

Section 1. SEC (Security of computer systems and networks. Fault-tolerant distributed computing)

UDC 004.8

Larysa Doroshenko,
Oleksandr Markovskiy, Andrii Honchar

ORGANIZATION OF RESERVATION AND RECONSTRUCTION OF DATA

Лариса Дорошенко,
Олександр Марковський, Андрій Гончар

ОРГАНІЗАЦІЯ РЕЗЕРВУВАННЯ ТА ВІДНОВЛЕННЯ ДАНИХ

In the article the method of additional packets formation during the information transfer on the Internet is offered. Also suggested is the way to use additional packages to recover lost or corrupted packets. Linear redundant codes are used to recover. An example of using the proposed method is given.

Key words: erasure codes, information packets, data packet reconstruction, linear codes.

Fig. 0. Tabl.: 1. Bibl.: 4.

У статті запропоновано метод формування додаткових пакетів при передачі інформації в Інтернет. Також запропоновано спосіб використання додаткових пакетів для відновлення втрачених або пошкоджених при передачі основних пакетів. Для відновлення використовуються лінійні надлишкові коди. Наведено приклад застосування запропонованого методу.

Ключові слова: відновлюючі коди, інформаційні пакети, відновлення пакетів даних, лінійні коди.

Рис. 0. Табл.: 1. Бібл.: 4.

Target setting. In the modern world the role of the Internet is dominant among the means of data exchange. Since there is a probability of loss, damage or delayed arrival of data packets over a global network, there is a need to create new methods and tools to solve these problems [1].

Actual scientific researches and issues analysis. Today the following technologies are used to reconstruct lost data during transmission: partial or complete duplication of data and their re-transmission, usage of the Reed-Solomon correction codes (the disadvantage is the reduction of the data to the solution of the nonlinear equations system, which is time-consuming), the usage of CRC-codes [2, 3]. The last two technologies perform error correction at the symbol level, which doesn't allow the recovery of lost packets during transmission [4].

Not investigated parts of the general subject. At present great attention is paid to erasure codes that work at the packet level. The main principle is the transfer of basic packages along with additional ones, which partially contain the information that is in the main packets. With the help of erasure codes. It is possible to reconstruct lost or damaged packets with a certain probability. However, the erasure codes that are used do not provide high data reconstruction efficiency.

Research objective. The purpose of the article is to investigate the dependence of the lost data packets reconstruction on the number of main and additional data packets. Analyze the extermination results and form on the basis of the analysis the method of data packets reconstruction with the highest probability.

Principal statements. The information packet is divided into n blocks, each of which is transmitted separately. The block, with the number j , where $j \in \{1, \dots, n\}$, consists of a sequence of words: $a_{j1}, a_{j2}, \dots, a_{jm}$. In the process of transmission in case of loss or distortion of block data with numbers q and r , where $q, r \in \{1, \dots, n\}$ it is necessary to restore the words $a_{q1}, a_{q2}, \dots, a_{qm}$ and $a_{r1}, a_{r2}, \dots, a_{rm}$ from information, stored in $n-2$ main and k backup blocks. To ensure the possibility of simple data recovery when one carrier is lost, it is proposed in the first backup block to store the sum of the module 2 of all the same words of the main blocks:

$$\forall j = 1, \dots, n : s_{1j} = \bigoplus_{i=1}^n a_{ji} \quad (1)$$

Considering the task of reconstruction the q and r main units of the words $a_{q1}, a_{q2}, \dots, a_{qm}$ and $a_{r1}, a_{r2}, \dots, a_{rm}$, their values can be found as a result of solving systems of two 30oolean equations of the form:

$$\begin{cases} a_{qj} + a_{rj} = z_{1j} \\ a_{qj} = z_{2j} \end{cases} \quad \text{or} \quad \begin{cases} a_{qj} + a_{rj} = z_{1j} \\ a_{rj} = z_{3j} \end{cases}, \quad \forall j \in \{1, \dots, n\}, \quad (2)$$

where the '+' symbol is denoted by a binary add operation in module 2. The first equation in systems (2) is transformed from (1), that is, the data stored in the first data backup block are used to obtain it. Obtaining the first equation of systems (2) is much more complicated, due to the fact that the a priori values of q and r are unknown. It is obvious that the second equation of system (2) can be obtained from the system of linear 30oolean equations, which for any values of q and r contains an equation in which only a_{q1} or only a_{r1} is included in the term. If n is a degree 2, then the example of such a system may look like:

$$\begin{cases} a_{1j} + a_{2j} + \dots + a_{m/2,j} = y_j \\ a_{1j} + a_{2j} + \dots + a_{m/4,j} + a_{m/2+1,j} + \dots + a_{3m/4,j} = y_2, \forall j \in \{1, \dots, n\} \\ \dots \\ a_{1j} + a_{3j} + a_{5j} + \dots + a_{m-1,j} = y_{\log_2 n, j} \end{cases} \quad (3)$$

In the general case of arbitrary value n , the number of equations of the system of the form (3) is equal to the nearest whole, equal to or greater than $\log_2 n$: $\lceil \log_2 n \rceil$. In order to reconstruct the information packet when any two blocks (main or backup) are lost, it is necessary to save in the same backup unit the sum of the two basic words of the same name (1), the amounts determined by the system (3) in the backup blocks and in one more backup unit duplicate the last information block.

Thus, if each word from a pair of blocks that is lost or damaged during data transfer is reconstructed independently, then the number of backup units (packets) of the media is:

$$kp = 2 + \lceil \log_2 n \rceil.$$

General structure. The reconstruction of the lost “pack” of information packets is proposed to be organized as follows. Assuming that the lost packets belong to one, for the definiteness of the i -th block, and have sequence numbers from j to $(j + 1)$ in the block, where $j \in \{1, 2, \dots, h \cdot (l-1)\}$, then it is quite obvious that the number of lost packets in each of the columns of the matrix M does not exceed one. Accordingly, the recovery of $p_{ij}, p_{ij+1}, \dots, p_{i(j+l)}$ packets is performed by calculating the amount by module 2 of the received packs columns and the corresponding backup code. Formally, the recovery procedure is described by the formula:

$$p_{ij} = C_{i,j \bmod l} \oplus \bigoplus_{q=0, q \neq j \bmod l}^{h-1} p_{i,q \cdot l + j \bmod l}. \quad (4)$$

The proposed method is actually specialized for the dominant type of information packets loss when broadcasting video information transmission in peer-to-peer networks. The analysis showed that the main reason for such losses is the exclusion of the network node, through which the main stream of video information is transmitted. Accordingly, the dominant type of the continuity violation of delivery to subscribers of video information is the loss of a group of consecutive packages that form a “pack.”

Testing. The analysis of the data presented in table 1 indicates that the proposed approach provides the reconstruction possibility of the loss of blocks of large multiplicity with a probability close to one for data blocks of real length. In this case, the number of operations required to determine the positions of blocks, distorted during the transfer does not exceed $\log_2 m$. This means that the error correction of a large multiplicity can almost be done at the rate of data transmission. Situations, in which the correction is impossible, can be easily detected by analyzing the differences between the components of the control code. In this case, the errors can be corrected by re-transmission, but the probability of this, as seen from table 1, is sufficiently small considering the actual probability of damage of the blocks during the transfer.

Table 1

Experimental probability of data recovery

<i>The number of main data blocks</i>	<i>Number of damaged blocks</i>	<i>The number of backup data blocks</i>			
		<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>
6	2	0.28	0.67	0.87	1
	3	0	0.34	0.52	0.86
	4	0	0	0.37	0.47
8	2	0.32	0.72	0.90	1
	3	0	0.42	0.63	0.82
	4	0	0	0.47	0.59
10	2	0.39	0.76	0.92	1
	3	0	0.44	0.64	0.79
	4	0	0	0.49	0.62

Thus, as a result of the conducted research, the theoretical substantiation and development of the method of data packet reconstruction during their transmission in global computer networks on the basis of linear redundant codes, which are executed in the form of reserving packets, is carried out. Development of a method of formation by reserving packages and methods for reconstruction lost packets. The simulation carried out for the experimental study of the use of linear codes for the reconstruction of data packets in order to select the optimal characteristics for different types of channels and data transmission errors confirmed, in general, the results obtained theoretically.

Conclusion. As a result of the conducted research, the method of reservation and reconstruction of data packets for transmission in the global networks was proposed.

To achieve the goal, a method is proposed for the formation of additional packets and reconstruction lost, damaged or delayed critical time of the main data packets using them.

References

1. Tanenbaum A.S. Computer networks. Prentice Hall PTR.- 2016.-960 p.
2. Стіренко С.Г., Габінет А.В., Костенко Ю.В. Забезпечення безперервного відтворення потокового відео в однорангових мережах з використанням erasures кодів. // Вісник НТУУ «КПІ». Інформатика, управління та обчислювальна техніка: збірник наукових праць. – К.: «Век+», 2015. – № 62. – С. 105–110.
3. Leong D. Erasure coding for real-time streaming / D. Leong, T. Ho // Proceedng IEEE Int. Symposium Information Theory – ISIT-2012 . – 200.
4. Brinkmeier M., Fischer M., Grau S., Schaefer G., Strufe T. Methods for Improving Resilience in Communication Networks and P2P Overlays. PIK // Praxis der Informationsverarbeitung und Kommunikation 32, 2009.

Autors

Doroshenko Larysa – student, Department of Computer Engineering, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”.

E-mail: larysad785@gmail.com

Дорошенко Лариса Юріївна – студентка, кафедра обчислювальної техніки, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського».

Markovskiy Oleksandr – associate professor, Department of Computer Engineering, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”.

E-mail: markovskyy@i.ua

Марковський Олександр Петрович – доцент, кафедра обчислювальної техніки, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського».

Honchar Andrii – student, Department of Mobile and video information technology, State University of Telecommunications.

E-Mail: andrew.gonchar98@gmail.com

Гончар Андрій Анатолійович – студент, кафедра Мобільних та відеоінформаційних технологій, Державний Університет Телекомунікацій.

РОЗШИРЕНА АНОТАЦІЯ

Л. Ю. Дорошенко,
О. П. Марковський, А. А. Гончар

ОРГАНІЗАЦІЯ РЕЗЕРВУВАННЯ ТА ВІДНОВЛЕННЯ ДАНИХ

Актуальність теми дослідження. У сучасному світі спостерігається зростання домінуючої ролі Інтернету в процесі передачі інформації, що вимагає розробки нових засобів та методів для забезпечення надійності пересилки пакетів даних.

Постановка проблеми. При передачі інформації через глобальні мережі можлива втрата, пошкодження або запізнення отримання (по настанню критичного моменту часу) пакетів даних. Використання запропонованого методу підвищує вірогідність відновлення втраченої інформації під час передачі.

Аналіз останніх досліджень і публікацій. На сьогоднішній день використовується декілька технологій для відновлення втраченої інформації. Найпростішою є часткове або повне дублювання втрачених пакетів даних і повторна їх передача, використання коригуючих кодів Ріда-Соломона та CRC-кодів для виправлення помилок на рівні символів, використання відновлюючих кодів.

Виділення недосліджених частин загальної проблеми. Стаття присвячена формуванню методу побудови додаткових пакетів даних при передачі інформації та аналізу залежності надійності передачі даних від способу формування додаткових пакетів. Найбільшу увагу приділено розробці методу формування додаткових пакетів для найбільшої вірогідності відновлення втрачених пакетів.

Постановка завдання. Завданням є розробити метод формування додаткових пакетів, який би дав можливість найбільш ефективно відновлювати втрачені та пошкоджені основні пакети даних.

Викладення основного матеріалу. Проведено дослідження та аналіз ймовірностей відновлення втрачених пакетів в залежності від кількості посилення додаткових пакетів та кількості втрачених пакетів. Визначено спосіб формування додаткових пакетів для найбільшої ймовірності відновлення втрачених пакетів.

Висновки. Досліджено залежності відновлення пакетів даних від способу резервування даних та кількості додаткових пакетів. Проаналізовано експериментальні дані та виділено найкращі результати, на основі яких сформовано запропонований метод.

Ключові слова: відновлюючі коди, інформаційні пакети, відновлення пакетів даних, лінійні коди.