PACS 87.19.lt, 85.60.Pg , 89.70.Kn

Increase of data protection level for visual information in control systems

A.V. Bushma
O.S. Popov Odessa National Academy of Telecommunications
1, Kuznechnaya str., 65029, Odessa, Ukraine

Abstract. Formalized description of data transmission between technical means and an operator from the viewpoint of information security is presented. The most widely used symbolic and bar graph data representation forms are analyzed and compared. An analytical estimation of redundancy inherent to the considered information models is offered.

Keywords: ergatic system, information security, digital data, redundancy, information model, symbolic form, bar graph form.

Manuscript received 24.02.10; accepted for publication 08.07.10; published online 30.09.10.

1. Introduction

Intensive development of information technologies in the whole area of modern society life, high level of responsibility in business processes and rigid requirements to protection of data used in them, wide implementation of digital technologies when controlling complex objects and technological equipment make man to be a key object in ergatic systems. Simultaneously, essentially increased are requirements to information security of a channel for visual data transfer. System approach allows adequate estimation of threats to information and provides its efficient protection [1 - 3].

In this situation, one should take into account that technology of visual data transfer generates practically all kinds of threats, namely: data corruption, disturbance in logic structure, contents, confidentiality and privacy [3, 4]. As a result, to provide a high level of information security they need a complex system for data protection. Our analysis shows that the highest level of vulnerability for a visual channel in ergatic systems is related with data corruption, disturbances in logic structure and contents of messages transmitted to a man.

Here, negative action on information is related with its corruption and unauthorized modification. One of the most efficient ways to protect digital data from a random or intended corruption made practically by any threat source is redundant coding [1, 4 - 7]. In these cases, using the codes with a high level of redundancy, when transferring data to an operator via the optical channel, is reasonable due to features of functioning human vision system [8, 9]. Therefore, practical interest consists of the analysis of information parameters inherent to signals that code messages in the channel for connection with the operator [10, 11]. But quantitative estimations of these characteristics for visual representation of digital messages are absent up to now.

This work is aimed at investigation of data protection in the visual channel of an ergatic system as well as ways to enhance it due to increase in the level of redundancy for codes used for formation of a digital optical signal.

2. Conception

The channel for visual data transfer functions is based on the system of rules that determine a correspondence between formed communications and a number of visual images, it means that there realized is a specific signal coding with using an information model (IM). Realization of IM is performed by an electrooptical converter (EOC) within the limits of the part of space being an information area (IA) in the facility for imaging (display).

The most widely used for visual representation of digital data in radio-electronic facilities and information-controlling systems are digital (symbolic) and discrete-analogous (scale) IM [12, 13].

Possessing a definite information redundancy, these IMs allow protection of digital data in the channel “technical means – man” from a random or intended disturbance generated practically by any threat source:

© 2010, V. Lashkaryov Institute of Semiconductor Physics, National Academy of Sciences of Ukraine

human factor, software-hardware means or external medium [8 - 10]. Let us estimate an information redundancy inherent to visual codes of widely used IMs, which will enable to juxtapose the levels of data protection in the optical transmission channel in various ergatic systems.

Formalized representation used for the process of data flow transfer to the operator implies presence of a structure where an information source, transmitter, transmission channel, receiver and addressee are connected in series, in accordance with the algorithm of signal processing [10]. In this approach, technical means of the system are the information source. In a specific case, as a rule, there generated is a finite set $\mathbf{I}$ of messages

$$\mathbf{I} = \{\, I_1,\ I_2,\ \ldots,\ I_\nu,\ \ldots,\ I_{l-1},\ I_l \,\}, \tag{1}$$

where I_ν is the ν-th message, and $\nu = \overline{1,\ l}$; l is the number of different messages circulating in the system.

It is assumed here that the message is formed in a random manner, and the addressee knows only that it belongs to the set $\mathbf{I}$. Finiteness of the latter defines a discrete character of information transfer in the system [10]. A syntactic analysis of messages implies abstracting from their semantic features, which allows the most objective estimation of properties inherent to various IMs. Then, information can be determined as a measure of decreasing uncertainty of knowledge about some subject. As the analyzed system provides a discrete data transfer, to estimate the amount of information one can use the probability measure offered by Shannon, the so-called “entropy’ [7, 9, 11]. In the considered case, one can write

$$H(I_\nu) = -\sum_{\nu=1}^{l} P(I_\nu) \log P(I_\nu), \tag{2}$$

where $H(I_\nu)$ is the entropy of the ν-th message I_ν; $P(I_\nu)$ – probability of I_ν appearance, and $\sum_{\nu=1}^{l} P(I_\nu) = 1$.

If after receiving the message I_ν the uncertainty is fully eliminated, then the amount of transferred information is equal to its entropy.

The messages formed by the information source in accord with the definition (1) are transferred to the transmitter. This function is provided by EOC converting this coming information to an optical (visual) form. The rules for this conversion are established by IM. As a result, synthesized inside IA is the corresponding symbol from the finite set

$$\mathbf{S}_{\mathrm{IM}} = \{\, S_{1\mathrm{IM}},\ S_{2\mathrm{IM}},\ \ldots,\ S_{\nu\mathrm{IM}},\ \ldots,\ S_{(l-1)\mathrm{IM}},\ S_{l\,\mathrm{IM}} \,\}, \tag{3}$$

where $S_{\nu\,\mathrm{IM}}$ is the ν-th symbol of IM, and $\nu = \overline{1,\ l}$; l is the length of the IM alphabet.

Representation of digital data implies that every element of the $\mathbf{I}$ set is corresponded by one element from $\mathbf{S}_{\mathrm{IM}}$, and vice versa, every element from $\mathbf{S}_{\mathrm{IM}}$ is corresponded by one element from $\mathbf{I}$. Then, between the sets (1) and (3) there exists one-to-one correspondence (bijection) and they are equivalent.

To realize IM from symbols belonging to the set $\mathbf{S}_{\mathrm{IM}}$, one can use IA formed by the finite set $\mathbf{A}$ of elements a_i

$$\mathbf{A} = \{\, a_1,\ a_2,\ \ldots,\ a_i,\ \ldots,\ a_{p-1},\ a_p \,\}, \tag{4}$$

where p is the total amount of IA elements.

When forming the message I_ν on IA from the elements a_i, synthesized is the corresponding symbol $S_{\nu\,\mathrm{IM}}$. These elements form the set $\mathbf{A}_{\nu\,\mathrm{IM}}$ that is a subset of the $\mathbf{A}$ set and on the level of technical means is created via excitation of IA elements on EOC by using the binary code Z_{IM}. This code can be described in two ways.

On the one hand, as a set of digits forming it

$$\mathbf{Z}_{\mathrm{IMd}} = \{\, z_1,\ z_2,\ \ldots,\ z_i,\ \ldots,\ z_{d-1},\ z_d \,\}, \tag{5}$$

where z_i is the i-th digit of the code, and $z_i = 0 \vee 1$, $i = \overline{1,\ d}$.

On the other hand, the excitation code of EOC can be considered as a set of words

$$\begin{aligned}\mathbf{Z}_{\mathrm{IMA}} = \{ &Z_{1\mathrm{IMA}},\ Z_{2\mathrm{IMA}},\ \ldots,\ Z_{\nu\,\mathrm{IMA}},\ \ldots \\ &\ldots,\ Z_{(q-1)\,\mathrm{IMA}},\ Z_{q\,\mathrm{IIMA}} \},\end{aligned} \tag{6}$$

where $Z_{\nu\,\mathrm{IMA}}$ is the ν-th word of the code Z_{IM}, and $\nu = \overline{1,\ q}$.

Realization of the information transmitter with technical means provides independent control over all IA elements as well as equivalency and bijection of the sets $\mathbf{A}$ and $\mathbf{Z}_{\mathrm{IMd}}$. In this case, from the information viewpoint, static and dynamical formation of visual images for $S_{\nu\,\mathrm{IM}}$ symbols does not imply any transfer of additional data to the operator, which defines invariance of the considered parameters relatively to the EOC excitation mode. Therefore, we shall analyze static synthesis of an image, which implies that the amount of digits in the code controlling IA Z_{IM} is equal to the number of its elements a_i (it means that $d = p$ in the definition (5)).

Technical means for controlling EOC form bijection between symbols and coding words in the sets described by (3) and (6) as well as define equivalency of the sets $\mathbf{S}_{\text{IM}}$ and $\mathbf{Z}_{\text{IMA}}$. It follows from the latter that the number of allowed p-digital words in the controlling code for IA Z_{IM} is equal to the number of symbols in the set $\mathbf{S}_{\text{IM}}$, thereof $q = l$ in the expression (6). If the code contains forbidden combinations of logic symbols, then they are blocked by technical means of the transmitter. Therefore, these words do not take part in data transfer and, from the formal viewpoint, do not influence on the amount of information transferred by these messages. But their presence influences on system parameters, which will be considered below.

3. Analytical model

The visual transmission channel provides message transfer to the receiver – visual analyzer of the operator. In a general case, also present in the channel are other signals and random processes that can introduce some differences between ν-th symbols at its input and output $S_{\nu\,\text{IM}}$ and $\tilde{S}_{\nu\,\text{IM}}$, respectively. To simplify our analysis, we assume that the data are not disturbed in the course of transfer, and new additional variants of symbols are not born. Then, the obtained information is an optical symbol belonging to a finite set

$$\tilde{\mathbf{S}}_{\text{IM}} = \left\{ \tilde{S}_{1\text{IM}}, \tilde{S}_{2\text{IM}}, \ldots, \tilde{S}_{\nu\text{IM}}, \ldots, \tilde{S}_{(l-1)\text{IM}}, \tilde{S}_{l\,\text{IM}} \right\}. \tag{7}$$

In assumption of an ideal character of the connection channel, one can state that the sets $\mathbf{S}_{\text{IM}}$ and $\tilde{\mathbf{S}}_{\text{IM}}$ are equivalent.

The visual analyzer recovers the message transferred from the information source by using the received signal that belongs to the set $\tilde{\mathbf{S}}_{\text{IM}}$ (7). When spatial, brightness and time characteristics of visual signals decline from the optimal values, the transferred symbols can be disturbed [8, 13]. Let us assume that the messages are recovered without errors. Formed in this case for the addressee – human brain – is the image of the initial ŕJ-th message I_ν in the form $R_{\nu\,\text{IM}}$ that belongs to the finite set

$$\mathbf{R}_{\text{IM}} = \left\{ R_{1\text{IM}}, R_{2\text{IM}}, \ldots, R_{\nu\text{IM}}, \ldots, R_{(l-1)\text{IM}}, R_{l\,\text{IM}} \right\}. \tag{8}$$

When the data are recovered in a correct manner, the sets $\tilde{\mathbf{S}}_{\text{IM}}$ and $\mathbf{R}_{\text{IM}}$ are equivalent.

As the adjacent sets in the considered series $\mathbf{I}$, $\mathbf{S}_{\text{IM}}$, $\tilde{\mathbf{S}}_{\text{IM}}$ and $\mathbf{R}_{\text{IM}}$ are equivalent in pairs, and the same is valid to the sets $\mathbf{S}_{\text{IM}}$ and $\mathbf{Z}_{\text{IMA}}$, it is easily to prove that any two of them are equivalent, too. Therefore, the absence of interferences and disturbances in the analyzed channel provides bijection between elements of all the sets that describe intermediate forms of message representation when transferring them from technical means to the operator. Consequently, this analysis of information parameters for signals used to code data in the connection channel with account of assumptions above can be performed in any its point, including the information transmitter. As an object for this study, we shall use the formed in this element message code Z_{IM} that is described with the set of words $\mathbf{Z}_{\text{IMA}}$ in accord with the expression (6).

As a rule, in visual alphabets there is some information redundancy. The reason for its appearance is synthesis of symbols from discrete elements of the IA indicator, the number of which exceeds the minimal necessary value for data coding free of redundancy and is defined by geometrical features of the formed visual images. Let us consider the most general case of equiprobable appearance for messages I_ν, when for any ν value the probability $P(I_{ŕJ}) = 1/l$.

Then, bearing the equivalency of sets in mind, and with account of the expression (2) as well as above assumption that the uncertainty is fully eliminated after receiving the message I_ν, the amount of information transferred can be defined as the entropy of the ν-th coding word $Z_{\nu\,\text{IMA}}$

$$D(I_{ŕJ}) = H\left(Z_{\nu\,\text{IMA}}\right) = \\ = -\sum_{\nu=1}^{l} P\left(Z_{\nu\,\text{IMA}}\right) \log P\left(Z_{\nu\,\text{IMA}}\right) = \log l, \tag{9}$$

where $D(I_{ŕJ})$ is the amount of information transferred by the ν-th message I_ν; $P\left(Z_{\nu\,\text{IMA}}\right)$ – probability of formation of the word $Z_{\nu\,\text{IMA}}$, and $\sum_{\nu=1}^{l} P\left(Z_{\nu\,\text{IMA}}\right) = 1$.

For an equal probability of every message I_ν, the amount of information $D(I_{ŕJ})$ transferred through the transmission channel takes a maximum value. This value is determined by only the number of different messages of the source or, which is equivalent, by the length of alphabet in the used IM.

This approach fixes the general properties of the system and its alphabet only in a formal manner and reflects the only information transferred to the operator in fact. Therefore, it is invariant relatively to the form of data representation. Information redundancy of a visual alphabet is not pronounced here and is not taken into account. However, visual image corresponding to the message and defined by IM influences essentially on the level of information security, which is directly bound

© 2010, V. Lashkaryov Institute of Semiconductor Physics, National Academy of Sciences of Ukraine

with fidelity of recovering the initial symbol by the addressee [6, 8, 10].

Therefore, let us consider the information redundancy of data in the system as being based on the analysis of the Z_{IM} code used in it. As shown above, on the one hand, the set $\mathbf{Z}_{\text{IMA}}$ describing Z_{IM} is equivalent to all sets that reflect transfer of messages. On the other hand, this code is related with visual images of the formed symbols, since it can be described with the set $\mathbf{Z}_{\text{IMd}}$, for which the equivalency and bijection with the set $\mathbf{A}$ of IA elements is realized by technical means. In this form of message representation, their theoretically possible amount increases up to the values inherent to the number of scramblings with unlimited reiterations $l_Z = 2^p$ [14]. Then, the set of words for the code exciting IA Z_{IM} can be written as

$$\mathbf{Z}_{\text{IMG}} = \mathbf{Z}_{\text{IMA}} \cup \mathbf{Z}_{\text{IMD}}, \tag{10}$$

where $\mathbf{Z}_{\text{IMA}}$, $\mathbf{Z}_{\text{IMD}}$ are the sets of allowed and forbidden words, respectively.

It is obvious that $\mathbf{Z}_{\text{IMA}}$ is described with the expression (6) for $q = l$, and the set of forbidden logic combinations can be represented as

$$\mathbf{Z}_{\text{IMD}} = \left\{ Z_{1\,\text{IMD}},\ Z_{2\,\text{IMD}},\ \dots,\ Z_{\mu\,\text{IMD}},\ \dots,\ Z_{(2^p - l - 1)\,\text{IMD}},\ Z_{(2^p - l)\,\text{IMD}} \right\}, \tag{11}$$

where $Z_{\mu\,\text{IMD}}$ is the μ-th forbidden word, and $\mu = \overline{1,\ (2^p - l)}$.

Besides, the probability for the words $Z_{\mu\,\text{IMD}}$ to appear when the system operates is $P\left(Z_{\mu\,\text{IMD}}\right) = 0$ for all the values $\mu = \overline{1,\ (2^p - l)}$. Therefore, if one estimates the amount of information in the case when the code Z_{IM} is represented with $l_Z = 2^p$ words, then using the formula (2) one can obtain the value $D\left(I_{\acute{r}J}\right) = \log l$. It is identical to the result obtained in the expression (9), which confirms the correctness of the above assumptions as well as legality for estimation of system information characteristics if using the properties of the Z_{IM} code described in the form (10).

From the equivalency of all the sets used for coding messages in the system as well as in the view of hardware realization of equivalency between the sets $\mathbf{A}$ and $\mathbf{Z}_{\text{IMd}}$, it follows that the $\mathbf{Z}_{\text{IMD}}$ set of forbidden words is corresponded with the equivalent set of forbidden symbols. It determines the redundancy of visual information that is related with geometrical features of synthesized visual images and can be quantitatively estimated with account of Z_{IM} code properties. With this aim, we shall use the factor of redundancy that takes into account blocked by technical means possibilities to form off-nominal visual images in IA,

$$K_{\text{R IM}} = \frac{H\left(\mathbf{Z}_{\text{IMG}}\right) - H\left(\mathbf{Z}_{\text{IMA}}\right)}{H\left(\mathbf{Z}_{\text{IMG}}\right)} = 1 - \frac{H\left(\mathbf{Z}_{\text{IMA}}\right)}{H\left(\mathbf{Z}_{\text{IMG}}\right)}, \tag{12}$$

where $K_{\text{R IM}}$ is the factor of redundancy for this IM, $H\left(\mathbf{Z}_{\text{IMG}}\right)$, $H\left(\mathbf{Z}_{\text{IMA}}\right)$ are entropies of messages that can be theoretically synthesized with the Z_{IM} code and are practically formed in this system, respectively.

For the considered code, the formula (12) with account of expressions (9), (10) and (11) takes a look

$$K_{\text{R IM}} = 1 - \frac{\log_2 l}{p}. \tag{13}$$

The obtained expression (13) can be used for estimation and juxtaposition of IM information properties. The factor $K_{\text{R IM}}$ is equal to zero in the case of data representation forms free of redundancy, which corresponds to the absence of forbidden words in the code Z_{IM}, i.e., the set $\mathbf{Z}_{\text{IMD}}$ is empty: $\mathbf{Z}_{\text{IMD}} = \varnothing$. In this information representation, the length of message alphabet and the number of IA elements are related with the expression $l_0 = 2^p$. The unity value of the factor $K_{\text{R IM}}$ is a theoretical limit and cannot be reached in real systems, as it is indicative of an unlimited redundancy of IM.

4. Results and discussion

It is practically interesting to juxtapose the bar graph and symbolic forms of digital information representation that are the most widely used in radio-electronic systems for various purposes. Despite a lot of versions in geometrical designing the visual images for symbolic IM, series articles are mainly based on *7*- and *9*-segment solutions for the polygramm as a consequence of an optimal relation between realization difficulties and the level of ergonomic characteristics. Less popular are the *5*- and *6*-segment symbols, but they are also interesting due to minimized technical expenses for information representation [12, 13].

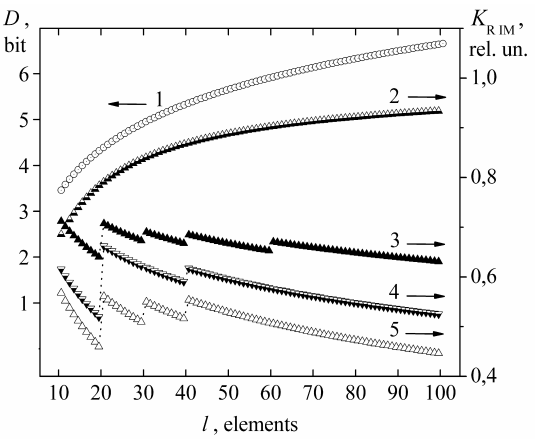

Fig. 1. The amount of transferred information in the system (*1*), the redundancy of a bar graph information model (*2*) and of the symbolic ones (*3 - 5*)

Depicted in Fig. 1 (plot *1*) are the results of calculations for the amount of transferred information $D(I_{f,J})$ in accord with (9) concerning the systems with the alphabet length up to 100 symbols. It is clearly seen that this amount grows with increasing the number of elements from the set of messages. It is valid for all the considered IMs. Adduced at the same figure are the dependences for the factor of redundancy $K_{\mathrm{R\,IM}}$ for various IMs, which were calculated using the expression (13): the plot *2* reflects properties of a bar graph form for imaging the message, while plots *3* to 5 correspond to the symbolic ones consisting of *9*-, *7*-, *6*-segment symbols.

The obtained data show that with increasing the alphabet length the redundancy of a bar graph IM increases (plot *2*), too, while for the symbolic ones (plots *3 - 5*) it is reduced. Moreover, with growth of the number of IA elements the character of $K_{\mathrm{R\,IM}}$ changes for bar graph IMs is qualitatively similar to that of the information amount D that is transferred by these messages (plot *1*). It is indicative of a different nature of visual images generated when synthesizing symbol or bar graph IMs. In the case of symbol forms imaging the messages, the growth in the number of elements in the polygramm of every digit is accompanied with the growth of the redundancy factor, and the sharp change of $K_{\mathrm{R\,IM}}$ corresponds to involving the polygramm elements not used earlier to form the image when the IM alphabet is successively lengthened.

5. Conclusion

The obtained quantitative estimation of the information redundancy when coding the optical signal in the connection channel between technical means and an operator allows juxtaposition of data protection levels in the visual channel of various ergatic systems in a rather simple way. The growth in the redundancy increases the probability that the operator can recover the message corrupted by any source of threats.

The numeric value of the information redundancy factor in the IMs the most widely used in control systems has shown that the bar graph form considerably outstrips the symbol representation by this parameter, when the message alphabet length exceeds 10 – 15. Moreover, when the volume of transferred information grows the redundancy of the bar graph IM grows, while that of all the digital forms reduces. These properties confirm the purposefulness of using the bar graph data representation to increase the security level of visual information in the ergatic systems.

References

1. V.A. Gerasimenko, *Information Protection in Computer-Aided Data Proceeding Systems.* In 2 vol. Vol. 1. Moscow, Energoatomizdat, 1994 (in Russian).
2. V.A. Gerasimenko, *Information Protection in Computer-Aided Data Proceeding Systems*. In 2 vol. Vol. 2. Moscow, Energoatomizdat, 1994 (in Russian).
3. V.A. Gerasimenko, A.A.Malyuk, *Fundamentals of information protection.* Moscow, MOPO, MIFI, 1997. 537 p. (in Russian).
4. A.A. Malyuk, *Information security: conceptual and methodological bases of the information security.* Moscow, Goryachaya liniya-Telekom, 2004 (in Russian).
5. B. Sklar, *Digital Communications: Fundamentals and Applications.* Prentice Hall, N.Y., 2001.
6. R. Leveugle, V. Maingot. On the Use of Information Redundancy When Designing Secure Chips // *IEEE Design and Diagnostics of Electronic Circuits and Systems,* 2006. – P.139-140.
7. A. J. Viterbi, J. K. Omura, *Principles of Digital Communication and Coding.* Dover Publications, 2009.
8. V.A. Litvinov, V.V. Kramarenko, *The control of reliability and recovering of the information in human-machine systems.* Kiev, Tekhnika,1986 (in Russian).
9. R. E. Blahut, *Theory and Practice of Error Control Codes.* Addison-Wesley, 1983.
10. L.F. Kulikovsky, V.V. Motov. *Theoretical bases of information processes*. Vysshaya shkola, 1987 (in Russian)
11. R. H. Morelos-Zaragoza, *The Art of Error Correcting Coding.* Wiley & Sons, Ltd., Chichester, England, 2006.
12. S. Gage, M. Hodapp, D. Evans, H. Sorenson, *Optoelectronics Applications Manual*. McGraw-Hill Book Company, New York, NY, 1977.
13. F.M. Yablonsky, Yu.V. Troitsky, *Information Imaging Tools*. Moscow, Vysshaya shkola, 1985 (in Russian).
14. V.P. Sigorsky, *Mathematical apparatus of the engineer.* Kiev, Tekhnika,1975 (in Russian).