Zbornik radova Ekonomskog fakulteta Brčko
Proceedings of Faculty of Economics Brcko

PROBABILISTIC EVALUATION MODEL OF INFORMATION EXPOSURE ON THE INTERNET

Irina M. Glotina[a,1]

[a]Federal State Budgetary Educational Institution of Higher Education "Perm State Agro-Technological University named after Academician D.N. Pryanishnikov", Perm, Russia

ABSTRACT

The article deals with the influence of information on the Internet. The growth and importance of resources of a new type – social networks, forums, blogs – as a means to discuss different events, problems, distribution of opinions, influencing actions of network users, have an increasing influence on the existence of modern man. A specific feature of Internet resources is the principle of attracting users to familiarize with the content and its multiple replication. While questions of credibility, reliability, objectivity of information are not taken into account. The use of information flows by hackers suggests the possibility of providing informational-psychological influence on emotional and sensitive, rational-volitional and intellectual spheres of the subconscious and psyche receptors, resulting in the formation of their predictable views, opinions, behaviour that directly or indirectly contribute to the objectives of an attacker. All this testifies to the topicality of the problems of assessing the influence of information on users of Internet resources. The article highlights the main factors complicating the analysis in the field of information effects: the rapid development of information technologies and means of communicating informational messages to the recipients; the inability to predict the patterns of behaviour of subjects of impact; inability to give an accurate estimate of what the message reached the recipient, and to assess the impact of informational influence. The author justifies a probabilistic approach to assessing the impact of Internet resources on users' opinions. The results of mathematical calculations are graphically illustrated and confirmed by the statistical information obtained with the help of modern web analytic systems: GfK Group, Rambler Top 100, Brand Analytics, Yandex Statistics. The results obtained by the author during the survey of the Internet users revealed resources that have a large audience and high frequency of attendance and are more potential threats to information security. Probabilistic assessment allows to identify the main areas where the buildup of forces and means of ensuring protection from the negative effects of information.

Introduction

Growing the importance of new types of resources - social networks, forums, blogs - as a means of discussing various events, problems, spreading opinions that affect the actions of network users, has an increasing influence on the existence of modern man, playing an important role in shaping him and as a person, and as a member of society.

Feature of Internet resources is the principle of attracting users to familiarize themselves with content and its repeated replication. At the same time, the issues of reliability and objectivity of information go to the background. Taking into account modern threats such as international terrorism and cyber fraud, the situation is not as harmless as it might seem at first glance, since information impacts can become a powerful weapon, and its application to energy, transport, finance and other critical technologies can to be catastrophic (Zegda, 2007, 28:33).

---

[1] glotina-i@yandex.ru

The use of information flows by malefactors presupposes the possibility of providing information-psychological influence on the emotionally sensitive, rationally-volitional, intellectual spheres of the subconscious and the psyche of objects of influence, the result of which is the formation of predicted views, opinions, behavioral reactions directly or indirectly contributing to the accomplishment of the attacker's goals. Information impact issues are considered in the works of Belov, Gubanov, Novikov, Chkhartishvili and Manoilo (Belov, 2010; Gubanov et al., 2009; Gubanov et al., 2010; Manoilo, 2003).

All of the above indicates the relevance of the problem of assessing the information impact on users of Internet resources. However, there are many factors that make it difficult to conduct analysis in such an area as the information impact. Among them we can distinguish:

-   rapid development of information technologies and ways of communicating information messages to addressees;
-   the inability to predict the behavior patterns of the subjects of information impact;
-   impossibility to give a reliable, sufficiently accurate assessment of the fact that the message has reached the addressee, and also to assess the impact of information impact in quantitative or value terms.

1. Research design and methodology

To solve the problem of quantitative estimation of information impact, we use the standard apparatus of probability theory.

-   To ensure that the information content has a proper impact on the user of the Internet, it is necessary to simultaneously fulfill several conditions, among which we have identified the following main ones:
-   the content should be brought to the user; this becomes possible if the user has independently switched to the address of the resource; for this resource must be in demand, popular, affordable. The second option is to send content to users; for this can be used advertising on popular sites, spam mailing to e-mail;
-   the user is familiar with the content; for this, the content should be relevant, attractive, have an intriguing title, contain interesting photo-audio-video material;
-   the user has perceived the information as reliable; for this, it must come from an authoritative source, which can be a media outlet, a television broadcast, an official, a friend or a group of friends whose opinions the user listens to.

For simplicity, we will assume that these events are independent, and the probability that the message will affect the user will depend on the likelihood of their simultaneous occurrence and can be expressed as follows:

$$p_0 = p_1 \, p_2 \, p_3 \tag{1}$$

where
p_1 is the probability that the content is available to the user;
p_2 - the probability that the user has read the content;
p_3 - probability that the user perceived the information as authentic.

The probabilities of the events p_1 and p_3 largely depend on the individual psychological characteristics of the user, such as age, social status, education, occupation, moral principles, habits, and the like.

Probability p_2 - depends on the position of the Internet resource in the search engines, which is determined by the capabilities of individuals trying to use the resource to provide information impact. If such person is the owner of the site or is engaged in its administration, then he has the opportunity to display a message on the most visited page and ensure that an audience is attracted to it. If an ordinary user submits a message, he can use the "gray" methods of "promotion" (frequent repetition, organization of a separate discussion topic, allotment) to attract attention to information, and also use various gaps and vulnerabilities when moderating the resource. In this regard, in order to maximize the likelihood of getting to know the information, recently attackers are increasingly using various automation tools and robots that

send messages to a huge amount of resources. The rating of the resource can be influenced by special technologies.

Assume that m of the independent users were acquainted with the message, then the probability of information impact provided by the message on these m users can be represented as:

$$p = 1 - \prod_{j=1}^{m}(1 - p_0) \tag{2}$$

where

m is the number of users who have read the message posted on the Internet resource.

If, for simplicity, we assume that expression (1) is equally probable for each user, then we get:

$$p = 1 - (1 - p_0)^m \tag{3}$$

For a graphic illustration of the behavior of the function p, several different values of p_0 were used: 0.001; 0.003; 0.005; 0.01; 0.02; 0.03 (Figure 1).

Thus, the greater the number of users who have read the attacker's message, the higher the probability that it will have an information impact (4):

$$\lim_{m \to \infty}(1 - (1 - p_0)^m) = 1 \tag{4}$$

In general, according to the Russian branch of the research group GfK (Gesellschaft fur Konsumforschung) Group, in Russia in 2016, 84 million people aged 16 and older are Internet users, accounting for 70.4% of the population (Research GfK, 2017). To assess the number of visits to Internet resources, content views, the number and nature of user requests, we used modern web analytics systems, whose capabilities and functionality allow us to obtain the necessary statistics.

Fig. 1: Changing the probability of information impact depending on the number of users
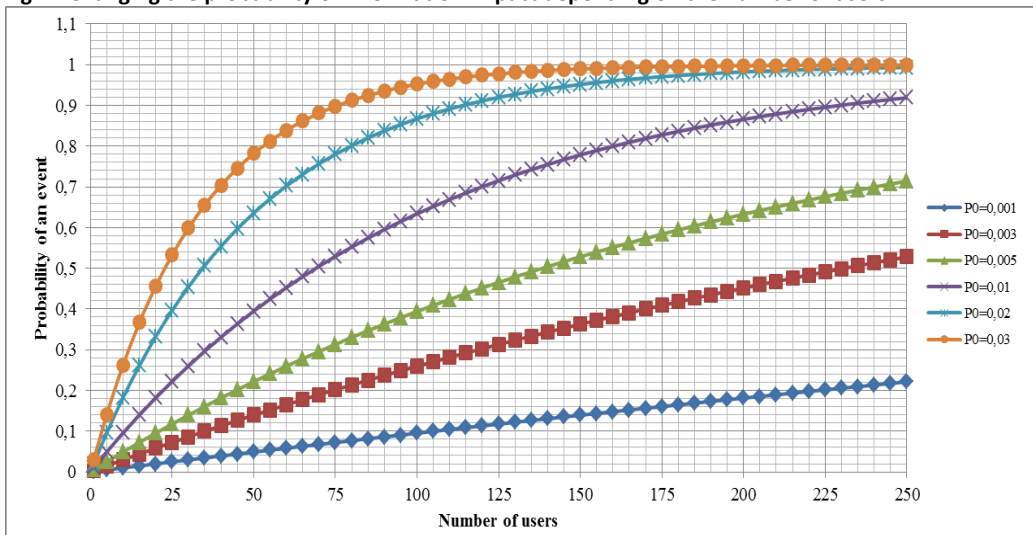
So, in order to study the popularity and availability of information resources, we used the Rambler / Top100 rating system, which allows users of the service to have detailed statistics on the website over the past 24 hours, as well as the dynamics of its visits in the last two months. The study was conducted within Russia, the research topic - news and media. Using the tools of the Rambler / Top100 system, statistics of the popularity of radio, television, newspapers and magazines were obtained. The sampling period is February 2017 (Rambler's Top100, 2017). The study used the following terminology:

Visitors – the number of unique visitors to the site for the selected period;
Views – the number of pages viewed on the site for the selected period.

The number of visitors and views presented in Table 1 of the most popular sources of information testify, firstly, about the rating of the popularity of resources among Internet users, secondly about the availability of information and confirm the above conclusion.

Table 1: Rating of the most popular sources of information

Rating	Radio	Visitors	Views
RADIO TOP-10 (among 100)			
1	Эхо Москвы	8 191 972	94 358 882
2	"Европа Плюс"	2 477 565	8 226 259
3	"Русское Радио" Москва	800 557	4 796 674
4	Love Radio	675 096	4 399 859
5	Радио DFM Москва	354 947	2 621 298
6	Авторадио - Первое автомобильное радио	312 937	731 245
7	Радио "Хит FM" Москва	165 389	695 298
8	Радио MAXIMUM — ведущая pop-rock станция	124 970	819 309
9	"Радио Дача"	145 504	482 771
10	Сайт ГТРК "Ставрополье" - новости Северного Кавказа	145 147	564 041
TV TOP-10 (among 198)			
1	Вести.Ru: новости, видео и фото дня	16 372 586	96 002 717
2	Новостной ресурс "RT News Channel"	10 809 940	51 297 348
3	Телеканал "Звезда"	9 101 304	36 996 004
4	НТВ.ru - новости, видео, прямой эфир телеканала НТВ	8 195 996	31 574 143
5	Вокруг ТВ	7 250 221	17 550 907
6	Телевидение Алтайского края "Катунь 24"	309 086	978 801
7	"Камелот Программа"	240 189	1 180 105
8	"Квартирный вопрос"	295 433	1 311 435
9	Вести Сочи	224 305	750 700
10	Телекомпания "Ариг Ус"	180 924	1 046 860
NEWSPAPERS AND MAGAZINES TOP-10 (among 1056)			
1	"Аргументы и Факты"	16 410 604	45 325 893
2	Газета "Московский Комсомолец"	11 729 422	66 700 175
3	"Русская семерка" - о России традиционной и современной	4 677 902	25 231 734
4	Газета "Вечерняя Москва"	3 672 600	6 304 243
5	"АвтоВзгляд"	3 211 366	7 427 669
6	ИД "Собеседник"	2 509 813	6 056 773
7	«Петербургский дневник» - ежедневное издание правительства Санкт-Петербурга	2 589 543	4 878 046
8	Журнал "Эксперт"	1 919 269	4 607 640
9	"Аргументы Недели"	1 540 811	4 052 736
10	Настоящая Леди - онлайн-журнал для женщин	1 728 704	2 266 074

Analytical system Yandex Statistics demonstrates the audience of users on Yandex.ru for the first half of 2017 (Yandex Statistics, 2017).

Table 2: The audience of users of the search engine Yandex.ru for 2017.

Month	Monthly audience,% of the Internet	Monthly audience	Average daily audience,% of the Internet	Average daily audience
January	90,0%	52 736 130	66,1%	26 954 140
February	89,2%	51 693 570	66,1%	27 614 730
March	90,4%	52 802 400	66,2%	27 838 100
April	89,3%	52 166 190	65,9%	27 212 830
May	89,1%	52 667 070	66,0%	26 926 970
June	89,0%	52 040 260	65,7%	25 668 840

Among the Internet resources, social networks play an increasingly important role in disseminating information, ideas, and providing information impact to users. This is due to the presence of communities of interest, the opportunity to express their opinions and learn the views of friends. The number of network participants is hundreds of millions. People, as a rule, are simultaneously participants of several social networks. To study the audience of social networks, the author used the system Brand Analytics. The study was conducted among Russian users, the study period was in February 2017 (Brand Analytics, 2017). The results of the study presented in Table 3 are comparable with those given above.

Table 3: Results of the study of the audience of social networks

Name of the resource	Number of Authors	Number of posts
Social network		1 113 535 014
Including VKontakte	17 781 232	371 902 588
Facebook	1 219 157	53 820 671
Instagram	4 656 991	98 925 636
Microblogging twitter.com		333 246 062
Video (youtube.com)		72 789 124
Messengers (telegram.org)		550 864
Blogging		12 320711
Forums		17 403 324
Reviews		10 337 457
News		15 204 115

The results in Tables 1-3 show that the Internet is the most favorable environment for information impact.

The results obtained by the author in the course of the Internet users' questionnaires allowed to construct a variational series and to reveal some regularities in the use of Internet resources. Variants of the x_i series are the number of Internet resources used by users on a regular basis. The number of observations n is the number of respondents who participated in the survey. According to the results of the study, n = 250. Users regularly view up to 20 resources, that is, x_i takes values from 1 to 20. The grouped row is presented in the table 4.

Table 4: Variation series

Number of Internet resources, x_i	Frequency (number of users), n_i	Frequency (share of users) $w_i = \frac{n_i}{n}$	Accumulated frequency $n_i^{\text{нак}}$	Accumulated frequency $w_i^{\text{нак}} = \frac{n_i^{\text{нак}}}{n}$
1	3	0,01	3	0,01
2	37	0,15	40	0,16
3	22	0,09	62	0,25
4	25	0,10	87	0,35

5	38	0,15	125	0,50
6	35	0,14	160	0,64
7	26	0,10	186	0,74
8	14	0,06	200	0,80
9	10	0,04	210	0,84
10	7	0,03	217	0,87
11	3	0,01	220	0,88
12	3	0,01	223	0,89
13	3	0,01	226	0,90
14	4	0,02	230	0,92
15	3	0,01	233	0,93
16	4	0,02	237	0,95
17	3	0,01	240	0,96
18	3	0,01	243	0,97
19	4	0,02	247	0,99
20	3	0,01	250	1,00
TOTAL	250	1,00		

The study shows that 87% of respondents regularly use up to 10 different resources, while the proportion of users who regularly view more than 10 resources is only 13%.

A graphic representation of the variation series is shown in Figure 2 in the form of a polygon.
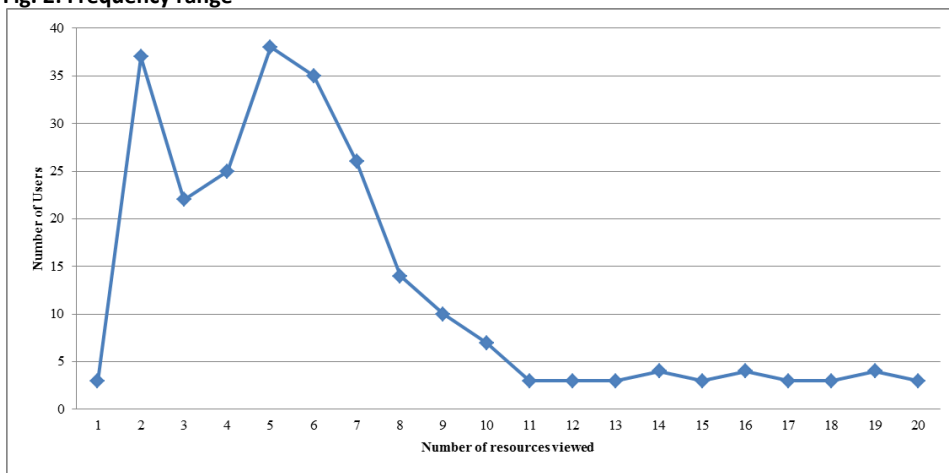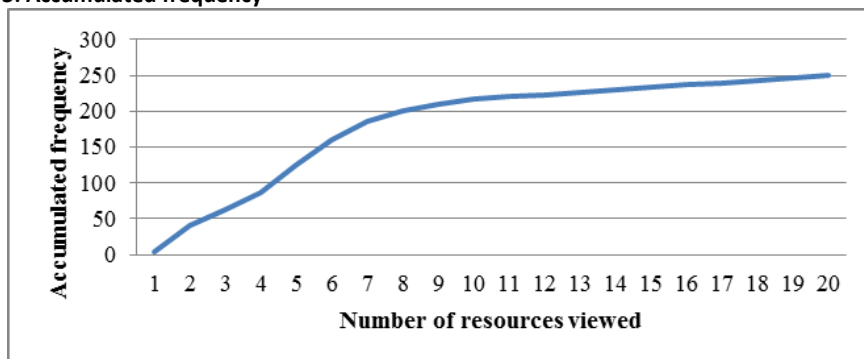
Fig. 2: Frequency range

Figure 3 shows the graphical interpretation of the accumulated frequency.

Fig. 3: Accumulated frequency

Based on the data presented in Table 4, ˉx was calculated, the arithmetic mean of the variation series:

$$\bar{x} = \sum_{i=0}^{20} x_i w_i = 6{,}44 \approx 6 \tag{5}$$

Thus, known resources that have a large audience and a high frequency of attendance carry more potential threats to information security. The uncertainty of the outcomes of the development of negative impacts on the Internet justifies the application of methods of probability theory. Probabilistic assessment allows defining the main directions, where it is necessary to build up the forces and means that provide protection against negative information impacts

Conclusion

The article highlights the main factors complicating the analysis in the field of information effects: the rapid development of information technologies and means of communicating informational messages to the recipients; the inability to predict the patterns of behaviour of subjects of impact; inability to give an accurate estimate of what the message reached the recipient, and to assess the impact of informational influence. The results of mathematical calculations are graphically illustrated and confirmed by the statistical information obtained with the help of modern web analytic systems: GfK Group, Rambler Top 100, Brand Analytics, Yandex Statistics. The results obtained by the author during the survey of the Internet users revealed resources that have a large audience and high frequency of attendance and are more potential threats to information security. Probabilistic assessment allows to identify the main areas where the buildup of forces and means of ensuring protection from the negative effects of information.

References

1. Belov, P.G. (2010). Forecasting the influence of Russian media on national security. *Space and time* No. 1, pp 58-68.
2. Brand Analytics. (2017). [Electronic resource]. URL: https://br-analytics.ru/ (circulation date March 10, 2017).
3. Gubanov, D. A., Novikov, D.A., Chkhartishvili, A.G. (2009). Models of Information Influence and Information Management in Social Networks. *Problems of Management*, pp. 28-35.
4. Gubanov, D.A., Novikov, D.A., Chkhartishvili, A.G. (2010). *Social networks: models of information influence, management and confrontation* / Ed. member corr. RAS, D.A., Novikov.Moscow: Publishing House of Physical and Mathematical Literature, p. 228.
5. Manoilo, A.V. (2003). Information confrontation in the conditions of psychological warfare. *Law and Law* No. 12, pp. 31-34.
6. Rambler's Top100. (2017). [Electronic resource]. URL: https://top100.rambler.ru (circulation date March 10, 2017).
7. Research GfK (2017). Trends in the development of the Internet audience in Russia. [Electronic resource]. URL: http: //www.GfK.ru (reference date is April 7, 2017).
8. Yandex Statistics. (2017). [Electronic resource]. URL: https://stat.yandex.ru/Russia/Portal (circulation date August 17, 2017).
9. Zegda, P.D. (2007). Ensuring the security of information in the context of creating a unified information space // Protection of information. *INSIDE* № 4. 4. pp. 28-33.