The question simulation of internet data exchange, according the features of functioning the hybrid content delivery networks

P A Filimonov¹, T Y Izotova¹, S S Gorelik¹, S S Eskov¹ and D I Mutin²

¹Russian Federation Security Guard Service Federal Academy, Pribirostroitelnaya 35, Orel, Russia, 302034, Russian Federation
²Mechanical Engineering Research Institute of the Russian Academy of Sciences, 4, Maliy Haritonievsky, Moscow, 101000, Russian Federation

E-mail: lev@academ.msk.rsnet.ru

Abstract. The article deals with an approach to modelling request routing and resource management processes of content-oriented services, based on the concept of hybrid content delivery networks. Proposed approaches to the formation characteristics of the Internet data exchange model in the framework of solving the problem efficiency increasing of the functioning information monitoring systems.

1. Introduction

Nowadays in some subject areas connecting with solving the problems of information monitoring exist the problem of rational using of limited resources, necessary for monitoring process organization in the conditions of prior uncertainty. In a general view mathematical and technological bases of difficult processes information monitoring are considered in the point [1]. Conceptual scheme of information monitoring systems is presented in the point [2].

Especially actual problem of limited resources rational using is shown in systems monitoring of primary channels broadcasting information in the Internet. A great number of loading sources causes its functioning efficiency decrease on certain rate in each case [3].

Taken research shows that existing methods and algorithms, directed on such systems informational monitoring efficiency functioning increased, are based in graph and stream models of Internet structure, determining the features of the set of nodes interaction – autonomous systems (AS) [4, 5].

The feature of the theoretic-graph approach to the modeling of the Internet logical structure is empirical selection of λ parameter, determining indicative character of the arbitrary AS probability distribution law and weak effect of preferable accessions principle [5]. Existing stream models of data transmission process in the Internet are oriented on consideration of principles of logical channels forming between AS subsets «from the end to the end» in the conditions of dynamic routing process and the existence of routing information collecting system [6].

Pointed approaches to the modeling do not consider existing trends of the Internet logical structure improvement, namely overlay networks active development and networks, based on the principles of equal peer-to-peer interaction [7]. It’s evident as functioning of such network on the base of existing
Internet structure demands coordination the questions of channels and buffer resources of transport network distribution. It has an impact on the type and parameters of the loading model for the existing information monitoring systems.

The article presents the approaches for the Internet logical structure model forming, considering hybrid (CDN-P2P) overlay content delivery network functioning in its frames and the assumption about information monitoring systems functioning principles are made, for that pointed model is a statistical model.

2. Information monitoring system functions, working within a priori uncertainty of data exchange process parameters by the set of observation objects.

Information monitoring system definition as a special information system intended for some process status assessment or a problem on the base of its fragmentary, unreliable, perhaps, doubtful information and modeling of its possible development options is given in [2, 4]. Wherein in [2] it’s emphasized that inputs of information monitoring systems are the information about the process or the object of observation. In general figure 1 presents the place and functions of information monitoring system.

![Figure 1. Information monitoring system place and functions.](image)

It’s obvious, that information monitoring system efficiency greatly depends on quality of access function to information field realization, which is a set of IO for which data exchange process is organized. Access to the Internet information field information monitoring system gets within a lot of communication channels with different physical transmission medium (PC).

Determine discussed above set of PC as \( C \) and the set of receiving devices of IO as \( P \). The sets of \( C \) and \( P \) not equipotent and the power of the \( C \)-set exceeds the power of the \( P \)-set. \( C \)-set is possible to present in the form of two subsets: \( C^R \) is the subset of available PC, within the information monitoring system receive IO, equipotent to \( P \)-set and \( \bar{C}^R \) – subset of available PC, within the information monitoring system doesn’t receive IO.

\[
C = C^R \cup \bar{C}^R, \quad C^R \cap \bar{C}^R = \emptyset
\]
Irrational division of \( C \)-set into \( C^R \) and \( C^R \) subsets reduces the efficiency of information monitoring system’s access function (figure 1). Its increase we can get by solving the problem of allocating from the \( C \)-set such \( C^R \)-subset within maximum quantity of IO getting is possible, containing information indicators of the process or the object of observation. Herewith IO extend within logical structure of the Internet, based on the process of logical channel forming between the source and the destination of the IO – LC\(_{io}\).

Let’s define the set of PC within LC is formed as \( C^S \). As in case of \( C^R \) and \( C^R \) subsets for \( C^S \)-subset the condition works \( C^S \subseteq C \). At the same time \( C^S \)-set and \( C^R \) and \( C^R \) subsets are not equipotent (2).

\[
|C^S| < |C^R \cup C^R|
\]  

(2)

Generally, \( C^S \), \( C^R \) and \( C^R \) are in relations composition \( \rho_1 \circ \rho_2 \) (3).

\[
\rho_1 \circ \rho_2 = \left\{ \left(c^R \circ c^R\right) : c^R \in C^R \land c^R \in C^R \land \left( \exists c^S \in C^S \left(c^R \circ c^S\right) \in \rho_1 \land \left(c^S \circ c^R\right) \in \rho_2 \right) \right\}
\]  

(3)

It’s possible to make a hypothesis that owing to the problem of effective IP-packets routing by the Internet transport system solving loading relations composition (3) in different periods of time \( t_m \) will be different and will depend on the features of the Internet logical structure. Thus, solving the problem of \( C^R \)-subset allocation mustn’t be considered determined.

To solve the problem of type of dependence determination, for the made above hypothesis it’s necessary to consider the existing and perspective approaches to the Internet logical structure modeling.

3. The existing approaches to the modelling of Internet logical structure.

The dynamic routing process modeling, maintained by the relevant protocols, in the existing transport networks with packet switching, is generally based on graph models and solving problems of optimal (the shortest) path searching with the use of algorithms of Dijkstra, Bellman-Ford and others.

In relation to the Internet, specified approach is based on its logical structure representation as a digraph \( G = (V, E) \), where \( V \) is a set of AS acting as a source and destination of IO data, and \( E \) is a finite family of ordered AS-pairs, connected by PC.

Digraph edge orientation reflects the direction of route announcement distribution. BGP protocol is used in switching nodes to maintain this digraph connectivity at the level of protocol implementation. The set of existing in network LC is determined of the attribute AS_PATH of BGP Update packet. In this attribute the beginning and the end of LC is extreme left \( v_i \) and extreme right \( v_j \) AS respectively.

LC between \( v_i \) and \( v_j \) in the moment of time \( t_m \) is the set of AS \( \{i,j,m\} \), which is a simple chain in digraph \( G \) (4). On condition of \( |i,j,m| \geq 3 \) transit AS \( v_k \) where \( k \neq i,j \) are alongside with extreme AS in the structure of LC.

\[
i_{i,j,m} = \{v_i,\ldots,v_k,\ldots,v_j\}_m, n_{i,j,m} \in R_m
\]  

(4)
Proceeding from (4) that the bundle of LC’s \( \hat{h}_{i,j,m} \) is determined as the set of LC from initial \( v_i \) to \( v_j \) AS in moment of time \( t_m \) (5). The power of bundle of LC’s \( \hat{h}_{i,j,m} \) is the amount of LC’s entering it.

\[
\hat{h}_{i,j,m} = \left\{ \left[ h_{i,j,m} \right]_{1} \ldots \left[ h_{i,j,m} \right]_{z} \ldots \left[ h_{i,j,m} \right]_{Z} \right\}
\]

(5)

At the same time from the point of view of BGP-routing protocol realization this model has number of disadvantages:

- a part of routes attributes for the route information collecting system is inaccessible;
- a part of local settings of the router from the point of view of route information collecting system is inaccessible.

To solve the problem of content delivery with the required quality, routing protocols maintain the functions of transport network channels resource management on the base of channels loading balancing. Obviously that the Internet structure model (4, 5) cannot adequately display this process because it’s impossible to define how many LC for the data transmission between \( v_i \) and \( v_j \) AS router uses.

This problem solves stream models using which alongside with the set of logical channels determination and the bunches of logical channels allow to define data stream optimal distribution between \( v_i \) and \( v_j \) AS. In [6] in the quality of methodological basis of multipath routing stream models, querying network device, algebraic, tensor calculus different integral-differentials equation and other is considered.

In general, in the frame of stream model, network structure is also defined by the digraph \( G=(V,E) \). However, in addition in each arc \( (i,j) \in E \) the parameter \( c_{i,j} \), characterizing its bandwidth, is assigned. In the frame of this network structure data exchange is presented by the set of streams \( K \), where each k-stream from this set is compared with the intensity parameter \( d_k \), as well as source and receiver nodes \( s_k \) and \( t_k \) respectively. Independent variable in this case is \( x_{i,j}^{k} \), value of which determines the part of k-stream in the arc \( (i,j) \in E \).

Within this model the problem of the set of optimal routes between \( v_i \) and \( v_j \) AS determination is solved. Usually this task boils down to solving the task of nonlinear programming with the optimality criterion (6)

\[
\min_{x} \left( \max_{(i,j) \in E} a_{i,j} n_{i,j} \right)
\]

(6)

where \( n_{i,j} \) is the size of packet queue on i-node of the arc \( (i,j) \in E \) and \( a_{i,j} \) is a coefficient determining the weight of the arc \( (i,j) \in E \) in general network digraph structure.

Within the process of informational monitoring system functioning using the considered graph and stream network structure models, parametrically identified to the real Internet structure allows to form the set of training samples allowing to build specified information indicators appearance sample probability distributions defining the data about monitoring objects activity on the set of physical channels to which the set of information receiving devices is connected. Further obtained the sample probability distributions can be used by the synthesis of the prediction algorithm of the informational monitoring system instead of unknown true distributions.

The common problem of considered models is the lack of accounting currently widely used overlay networks functioning within the described Internet structure. Such networks having a significant
impact on the routing of the monitoring objects request to the Internet informational resources should include content delivery networks (CDN), peer-to-peer networks (P2P) and its hybrid versions.

4. Approach to the logical Internet structure modelling on the base of hybrid content delivery network.

In general the main task of CDN establishment is duplication of the content and/or services by the base of AS-source (origin server in terms of CDN) between several peripheral AS, which are named surrogate AS (surrogate server). The principle of surrogate AS location is connected with geographical proximity of the client AS. In this case the requests from clients AS are routed to the nearest AS with a minimal load (origin or surrogate). Such requests redistribution generally allows to plan the resource of transport network bandwidth more efficiently and to solve the tasks of scalability and improving customer service efficiency. Taxonomy and the architectural features of CDN networks are presented in [9, 10].

The works [11, 12] are detected to the modeling the process of requests routing in CDN. Generally the requests routing model determines the process of finding variables \( y_g^{kl} \), defining the using g server, \( g \in G_l \), as a source of \( l \)-th type content in \( k \)-th sender-receiver pair on the base of unknown variables \( x_{ij}^{kl} \) determining the share of \( l \)-th type content traffic in the \( k \)-th pair of the sender-receiver transmitted over the transmission path \((i, j)\) from \( i \)-th to \( j \)-th Internet transport router.

Development of such model is review of hybrid content delivery network (Hybrid CDN – HCDN). Logical HCDN structure is based on the integration CDN and peer-to-peer network (P2P). Approaches to the HCDN modeling are presented in [13]. Generalized HCDN network diagram is shown in figure 2.

![Generalized HCDN network diagram](image)

Figure 2. Generalized HCDN network diagram.

Modeling results, presented in [11-13] allow to assume that functioning CDN and HCDN within Internet have a significant effect on the model of its logical structure (4-6) that in its turn influence on the type and parameters of the load model for the existing systems of informational monitoring. Generally, subsystem functioning of these systems access should be considered in priori uncertainty of the types and parameters probability distribution laws of occurrence of IP-packets of the set of IO, generated by the set of objects or processes of observation on the set of physical channels \( C \) of Internet transport. Due to the fact that the dependence of losses in the operation of the information monitoring system is also unknown, the synthesis of traditional statistical model is impossible, that requires conducting the research based on the training samples forming on the base of empirical data about the decision, taken by the informational monitoring system in the past, which allow to construct sampling distributions for obtaining estimates of the average risk of losses and its minimization by the choice.
the rules of solving the problems of the distribution of the set of access devices $P$ on the subset of physical channels $C^R$.

5. Conclusion

Conducted research of the existing versions of the informational monitoring systems showed that suggested within its perfection models of logical Internet structure, based on the process of the feature space forming with the use of data packets parameters, used by the routers, not fully considers the presence of prior uncertainty of the data exchange process parameters, occurred in particular content delivery network functioning process.

From this the article presents the approach to the logical Internet structure modeling on the base of the approaches within hybrid content delivery network modeling.

References

[1] Okhotnikov E S 2006 Technological process monitoring systems of oil-and-gas producing enterprises: classification and mathematical modeling Oil and Gas Business 2 1-13
[2] Ryzhov A P 2007 Information monitoring of complex processes: technological and mathematical foundations Intellectual systems V 11 N 1-4 101-36
[3] Romansky R, Kirilov K 2018 Architectural Design and Modelling of a Web Based Application for GDPR Clarification AIP Conf. Proc., Proc. of the 44th International Conference on Applications of Mathematics in Engineering and Economics (AMEE’18) Vol 2048 p 060006
[4] Ryzhov A P 2015 Mathematical problems of the systems of evaluation and monitoring of complex processes Intellectual systems 19 1 6-20
[5] Raygorodsky A M 2012 Extreme problems of graph theory and the Internet: Tutorial (Dolgoprudny: Publishing House Intellect) p 104
[6] Romansky R 2019 A Survey of Informatization and Privacy in the Digital Age and Basic Principles of the New Regulation International Journal on Information Technologies and Security 11(1) 95-106
[7] Evseeva O Yu and Garkusha S V 2012 Overview of technological and theoretical solutions for QoS-based routing Telecommunication Problems 3(8) 24–46
[8] Zhi H L, Wang Y and Yang R 2012 An analysis and comparison of CDN P2P-hybrid content delivery system and model Journal of Communications 7(3) 232-45
[9] Dilley J, Maggs B, Parikh J, Prokop H, Sitaraman R and Weihl B 2002 Globally distributed content delivery IEEE Internet Computing September/October 50-8
[10] Bartolini N, Casalicchio E and Tucci S 2003 11th IEEE/ACM Int. Symp. On Modeling, Analysis and Simulation of Computer Telecommunications Systems, Orlando: Florida, USA
[11] Evseeva O Iu and Kader M B 2012 Mathematical model of request-routing in content delivery networks Information Processing Systems 9(107) 165-8
[12] Kabir M H, Manning E G and Shoja G C 2002 Int. Conf. on Computer and Information Technology: proc. of conference (Dhaka: Bangladesh) pp 315-20
[13] Evseeva O Iu and Kader M B 2014 Multi-level hierarchical method for complex resource management in hybrid content delivery network Information Processing Systems 8(124) 113-20