-hash, symmetric or asymmetric encryption and digital signature algorithm based codes), and
7. Aggregate multi-type data organization and final data presentation codes, the so called document codes, with document codes being:
   - Document structure or data layout codes, (e.g., SGML and XML), and
   - Document style codes, (e.g., HTML-attribute or CSS as document style codes).

In the code taxonomy, the first and the seventh type of a code; *data* and *document codes*, are semantic codes. As Figure 1 illustrates, semantic codes are the highest level codes among all itemized in the list above. Semantic codes are human user bound. These codes are initially and finally processed by the human user via s "*information* loading or offloading on/from the data."

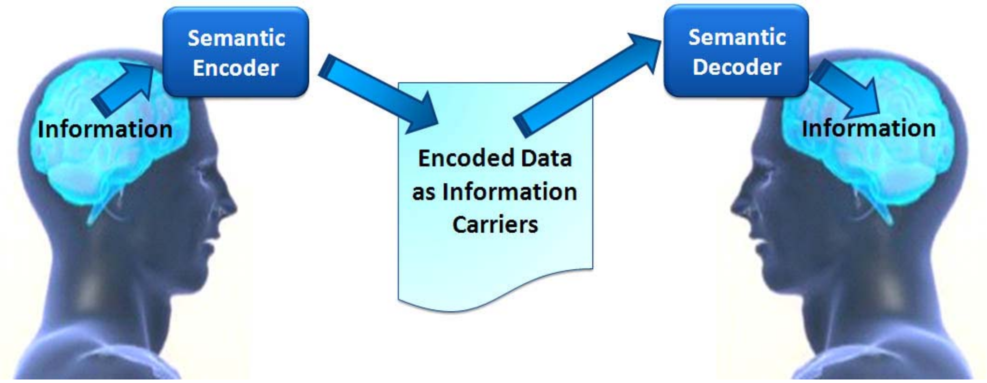


Figure 1. Semantic encoder/decoder is an *information/data* converter.

Original *data codes*, the first in the above list, are data input codes unavoidable at the computing systems data input point (e.g., keyboard). *Document codes*, the last in the above code list, are data presentation or data rendering codes which must be present at the output point of the user interface, (e.g., display or printer).

All simple data is considered using just one type of code with a very simple structure as trivial data. Trivial data rendered from trivial documents. Typical trivial data is plain ASCII encoded text. Figure 2 illustrates unstructured single byte ASCII character code next to the structured four byte signed integer and eight byte floating point numeric data codes. When rendered by the text editor graphical engine, ASCII character strings appear as a trivial document with the sole structure element being an end-of-line or new line character. Semantic text data enhancement is accomplished with the blank character acting as a word separator and sentence beginning and ending conventions. An upper case character indicates the beginning and a period indicates the end of a sentence. These semantic code symbols are processed by the human user. Text document structuring requires human encoding intervention. All documents including even simple text contain document format and style metadata characters embedded within a document. For instance, Web document may have a complex structure and style encoded to be used as browser instructions on how to render document semantic content.

Losing structure metadata code symbols or any of the individual document element data codes results in losing document semantic content even with all individual data elements preserved.

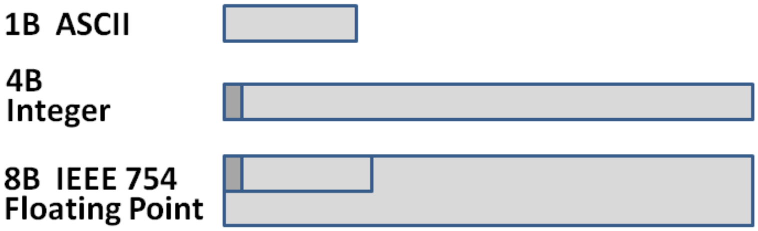


Figure 2. Simple data code examples; unstructured single byte ASCII character code, two-field structured four byte signed integer numeric code and three filed eight byte IEEE-754 floating point numeric “code words.”

3. DIGITAL DOCUMENTS PRESERVATION PROBLEMS

Everyone is familiar with the problems of software portability among operating systems and hardware platforms. However, more significant problems of modern computing are problems of digital data portability from one application to another. Digital data encoded by some basic and semantic codes, i.e., in some formats, can be ported between:

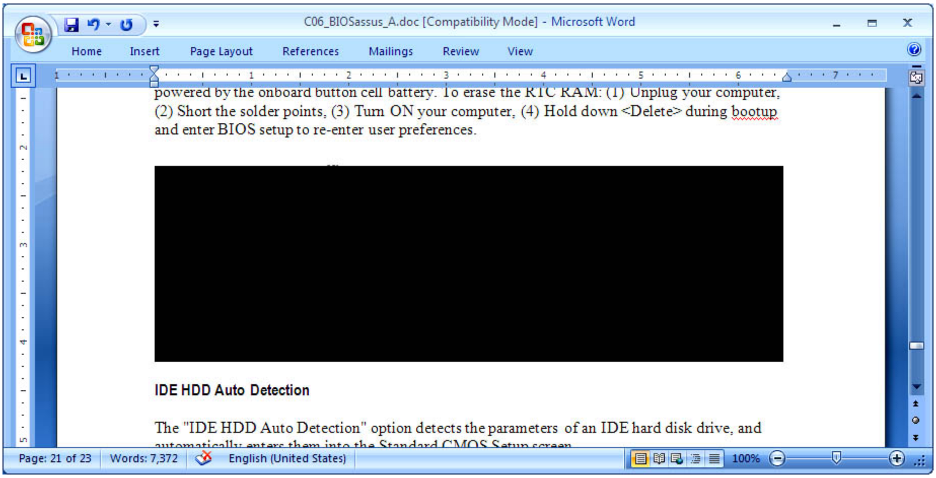
- Different contemporary applications, (e.g., from the Web browser to the Microsoft Word), or
- Different versions of the same application, (e.g., Microsoft Word 2007 and Word 2010).

These portability problems are most prominently visible with document processing programs of the Microsoft Suite. Microsoft document processing suite was delivered in many versions accompanied with many document formats. Figure 3 illustrates an example of a document encoded in the Microsoft Word 1997 and opened ten years later in the Word 2007 version. The originally embedded image element was not properly render using Word 2007 image decoder.

To better understand issues of the data and engine-code loss, the document is defined.

Definition 2: A document is a structured data object with the following basic attributes:

- Creation and recreation or rendering involve multiple data codes, and
- The structure of the original or rendered document is user oriented, i.e., document is expected to enhance and simplify *information* extraction process, in this taxonomy, the purpose of the document is effective semantic decoding.



powered by the onboard button cell battery. To erase the RTC RAM: (1) Unplug your computer, (2) Short the solder points, (3) Turn ON your computer, (4) Hold down <Delete> during bootup and enter BIOS setup to re-enter user preferences.

IDE HDD Auto Detection

The "IDE HDD Auto Detection" option detects the parameters of an IDE hard disk drive, and automatically enters them into the Standard CMOS Setup screen.

Figure 3. An example of the Microsoft Word 1997 document displayed using Word 2007 rendering engine.

The first attribute implies that documents contain data elements of multiple types, i.e., a document is itself of a multimedia data type. Each document element of distinct data type requires specific data encoding/decoding. Apparently one document element decoder has failed in the example shown in Figure 3.

As a structured or semantically encoded data super object, a document has two versions:

- Storage, communications and processing bound version, and
- User interface I/O, rendering or user bound version.

With these clarifications made some possible problems with digitized data may be analyzed.

4. DIGITAL DARK AGES

Currently, there are many large scale data digitization projects going on the corporate or national levels. The process of digitization is commonly based on a single input data code which is necessary for conversion of the hard copy archive or historical documents held on paper or micro film. Digitization efforts are motivated by the safer and more efficient document exposure and access as well as by the development of the e-government and e-publishing.

To preserve digital versions of unstructured data and structured documents, one is forced to consider complex backup systems and even the inverse process of the data de-digitization, meaning backing up digital data with hard copies. The newly identified threat that all digitized data are facing is slow fading from use and totally disappearing data codes with

data rendering engines. In 1997, Kuny has sounded the alarm about this serious threat as the threat of the "*digital dark ages*," [5]. Recently [2], wrongly advertised as the "father of the Internet," Vint Cerf, of Google made several public statements that he is worried that all the images and documents we have been saving on computers will eventually be lost.
In this paper, an attempt is made to place the problem of "*digital dark ages*" in a wider and more technical perspective. To avoid misconception and misuse of terminology formal definition of this new sort of dark ages are presented:

Definition 3: "Digital dark ages" are times with partial or total denial of digital services and digitized data access.
In light of this compact definition, an adjunct term of the "digital darkening" as partial and temporary loss of digital services and/or digitized data and as an event possible in the *digital dark ages* is considered. Internet access is considered as one of the most vital digital services today.

5. DIGITAL DARKENING

Digital darkening from inside could represent the physical state and the activities resulting in the reduction of digital processes in a well-defined and bounded portion of the cyber space referred to as the cyber domain. A cyber domain may cover one physically connected territory, such as a country or it may cover a logical territory with multiple disconnected physical regions. Only the former is considered within this paper. In one cyber domain covering a country, levels of digital activities may be reduced because of:

- *Educational reasons, (e.g., small number of educated people and English language users),*
- *Economic reasons and inability of the physical domain to financially support higher levels of digital activities, and*
- *Political reasons driven by the resolution of the political elite in charge of governing particular domain.*

Some authors point to the first and the second cause of low level of digital activities as "digital divide," [6]. The digital divide inhibits the use of computing and Internet resources by large sections of the population for reasons of low literacy, conservative culture or poverty. The digital divide exists among genders worldwide and schools contribute less than work in narrowing this gap. The lack of computing driven work is in direct relation to the digital darkening phenomena. Jobs that require computer use impose as mandatory learning and even self-education creating strong emotional drive to acquire digital age skills faster. Educational institutions, having both, educators and students mostly uninspired and not pressured, contribute significantly less to the erasure of the digital divide boundaries between genders and socio-economic classes [7].
It has been noticed that entire countries like Albania, Yemen and Mongolia suffer from digital darkening primarily due to educational and economic reasons, while countries like China and Iran have political reasons as the main obstacle to using Internet as a platform more open to the external world. Politically motivated internal digital darkening is well illustrated by the statement of Pan Shiyi, chairman of the Chinese SOHO corporation, made at the "CVW 2014," conference on Internet's impact on industry, held in Beijing on December 12, 2014, [8]. Pan Shiyi said: "I want to make an appeal to the government …

not to make us climb over the Internet great wall of China. The wall does not comply with the spirit of the free Internet!" The "wall" he was talking about was the "Great China Firewall," i.e., the great gateway data traffic semantic filtering system, limiting the spectrum of the document data available to Chinese Internet users.

The Chinese government, to which in a derogatory manner Western media refers to as the regime, officially exercises strict control and censorship over the Internet which has been a source of frustration among the public for many years. However, the Chinese government turns a blind eye to many technical solutions available to access even forbidden foreign data content.

Another prominent example of digital darkening from inside is the case of North Korea. Computing devices and Internet access are available in the North Korea, but due to both, internal educational, economic and political reasons, in a very limited amount. Internet access is available to private citizens for free but only with special authorization leaving government, educational organizations and state owned businesses as primary users. Frequent claims that hacker groups are attacking US based high profile targets are not realistic.

There are several examples of total digital darkening in recent history. For instance, Egypt shut off the country's Internet for five days in 2011 during the massive protests, Syria's Internet was shut off three times in 2011, Nepal and Burma have been briefly disconnected, and China shut off Internet access to the Xinjiang region during the political unrest in 2009, [9].

It seems that the digital darkening from the outside is far more dangerous and threatening to the world than the above mentioned darkening from the inside. Examples of the darkening caused by the outside forces upon some cyber domain are countries of Cuba, North Korea and potentially Russia. All of them have faced the USA as the outside force. The long lasting US laws on the subject of Cuba isolation kept for years all telecommunications between the USA and Cuba as illegal. In such a situation, with the economic stagnation, Cuba has been for years exposed to the digital darkening from the outside. Fortunately, digital darkening and economic under development has not prevented Cubans from having one of the finest educational systems in the region. In case of North Korea, several Korean Internet blackouts caused by the well-organized sophisticated cyber-attacks approved by the US government [10] and launched most likely from the South Korea and Japan.

With the ongoing Ukraine civil war and extremely strong stance of the USA, the media report on possibilities of having Russia in the total Internet darkness imposed from the outside. On September 30, the Kremlin made an official statement that Russian cyber forces are rehearsing responses should their esteemed partners, meaning NATO led by the USA, decide to switch Russia off from the Internet [11].

The position of this paper is that any cyber-attack resulting in Internet blackout or any other sort of digital darkening, from the outside, has to be classified as a crime, and regulated by the international laws and United Nations. Even in the worst political or military conflicts, all parties involved must be kept in contact with each other and the world.

6. PRESERVATION OF DIGITAL HERITAGE AND PREVENTION OF THE DIGITAL DARK AGE

Digitized data are not only scanned hard copy data but also digitized analog electronic data such as analog audio or video tape recordings. Saving codes and players of such data is necessary. After being digitized and stored in the digital media, data is only retrievable through the use of technology, i.e., via special computing devices referred to as rendering engines. Preserving of such engines and preserving of their compatibility with the original data codes used, is being omitted in most discussions on the topic of the "*digital dark ages*." Besides the preserving detailed documentation on all data formats, i.e., codes, a law is proposed, the preservation of digital data rendering engines. An effort has been made in that direction by the organizations such as forgottenbooks.org and archive.org. In the spirit of the presented idea that all legacy rendering engines, all legacy software be preserved, archive.org maintains an archive of old software [12] [13].

To preserve legacy hardware platforms and operating systems a project is proposed for developing and making publically available any virtualized versions of legacy machines. Virtualized machines and sand boxed software is easy to run on powerful modern workstations. One of the projects of this kind is the Cygwin project [14]. Cygwin emulates the once very popular MS-DOS operating system and legacy PC platform. It is surprising that Microsoft does not invest any effort in that direction and does not maintain a library of the original documentation of their past products.

7. SUMMARY

The purpose of this paper is to provide analytical framework for further work on the problems defined as *digital dark ages* and digital darkening. The complementary purpose is to propose a general solutions guideline to be applied in order to minimize the risks of losing data and services due to the age related transparent disappearance of the available computing resources and capacities. The problem of digital darkening and having aggressive parties exercise digital disconnect type of the cyber-attacks is addressed. It may be interesting to mention that the Kingdom of Serbia through the entire First World War with Austro Hungarian and German empires has never failed to make scheduled loan payments to the banks in these empires as well as to respond to any communication message addressed to the Serbian government from the mentioned empires. Such a gentlemanly conduct should be maintained in any current or future conflict among nations.

The opinion that this paper expresses is that relaxing the external cultural and political pressures and demands to rapidly transform the way of living and the way of doing business will relax the defensive posture of countries like China and North Korea and so reduce digital darkening trends from the inside due to political reasons.

As the legacy device and software documentation is rapidly disappearing and as the people that knew about them are leaving us too, we are making an appeal that all companies that have ever made an impact on the history of computing and data encoding, act urgently.

Acknowledgement

The work presented in this paper has been partly supported by the Serbian Ministry of Education, Science and Technological Development, under Project No. - III44006, of development of new *information* and communication technologies, based on advanced mathematical methods, with applications in medicine, telecommunications, power systems, protection of national heritage and education.

BIBLIOGRAPHY:

[1] Kurt D. Bollacker, "Avoiding a Digital Dark Age," American Scientist, Computing Science, Volume 98, 2010. pp.106-110. www.americanscientist.org/issues/pub/avoiding-a-digital-dark-age/1

[2] Pallab Ghosh, "Google's Vint Cerf warns of 'digital Dark Age', BBC News science & Environment, 13 February 2015, www.bbc.com/news/science-environment-31450389

[3] Steve Connor, "Vint Cerf: The 21st century could become a digital black hole," The Independent, 13 February 2015 http://www.independent.co.uk/life-style/gadgets-and-tech/news/vint-cerf-the-21st-century-could-become-a-digital-black-hole-10045596.html

[4] A. Evans, "Sound Ideas: Music, Machines and Experience," 121–122; 2005, Minneapolis, MN, University of Minnesota Press. xenopraxis.net/readings/evens_soundideas.pdf

[5] Terry Kuny, "A Digital Dark Ages? Challenges in the Preservation of Electronic Information," 63rd IFLA General Conference, Copenhagen, Denmark, August 31 - September 5, 1997. http://archive.ifla.org/IV/ifla63/63kuny1.pdf

[6] Book Chapter. Hacker, K, Mason, S. and Morgan, E. , "Digital disempowerment," In Emma Rooksby (Ed.) Information Technology and Social Justice, Idea Group, Inc., 2007. http://www.igi-global.com/book/information-technology-social-justice/582

[7] Jan van Dijk, Kenneth Hacker, "The Digital Divide as a Complex and Dynamic Phenomenon, " The Information Society, Copyright c Taylor & Francis Inc., pp. 19: 315–326, 2003, http://web.nmsu.edu/~comstudy/tis.pdf

[8] Fang Xiao, Lu Chen. "Member of Chinese Business Elite Calls for an End to Internet Blockade," Epoch Times, Chinese Regime, HK Current Affairs, December 15, 2014 http://www.theepochtimes.com/n3/1144733-member-of-chinese-business-elite-calls-for-an-end-to-internet-blockade/

[9] Luke Johnson, "Explainer: Can Russia Disconnect From The Internet?, Radio Liberty, October 02, 2014. www.rferl.org/content/can-russia-disconnect-from-the-internet/26617176.html

[10] Chris Strohm, "North Korea Web Outage Response to Sony Hack, Lawmaker Says," Cyber-security, Bloomberg, March 17, 2015,

[11] Анастасия Голицына, "Совет безопасности обсудит отключение России от глобального интернета," Статья опубликована в № 3678 от 19.09.2014 под заголовком: Суверенный интернет. www.vedomosti.ru/politics/articles/2014/09/19/suverennyj-internet

[12] Historical Software Collection, 2015. https://archive.org/details/historicalsoftware

[13] The Internet Archive Software Collection , 2015. https://archive.org/details/software

[14] Cygwin project, https://www.cygwin.com/