DATA STORAGE AND TRANSMISSION CONVERGENCE CONCEPT

Tomasz Bilski

Poznań University of Technology

Pl. Skłodowskiej-Curie 5, 90-965 Poznań, Poland

Tomasz.Bilski@put.poznan.pl

ABSTRACT

Computer systems process, store and transmit data. Seemingly, storage and transmission are different. Despite the difference, the properties of services are similar - effectiveness, security, reliability. The requirements are accomplished with the same methods: data structure, coding, compression, redundancy. The convergence between the services is noticeable. In the paper, some concepts, such as how to utilize the convergence to optimize the system performance, are presented. For example, in a file transmission procedure: sectors of disk are read, their checksums are examined and a file is assembled; communication software divides the file into packets, assigns new checksums and transmits data; on the receiver side the checksums are examined, the packets are assembled into a file and forwarded to the storage system; during write process a file is divided into sectors; sectors are given checksums and stored. In the paper, more effective scheme is proposed. First a disk sector is read and its checksum is examined. Then the sector is placed in the packet data field and the sector checksum is placed in the packet checksum field. The packet is transmitted. On the receiver side the sector and its checksum are stored. The unification of other elements of storage and transmission such as: channel coding, compression technique is also evaluated in the paper.

Key Words: knowledge and information management, Internet, information technology.

1. INTRODUCTION

1.1. Storage versus transmission

Computer systems perform three main tasks. They process, store and transmit data. Seemingly, the tasks of storage and transmission are different. Seemingly, the two tasks are performed with a use of diverse software and hardware technologies. But, the basic properties of these services are the same.

Data storage process may be described as data transmission through time. In this case, “the sender” and “the receiver” are usually not separated. On the other hand data transmission process may be described as short-lived data storage in the communication medium. The “write” and “read” process are performed by two separated and distant elements. In both cases the processes are based on the changes of the medium states. The changes of the states are stimulated by the write/send element and sensed by the read/receive element.

Both of the services should be effective, secure and reliable. The requirements are usually accomplished with a use of the same general methods and similar or identical particular techniques. Among them are: physical processes, data organization, multilayered structure, coding, ciphering, compression, redundancy. Occasionally, the solution of the storage problem is introduced to the transmission systems and vice versa. There are many examples of the convergence. For example, PRML (Partial Response Maximum Likelihood) signal analysis algorithm has been primarily developed for long range communication (it has been used by Viking spacecraft). Nowadays, the algorithm is implemented in all contemporary hard disks.

The question is, if the noticeable convergence between storage and transmission could be better utilized to further optimize the entire system performance? In the past, transmission phenomenon has been used for data storage. Similarly storage media have been used for data transmission. In some circumstances, even today, physical transport of removable media volumes remains an economical procedure.

1.2. Transmission used for data storage

It should be mentioned that, in the past there were computer systems that used transmission phenomenon for data storage. In the first years of computers, acoustic storage devices and magnetostrictive devices were used. The acoustic device consisted of an acoustic delay lines. Stored data have been represented by a sequence of acoustic waves cyclically transmitted between the two ends of a tube filled with air or mercury. Data from the computer was sent to the piezocrystal transducer at one end of the tube, the piezo pulsed and generated a small wave in the mercury. The wave was subsequently propagated to the far end of the tube, where it was read back out by the other piezocrystal transducer and sent back to the computer. In this way the pattern of waves sent into the system by the computer could be kept circulating as long as the power was applied. The computer would count the pulses by comparing to a master clock to find the particular bit it was looking for. One mercury delay line had a data capacity of a few kilobits.

Another type of transmission-based storage was magnetostrictive delay line. Mercury tube had been replaced with 10's of meters of wire. The wire was wound into a loose coil and pinned to a board. Transducers used the magnetostrictive effect; small pieces of a magnetostrictive material, typically nickel, were attached to either side of the end of the wire, inside an electromagnet. When bits from the computer entered the magnets the nickel contracted or expanded (based on the polarity) and twisted input end of the wire. The resulting tortional wave then moved through the wire. Unlike the acoustic wave however, the tortional wave was considerably more resistant to problems caused by mechanical imperfections. Due to their ability to be coiled, the wire-based devices were built as long as needed, and held more data. The time needed to find a particular bit was longer as it travelled through the wire, and access times on the order of 500 μs were typical.

1.3. Storage media used for data transmission

There are two types of storage media: irremovable and removable. Irremovable media such as hard disk are written and read in the same location. Removable media such as tapes, optical disks, diskettes, flash modules may be used to physically transfer data from one place to another. This way of data transmission was popular before computer networks were widely available.

Today, in the case of large volume of data, the speed of such transmission is greater than the speed of such transmission technologies as V.90 (modem 56 kbit/s), ISDN, ADSL. But obviously the speed of such transmission is not comparable with the throughput of modern computer networks (table 1). Even in the mid 1990's there were opinions that it has been economically justified: "*although significant advances in electronic data transfer communication networks can be expected in the next decade, because of band width limitations and telecommunication costs, data interchange via physical transport of removable media volumes will remain the most economical procedure for many applications*" (Gniewek, (1995)).

Table 1. 10 GB volume transmission time for 100 km distance

Technology	Approximate time
Modem 56 kbit/s	400 h
ISDN BRA 128 kbit/s	180 h
ADSL 2 Mbit/s	11 h
Physical transport of storage medium	2 h
Gigabit Ethernet 1 Gbit/s	1,5 min

2. STORAGE AND TRANSMISSION CONVERGENCE

2.1. Storage and transmission resemblance and differences

There are many general similarities between storage and transmission processes (figure 1). The requirements, such as performance, reliability, security are usually achieved by the same methods and techniques that differ only in some details. In the chapter we will discuss, the following aspects of storage and transmission: system architecture, physical processes, data organization, coding, modulation, redundancy. We will focus on the resemblances of the two systems and marginally note the differences.

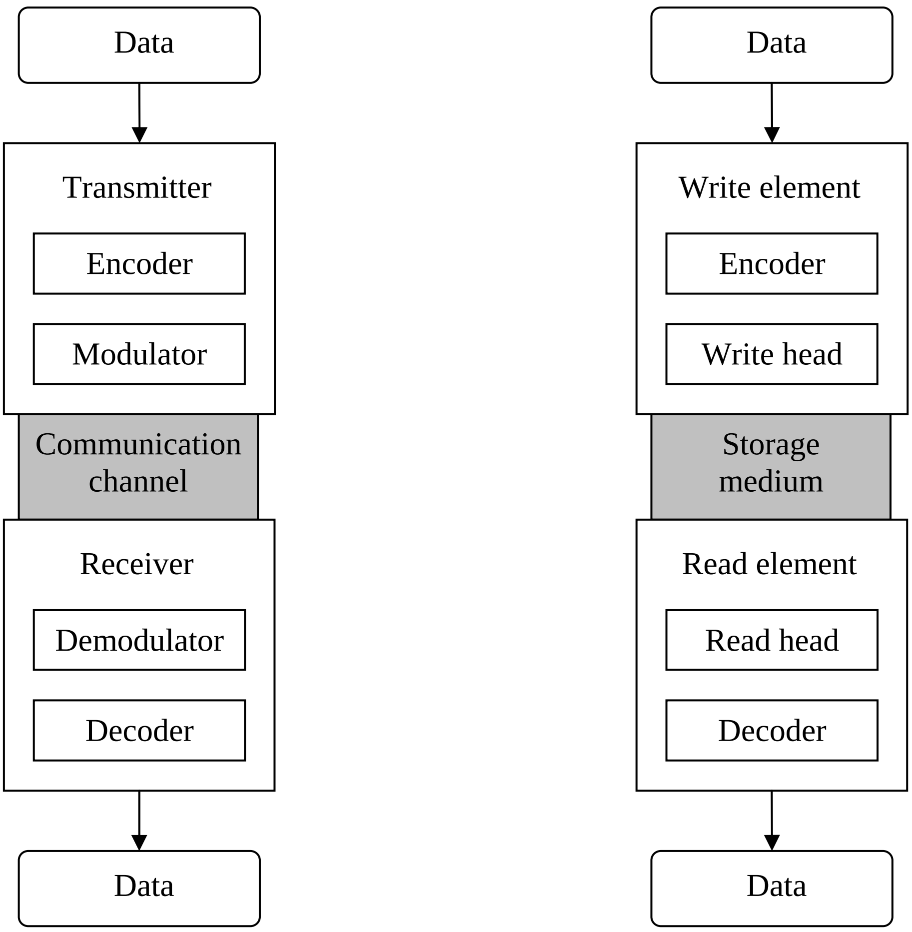

Figure 1. Communication system and storage system comparison

2.1.1. System architecture

Multilayered, hierarchical transmission models such as ISO/OSI model and TCP/IP model are widely implemented in computer networks. Corresponding models had been created for data storage systems. The most common known storage system model is OSSI (Open Storage Systems Interconnection) represented by IEEE 1244 standard set (Cole, 2000).

2.1.2. Physical processes

Storage and transmission processes are based on the changes of the medium states. The changes of the states are stimulated by the write/send element and sensed by the read/receive element. Several physical phenomena are used in the processes. Data transmission systems use:

- electrical signal propagation in a wire,
- laser impulse propagation in a fiber or space,
- electromagnetic field modulation in radio transmission.

Storage systems use:

- electrical charge (flash memory),
- changes of optical properties,
- magnetic field modulation.

In order to transmit data, pulses of “high” and “low” voltage are sent across communication wire. In order to store data in semiconductor flash medium, changes are induced in the electrical charge of the transistors.

Write process in optical media as well as send process in optical fiber are based on laser light modulation. The write head and the laser transmitter generate laser impulses, which are then stored in a form of changed optical properties of the storage medium material or propagated in a fiber (or space). The read and receive processes are based on detecting the laser light changes. Furthermore, WDM (Wavelength Division Multiplexing) used to multiple communication channels in a single optical fiber is similar to data multiplexing in hybrid disks and holographic storage modules. Hybrid, optical disk uses two data layers one layer may be read with 780 nm laser in a CD reader, second layer may be read with 635 nm laser in a DVD reader. Holographic memory modules use laser impulses with different phases to write/read different pages of data stored in a single module. The main difference between the storage and the transmission processes rests on the laser wavelength used in both processes (table 2).

Table 2. Laser wavelength used in transmission and storage systems

Application	Wavelength (nm)
Data transmission	
Optical fiber	980
	1310
	1550
	1625
Data storage	
CD	780
DVD	635
LOTS	532
Blu-ray	405

Write process in the magnetic media as well as send process in the wireless radio networks are based on electromagnetic field modulation. The write head and the radio transmitter generate electromagnetic field, which is then stored in a ferromagnetic material or propagated in space. Similarly, read and receive (in the wireless networks) processes are based on detecting the electromagnetic field changes.

2.1.3. Coding and modulation

The fundamental part of any communication/storage system is the channel/medium, which is the physical element by which information is transferred/remembered. In nearly all cases some devices are required to convert binary data into a suitable form for transmission/storage and then back into a form that is comprehensible to the end-user. This adaptation is done by coding and modulation at the transmit/write side, and respectively by decoding and demodulation at the receive/read side of the processes (Burr, 2001).

One of the main tasks in coding and modulation is to provide read/receive bit synchronization. Reading/receiving device should be able to look into bit values in the appropriate moment of time. The synchronization problem may be solved by special synchronization line in the communication channel (synchronization path in the storage medium), synchronization bits or by a self-clocking code. The synchronization bits are used in such technologies as: Ethernet, FDDI, HD, MO, DLT, CD, DVD. The third solution is generally more valuable and is used both in storage and transmission.

The arbitrary input sequences need to be encoded into sequences that satisfy the synchronization constraint. The code must provide sufficient signal transitions to enable reliable clock recovery during receiving/reading. An encoding model that is commonly used is that of a finite-state encoder, where the input binary sequence is divided into blocks of a fixed length, and each block is mapped in a stateless or state-dependent manner into a channel codeword of another (and greater) length. The sequence of generated channel codewords forms a word that satisfies the constraint. An obvious requirement is that it should be possible to decode (reconstruct) the input binary sequence from the output channel codewords.

RLL (Run Length Limited) codes are usually used for the purpose. The run length is defined as the number of identical contiguous symbols ("0" or "1") which appear in the channel codewords. Channel symbol "1" changes the state of medium/communication channel and "0" doesn't change the state. RLL code is defined by the shortest and the longest run length that appear. These two parameters are often given in the form (d, k) pair, where d is the minimum and k maximum number of "0" channel bits between two successive "1". For example: RLL (2,7) is a code, in which every two consecutive "1" are separated by not less than 2 and not more than 7 bits "0". RLL codes designed for digital transmission usually have a parameter d set at 0. The preferred codes for data storage usually have a parameter d of 1 or greater (table 3). RLL codes with $d>1$ have additional advantage. Two consecutive changes of state are separated by 1 or more signal elements – this decreases intersymbol interference.

Table 3. RLL codes used in transmission and storage systems

Application	RLL code
Data transmission	
FDDI	RLL (0,3)
Ethernet 100 Mbit/s	RLL (0,3)
Ethernet 1 Gbit/s	RLL (0,5)
IrDA 16 Mbit/s	RLL (1,13)
Data storage	
Blu-ray, LTO	RLL (1,7)
QIC, DLT	RLL (2,7)
CD, DVD	RLL (2,10)

Another important feature of a code is DC balance. To accomplish DC balance the average number of 1s and 0s in the serial stream must be maintained at equal or near equal levels over every span of length N. The value of N and the degree of balance obtained vary with the coding. DC-balanced codes are used in both communication and storage systems (table 4).

The advantage of DC balance in the transmission is the lack of informational content in the DC level, meaning that it is possible to strip it out (with a filter) and re-insert it later without violating the integrity of the signal. In the storage systems a DC-balanced code makes it possible to use simple servo systems that extract tracking information from the data track without any specific additional consideration and circuits.

Table 4. DC-balanced codes used in transmission and storage systems

Application	DC-balanced code
Data transmission	
FDDI	4B5B
Ethernet 100 Mbit/s	4B5B
Ethernet 1 Gbit/s	8B10B
Data storage	
CD	EFM
DVD	EFM+

2.1.4. Redundancy

In both (storage and transmission) processes data degrades. In the storage systems data degradation is relative to the storage time. In the transmission system data degradation is relative to the transmission distance. Likelihood of error increases as more bits are packed onto a square centimeter of medium and as communication transmission speed increases. In both cases redundancy is used to minimize error rate. A typical form of redundancy is a CRC code added to the transmitted/stored data. The same particular CRC codes are used in data transmission and storage systems (table 5).

Table 5. CRC codes used in transmission and storage systems

Application	CRC code
Data transmission	
ATM	CRC-8
Ethernet	CRC-32
FDDI	CRC-32
Fibre Channel	CRC-32
Token Ring	CRC-32
SDLC/HDLC	CRC-CCITT
Data storage	
CD	CRC-CCITT
FD 3,5”	CRC-CCITT
DLT	CRC-16
MO	CRC-32

CRC code is one of the several redundancy levels. In some transmission protocols the additional redundancy level is based on time. Data transmission errors are relatively easy to fix once an error is detected. For example if an IP packet that has been sent in a particular moment of time is lost or damaged then another copy of the packet is sent in the next moment of time.

On the other hand, the storage systems redundancy is generally based on the location. If a particular hard disk sector is damaged then another disk sector in a different location is used. Additionally, to provide data integrity over the long term, error correcting codes are required.

2.1.5. Compression and cryptography

Both communication and storage systems widely utilize data compression techniques. In each case standard, well-known algorithms are used. For example, video transmission (Digital TV in DOCSIS system) and video storage (DVD) systems use the same MPEG 2 algorithm.

The same applies to cryptography. For example, data stored on magnetic disks under MS Windows 2000 may be ciphered with a use of hybrid cryptography system based on such algorithms as 3DES and RSA. Data transmitted over SSL (Secure Socket Layer) connection are ciphered with a use of hybrid cryptography system, which may be based on the same algorithms.

2.1.6. Data organization

Typically data are stored and transmitted in blocks (frames, cells, packets, sectors). Each block is built from several fields: data, address, control, and error detection/correction. The transmitted and stored blocks of data have equivalent organization. Typically the block consists of a header, user data and control bytes. For example, Ethernet frame and hard disk sector have similar constitutive fields (table 6).

Table 6. Examples of transmission and storage block structures

Ethernet frame structure	**Hard disk sector structure**
Synchronization bits	Synchronization bits
Target physical address	Sector address (head number, cylinder number, sector number)
Source physical address	
Length of frame	
Data	Data
CRC	ECC

Transmitted or stored data blocks may have constant or variable length. In a great number of magnetic and optical storage systems data blocks have constant length. On the other hand in many communication protocols (e.g. Ethernet, IP) transmitted data blocks have variable length (table 7).

Table 7. Data blocks in transmission and storage systems

Data block	**Data block length (without a header)**
Data transmission	
ATM cell	constant (48 bits)
HDLC frame	Variable
Ethernet frame	variable (46÷1500 bytes)
IP packet	variable (0÷65515 bytes)
Data storage	
CD sector	constant (2048 bytes)
HD sector	constant (512 bytes)
FD sector	constant (512 bytes)
DLT block	Variable
MO sector	Selectable

It is worth observing that although the IP packet allows the total length of a datagram to be up to 65535 bytes such long datagrams are impractical for most networks. It is assumed that all hosts must be prepared to accept datagrams of total length up to 576 bytes. It is recommended that hosts only send datagrams larger than 576 bytes if they have assurance that the destination is prepared to accept the larger datagram. The length 576 bytes is selected to allow a reasonable sized data block to be transmitted in addition to the required header information. Since the maximal Internet header is 60 bytes there are at least 516 bytes for user data (Hall, 2000). This value is a little bit greater than magnetic disk sector length.

2.2. Replication of the system functions

There are many computer systems, in which the particular functions of the storage and transmission systems are duplicated. The question is, if this duplication of the functions is always necessary? For example, let's look at some phases of a file transmission procedure. Disk files are often transmitted in such network services as back-up, e-mail, FTP, WWW. In addition, it must be observed that storage dedicated networks, such as Storage Area Network (Toigo, 1999) emerge in large information system. Data grids are being built around the world as the next generation data handling systems. The grids are based on storage systems located in remote sites and connected by networks (Moore, 2004).

During the process, on the sender side (figure 2), some sectors of disk data are read, data are decoded, their checksums are examined and then a file is assembled. Next, the communication software divides the file into transmission packets and assigns new checksums for packets. Data are coded, modulated and transmitted. It must be observed that usually transmission packets and disk sectors have different sizes but it is possible to encapsulate a sector in a single packet. On the receiver side (figure 3) data is demodulated, decoded and the checksums are examined. The user data from the consecutive packets are assembled into a file and forwarded to the storage system. During write process a file is again divided into proper blocks (sectors). Every sector is given an appropriate checksum. Data are coded and stored on a disk.

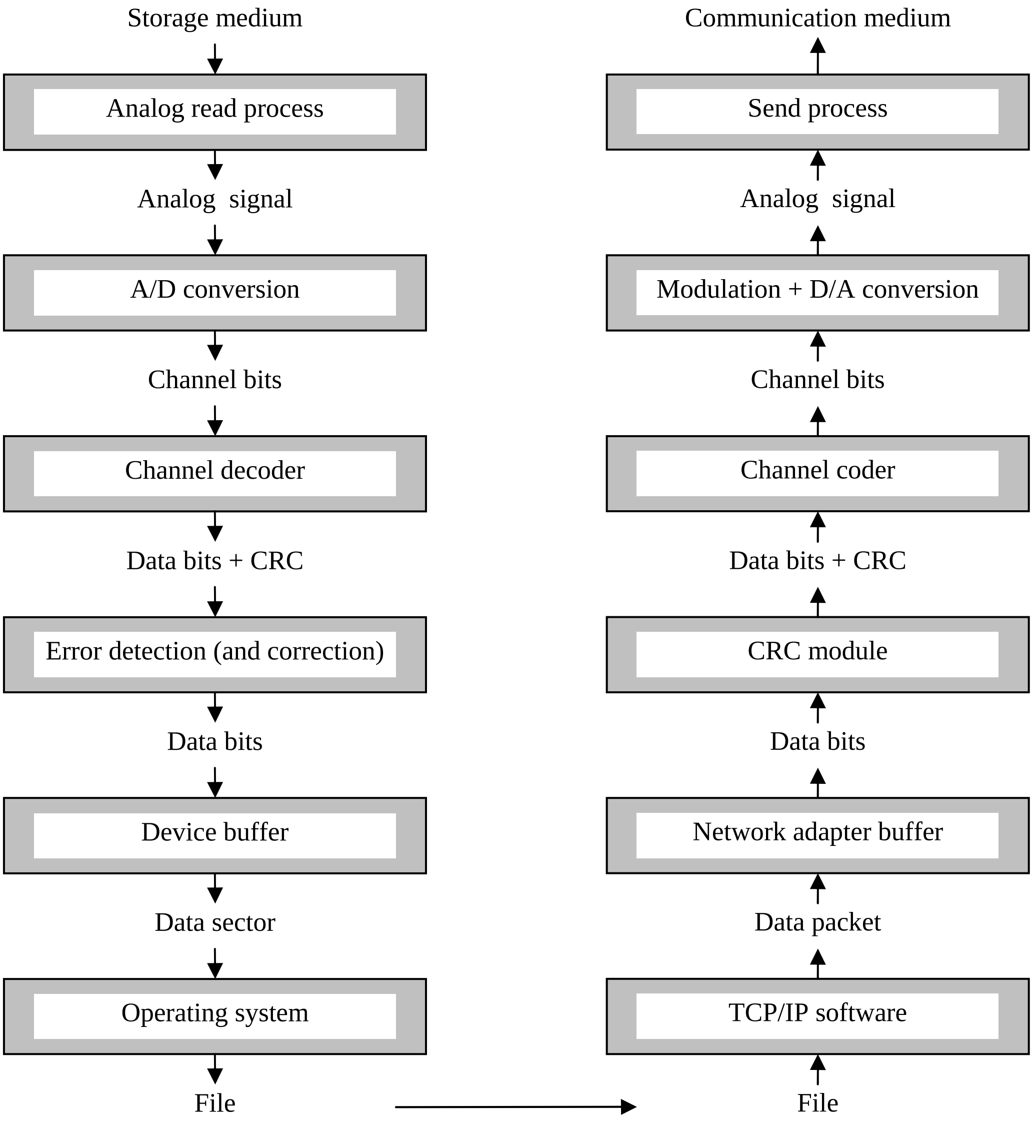

Figure 2. Send file procedure phases

During this multistage process some stages are performed with similar or identical methods and algorithms. The particular stages of the process (e.g. data channel coding and decoding, analog/digital conversion, error detection and correction, CRC computing) are replicated (Figures 2 and 3).

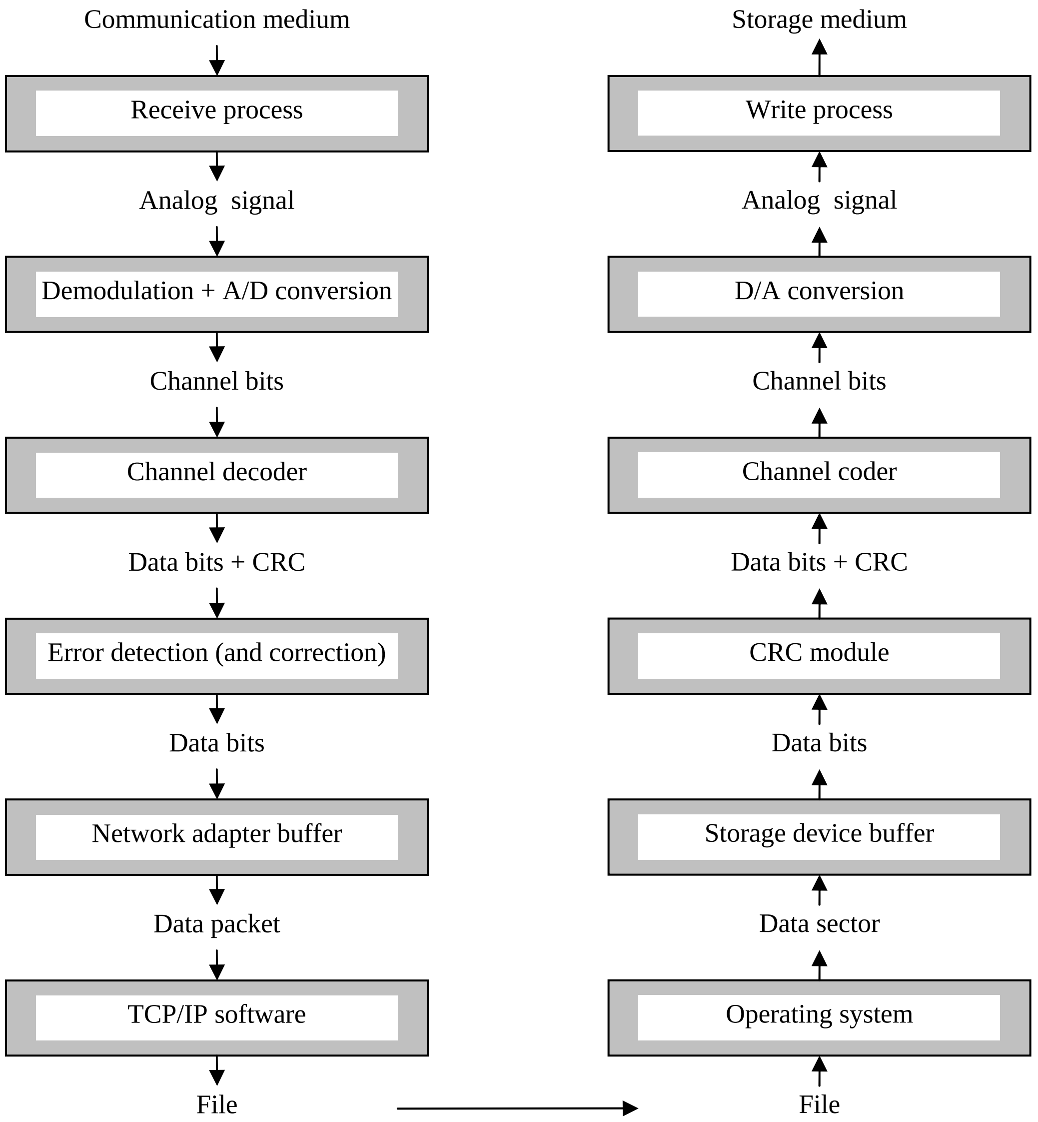

Figure 3. Receive file procedure phases

3. CONVERGENCE EXPLOITATION CONCEPT

3.1. General remarks

In the paper new shorter schemes are proposed. The schemes of the process with uniform sizes of packets and sectors and a common error detection/correction solution are evaluated. The proposed scenarios remove one or more replicated stages of the processes depicted on figures 2 and 3. This reduction of the number of process stages will result in shorter access time and less system overload.

In the traditional system communication between the subsystems is done using a bus. The bus is cheap and versatile. But, the disadvantage of a bus is that it creates a communication bottleneck. Recently, buses are increasingly being replaced by computer networks (Hennessy, 2002). Dedicated protocols, such as iSCSI, FCIP, iFCP are created to provide access to remote storage resources by TCP/IP networks (Black, 2001). The concept proposed here is a step in this direction.

Some different scenarios are presented here. The difference between them is the number of process stages that are removed from the procedures given in the previous section. In the first scenario we remove one stage in the send process and one in the receive process. In the last scenario we remove almost all stages of the process. Although, every possible scenario is presented here, not all of them may be used in practice. In a particular case one may choose the optimum scenario for a given storage/transmission system. Furthermore, it should be possible to implement different scenarios on the sender and the receiver end of the communication channel.

To implement the scenarios we must change the functionality of storage and network devices. For example, hard disks (on the sender and the receiver side) should have their own network connections, network interfaces with physical and IP addresses. Some research centers are now working on IP disk controller architecture (Wang, 2003). We may observe, that the IPv6 address space is adequate for this purpose (Hinden, 2003). The so-called "Internet-on-a-Chip" technology makes possible to combine core components of TCP/IP communication protocols and handheld computers onto a single chip without compromising performance or density (Lai, 2002). Additionally, in the connection oriented transmission protocols (such as TCP) a connection set up is required. We assume that this is done either by the host operating system or by the communication software integrated with the storage device (such as HD, RAID, tape library).

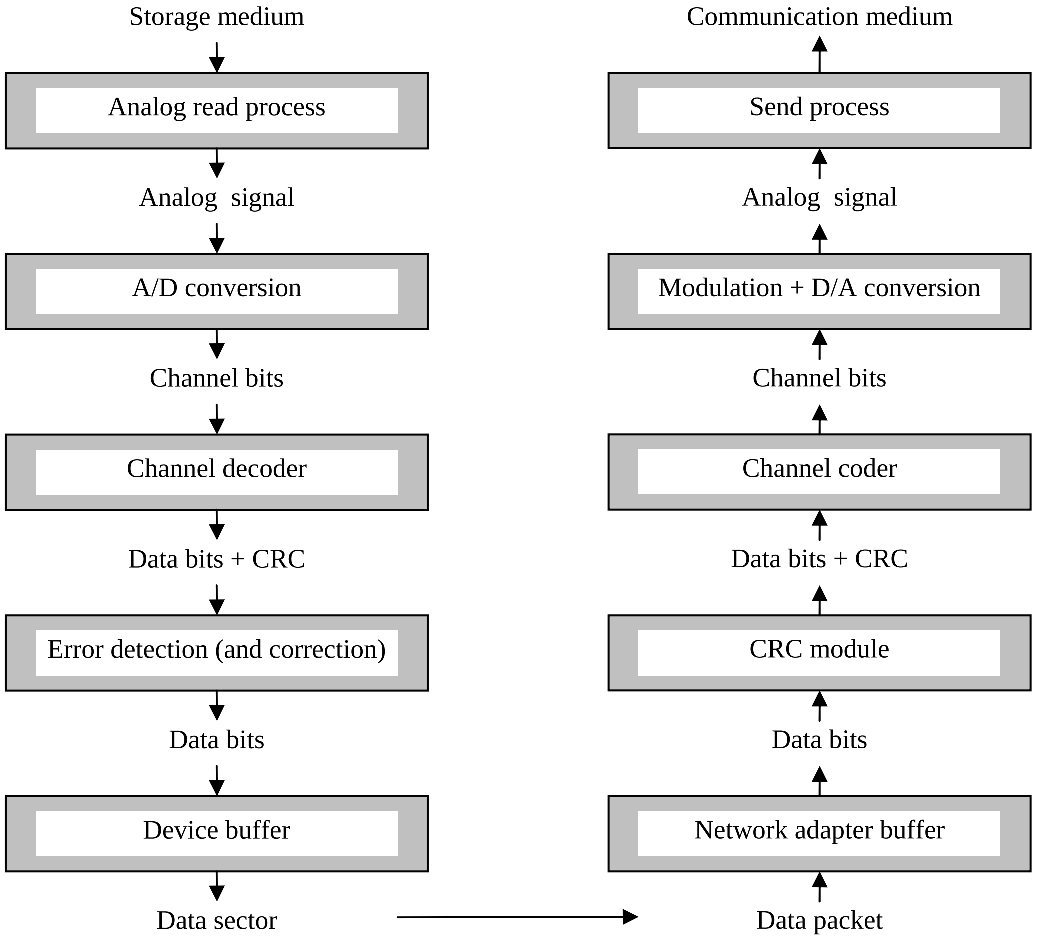

Figure 4. Send file procedure phases in scenario 1

3.2. Scenario 1

In the first scenario operating system on the sender side issues file transmit command. The command is performed by the disk drive. The appropriate data sectors are read and converted to data packets by drive. It is assumed that both the data sector and data packet have the same data field size. The given packet is transferred directly from storage device to network adapter, which sends it via the communication channel (figure 4). In this scenario, transmitted file is not assembled on the sender side.

On the receiver side of the communication channel, received data packet from network adapter buffer is transferred to the storage device. The storage device converts it into a sector and stores (figure 5). Finally storage device informs receiver operating system that a new file has been created.

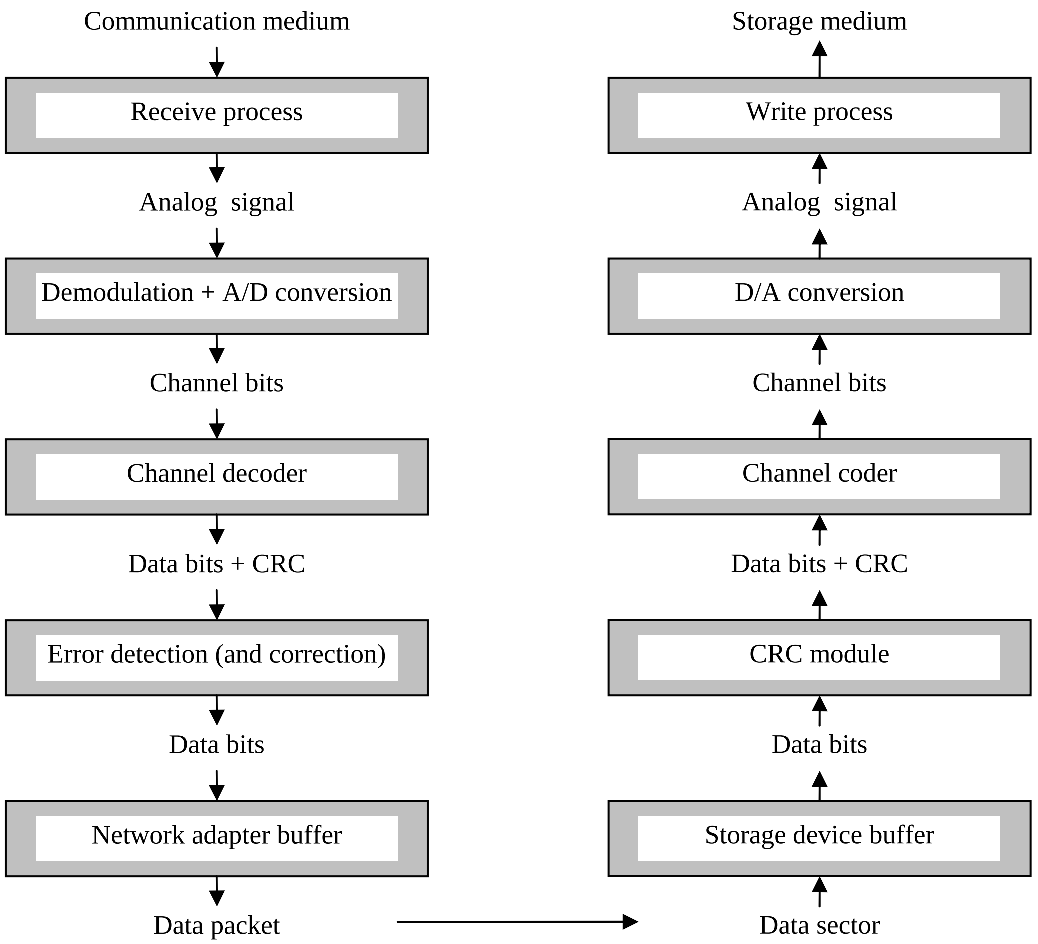

Figure 5. Receive file procedure phases in scenario 1

3.3. Scenario 2

In the second scenario the sequence of data bits read from the storage medium are not assembled to form a sector but transferred as a stream to the sender communication device, which transmits them via the communication channel (figure 6). On the receiver side the same stream of bits is transferred from the communication device to the storage device and stored in the storage medium (figure 7).

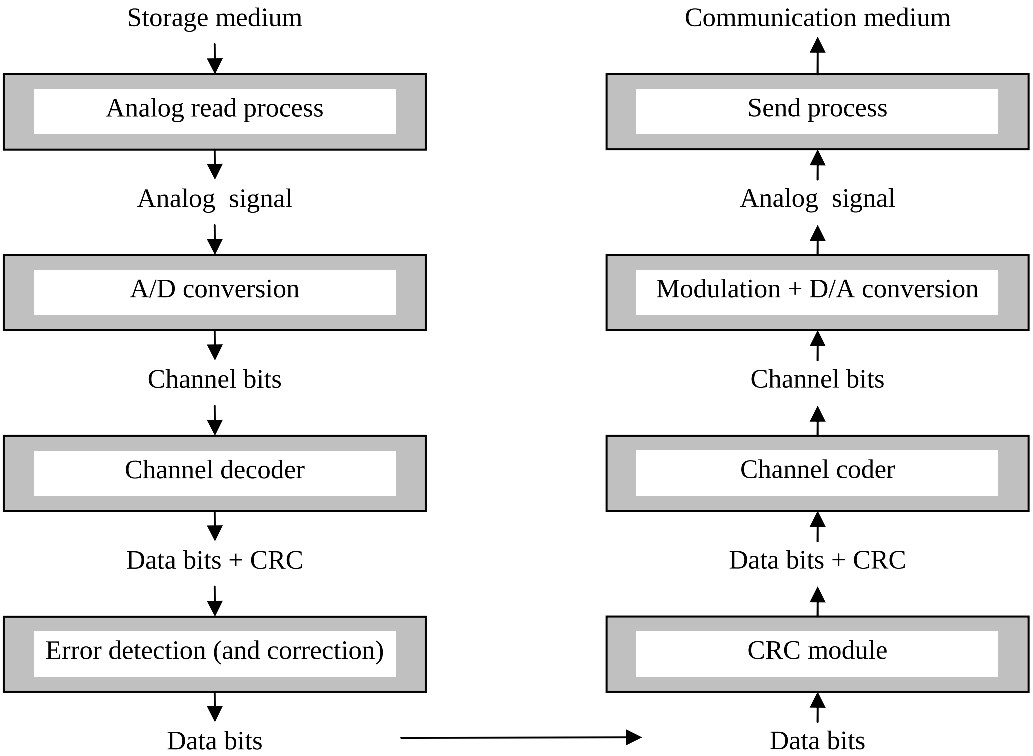

Figure 6. Send file procedure phases in scenario 2

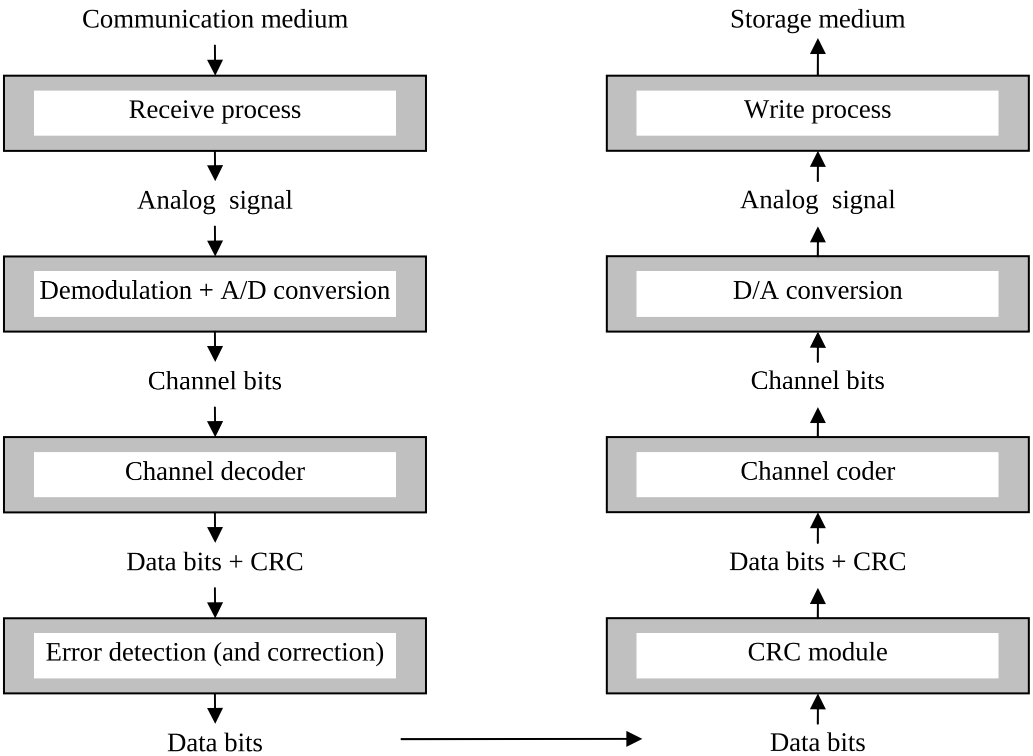

Figure 7. Receive file procedure phases in scenario 2

3.4. Scenario 3

The third scenario is similar to the second one but without CRC handling. Original data bits read from medium together with its CRC are transferred to the network device and transmitted to the receiver. The sender network adapter doesn't calculate new CRC bits (figure 8).

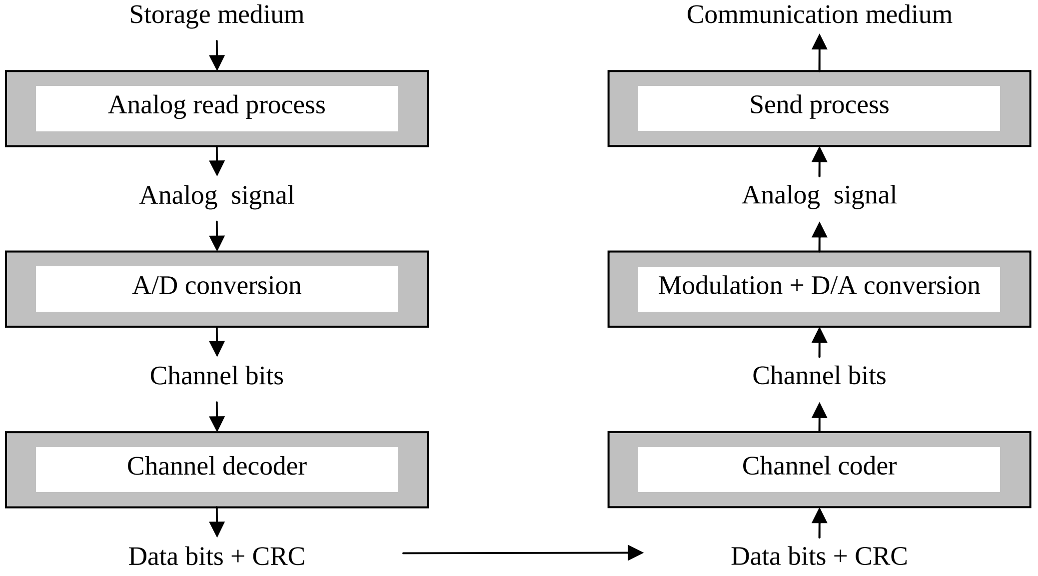

Figure 8. Send file procedure phases in scenario 3

The receiver network device transfers data bits together with its orginal CRC to the storage device. Data bits together with transmitted CRC are stored in the storage medium (fig. 9).

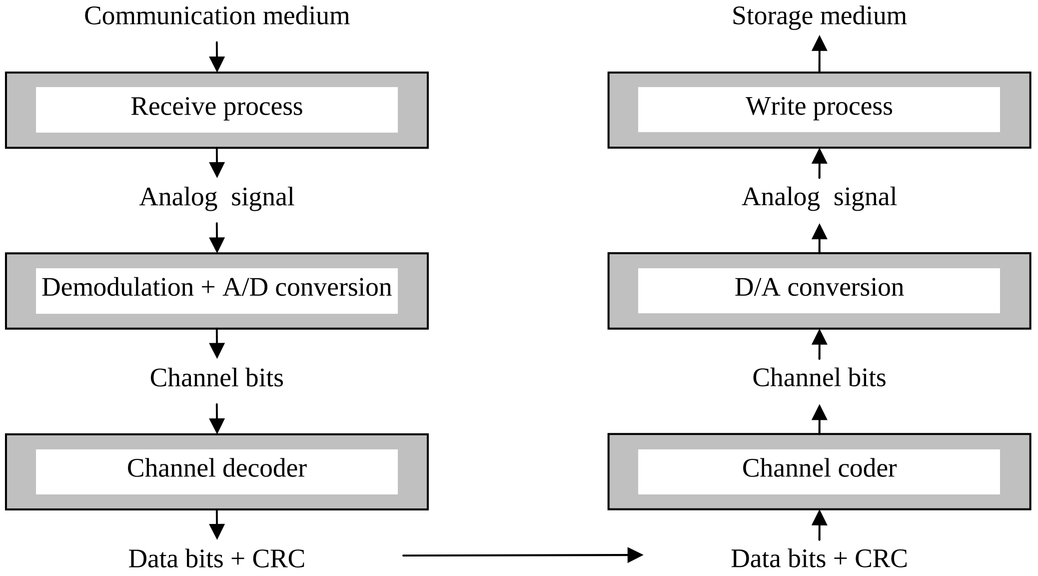

Figure 9. Receive file procedure phases in scenario 3

3.5. Scenario 4

In this scenario channel bits read from the storage medium are transferred to the network device and sent without any additional processing (fig. 10).

Similarly, during the receive process the channel bits from the communication medium are transferred to the storage device and stored without any additional processing (fig. 11).

The evident weakness of this scenario is the lack of any error correction and detection mechanisms for transmission and storage processes. It must be assumed that both storage and communication media are fully reliable.

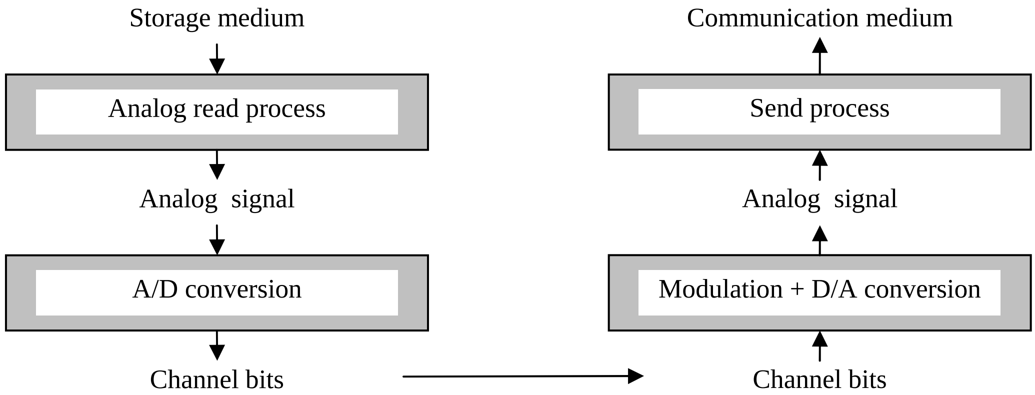

Figure 10. Send file procedure phases in scenario 4

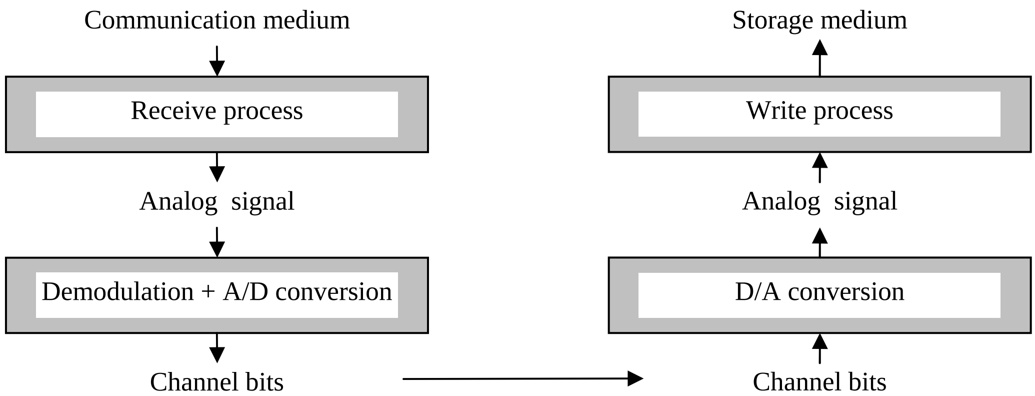

Figure 11. Receive file procedure phases in scenario 4

3.6. Scenario 5

The last scenario is based on an analog signal transmission. In fact, in each transmission system data are represented by the states of analog communication channel. Analog signals read from the storage medium are amplified and transmitted (figure 12). Additional modulation before the transmission is possible.

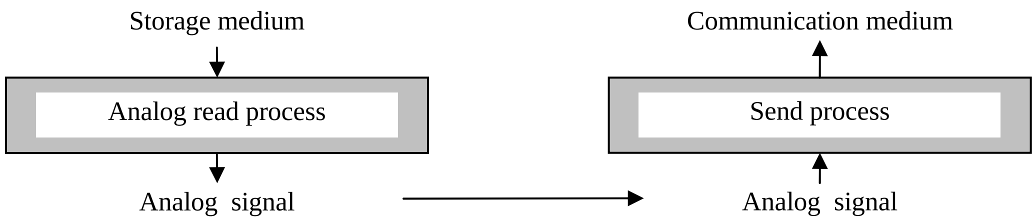

Figure 12. Send file procedure phases in scenario 5

At the receiver side the analog signal is transferred directly (or after demodulation) to the write head (figure 13).

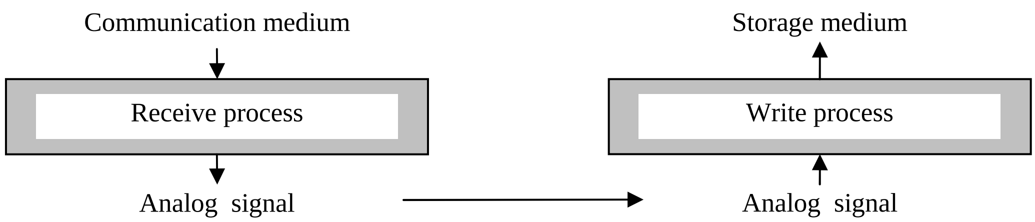

Figure 13. Receive file procedure phases in scenario 5

The evident weakness of this scenario is the lack of any error correction and detection mechanisms for transmission and storage processes. It must be assumed that both storage and communication media are fully reliable.

4. SUMMARY

In the paper we have shown many examples of convergence between storage and transmission processes. The similarities are evident in such aspects of the two systems as: architecture, data organization, physical phenomena, coding and redundancy. Furthermore, in some jobs performed by computer system (such as a file transmission) both processes are employed. It has been shown, that in such task the same elementary steps are executed many times. Removal of the replicated steps together with some changes in the system architecture and tasks performed by specific modules (especially storage and communication devices) may lead to the system performance optimization. Although, every possible scenario is presented here, not all of them may be used in practice. In a particular case one may choose the optimum scenario for a given storage/transmission system. Furthermore, it should be possible to implement different scenarios on the sender and the receiver end of the communication channel.

REFERENCES

Black D. (2001), IETF's IP Storage (IPS), Ninth NASA Goddard Conference on Mass Storage Systems and Technologies, Eighteenth IEEE Symposium on Mass Storage Systems.

Burr A. (2001), Modulation and Coding for Wireless Communication, Pearson Education Limited, Harlow.

Cole J.L. (2000), Project 1244: IEEE Storage System Standards, Eighth NASA Goddard Conference on Mass Storage Systems and Technologies, Seventeenth IEEE Symposium on Mass Storage Systems, 429-441.

Gniewek J.J., Vogel S.M. (1995), Influence of Technology on Magnetic Tape Storage Device Characteristics, Papers from the Fourth NASA Goddard Conference on Mass Storage Systems and Technologies.

Hall E.A. (2000), Internet Core Protocols, O'Reilly, Sebastopol.

Hennessy J.L., Patterson D.A., Goldberg D. (2002), Computer Architecture: A Quantitative Approach, Morgan Kaufmann, San Mateo.

Hinden R., Deering S. (2003), Internet Protocol Version 6 (IPv6) Addressing Architecture, The Internet Society, RFC 3513.

Lai S., (2002), Future Trends in Non-Volatile Memory Technology, http://www.ece.ualberta.ca/~salamon/OUM/StefanLaiIDF0202.pdf

Moore R.W., Jagatheesan A., Rajasekar A., Wan M., Schroeder W. (2004), Data Grid Management Systems, 21st IEEE/12th NASA Goddard Conference on Mass Storage Systems and Technologies.

Toigo J.W. (1999), The Holy Grail of Data Storage Management, Prentice Hall PTR, Upper Saddle River.

Wang P., Gilligan R., Green H., Raubitschek J. (2003), IP SAN – From iSCSI to IP-addressable Ethernet Disks, Twentieth IEEE/Eleventh NASA Goddard Conference on Mass Storage Systems and Technologies.