Load Optimization in Mobile Network using Traffic Engineering

Mohammad. A.A Al- Rababah
Northern Border University, KSA,

ABSTRACT
Modern computer networks use dynamic routing algorithms, since the static does not take into account the current load on the network.
The approach The application of gradient-projection algorithm for optimization of loading in mobile network is offered. **Conclusion the** problem of constructing the traffic and formulate the classical problem of load balancing and optimal routing.
Keyword: Network ,Mobile, Traffic, optimization , routing

1. INTRODUCTION
In a typical modern computer networks the presence of mixed traffic, and there is a steady increase in the share multimedia traffic [1, 2, 3]. Each type of traffic has its own requirements for quality of service (QoS) [4]. With regard, to that we formulate the basic requirements for QoS for voice and video. In [5, 6] outlines the parameters of QoS for voice.Voice quality, three factors affect the quality of QoS [7]: packet loss, delay and delay variation. Packet loss causes short-term gaps in conversation.

In [8] described two main types of video applications: interactive video and streaming video. Applications include interactive video audio codec G.711 for speech, so the requirements for the QoS parameters include requirements for voice traffic, namely, packet loss, delay and delay fluctuation. However, despite this, the traffic is video conferencing is radically different from voice traffic. For example, video conferencing traffic using variable packet sizes and variable packet rate. Speed video conferencing - it's sample rate of the video stream, but not the actual bandwidth, which requires a video call. Since the variables are used packet sizes and packet generation rate, it is difficult to accurately estimate the absolute value of overhead [9].Application streaming video to a less demanding QoS, since they are less sensitive to delays and are insensitive to fluctuations in the delay (due to buffering at the application level) [10]. However, streaming video may contain important information and, therefore, require guarantees QoS. Not great video content can be viewed as an Internet service (the service is worse than "Best Effort"), this suggests that these flows are working as long as there is bandwidth, but will be replaced in case of network congestion.

Based on the QoS requirements for voice and video, we define the basic requirements for QoS [11]. These include requirements for bandwidth, delay, change (fluctuations) of the delay, reliability, and relative load node. Unlike the multimedia traffic data is transmitted, usually with non-uniform flux density, in batches, and arrive at unpredictable intervals. In this connection, the bandwidth used for multimedia traffic, should be protected from data traffic, and vice versa.

2. REVIEW AND ANALYSIS OF EXISTING SOLUTIONS
Widely used two methods: a distance-vector routing and routing of the channel, their comparative characteristics are given in Table 1.

Table 1. Comparison of two methods of route atmen of the Problem.

Distance Vector Routing	Routing of channels
Each router sends routing information to its neighbors	Each router sends routing information to all routers
Sent the information is an estimate of the cost of the paths to all networks	Sent information - the exact value of the cost to the adjacent channels networks
The information is sent on a regular basis	The information is sent only when a change
The router sends information about the next hop using the distributed Bellman-Ford algorithm on the basis of the valuation of the way	The router builds the first description of the topology of the internetwork, and then can use any routing algorithm to determine information about the next hop.

Tools of the traditional route does not meet the quirements of QoS, the request load balancing of communication channels do not provide enough speed when you change routes on the network caused by the movement of bscriber systems. Another drawback is the need to" send regular updates of routing information even with a slight ovement of subscriber systems, in which the amount transferred via the network service traffic increases dramatically, reducing the bandwidth

One of the main objectives of the operation management of computer networks in the transmission of multimedia traffic is to organize an effective system of information delivery, which becomes especially important in mobile networks due to the constant movement of subscriber systems. To provide the necessary requirements to QoS parameters used by different modes of data transfer. One approach to solving this problem is to optimize the network by distributing the traffic, ie in real time to solve the problem of the dynamic assignment of routes that meets the requirements to the parameters of QoS, and to ensure uniform loading of the network. Providing these requirements by means of traffic engineering procedures (TE - Traffic Engineering) [12], whose main purpose is to provide a uniform network load. This is the main difference between this task from the task of optimal routing. The problem of optimal routing is to minimize the estimated (target) as a channel for each pair of sender and recipient (IE) subject to the restrictions

3. Meeting the challenge

Consider a single domain network consisting of nodes and links connecting them, which is represented as a directed graph $G = (N, L)$. Let N be the set of nodes (routers) n and L the set of links l network. As an alternative, we use the representation (i, j) to communicate with the node i node j. Bandwidth connection is denoted l bl. Let K be the set of pairs of IE k,, where *k,* $k \in K$, - (s_k, t_k), s_k input nodes , t_k - output nodes IE-pair k. The requirements of traffic, denoted by, d_k defined by the incoming load between nodes s_k and t_k

Moreover $A \in R^{N \times L}$, where N = | N | and L = | L |, is determined by the incidence matrix of the host connection, for which
Anl = -1, l if the channel is included in the node n,
Anl = 1 if channel l leaves from node n,
Anl = 0, in other cases;
, Refers to a vector load channels with the elements;

$R^k \in R^{N \times 1}, k \in K$, determined by the vector for which

$R^k_{s_k} = d_k$, $R^k_{t_k} = -d_k$ and $R^k_n = 0$ otherwise.
.

The intensity of traffic, released IE-pair k by the relation l, is defined as x_l^k. This intensity is the control variable problem of static load balancing. For a given x_l^k, induced load on the channel l, in the form:

$$y_l = \sum_{k \in K} x_l^k \text{, for all } l \in L. \qquad (1)$$

It is easy to see that, generally speaking, the mapping between the requirements of traffic d_k and download channel

y_l not a "one to one", thus while d_k uniquely identifies the

y_l, the reverse is not true. Download y_l channel l should be of the value of the channel $C_l(y_l)$, which is permissible increases and y_l convex function of load

convex function of load [13] the problem of uniform traffic distribution is solved by minimizing the cost function by optimizing the distribution of traffic

$$x^* = (x_l^k; k \in K; l \in L) \qquad (2)$$

and is represented as follows: Minimize

$$C(x) = \sum_{l \in L} C_l(y_l), \qquad (3)$$

subject to the restrictions

$x_l^k \geq 0$, for each $l \in L$ and $k \in K$,

$\sum_{k \in K} x_l^k \leq b_l$, for each $l \in L$,

$Ax^k = R^k$, for each $k \in K$,

Where $y_l = \sum_{k \in K} x_l^k$, for all $l \in L$, download the channel *l*, x_l^k The intensity of traffic on the connection k

In (3) placed three restrictions, the first indication that the value of load channels must be positive, the second - bandwidth limiting, and the third - the so-called flow conservation constraint, which indicates that the incoming traffic to the site each IE-pair must match traffic coming out of the site.

Download y_l channel should be of the value of $C_l(y_l)$, which allowed increases and y_l is a convex function of the load. B result, the alignment is the intensity of traffic x_l^k and traffic requirements d_k, However, routes for traffic with such requirements is not necessarily unique, but they are guaranteed not to have loops.

In [14] discussed ways of solving optimization problems based on different objective functions. It is assumed that the objective function is convex and increases depending on the output parameters. In this paper the simple optimization problem, and the nominal value of the problem of minimizing the flow. It is believed that each connection has a certain value. Then the cost function channels can be represented as:

$$C_l(y_l) = a_l y_l \text{ for all } l \in L. \qquad (4)$$

In this case, if the cost of communication adopted for all l, then the optimization problem minimizes the amount of resources used. However, when there is only one route, then the problem becomes analogous to the so-called problem of finding the shortest path. To describe the value of the channel is often used non-linear function of the form:

$$C_l = \begin{cases} \dfrac{y_l}{b_l - y_l}, & y_l < b_l, \\ \infty, & y_l \geq b_l. \end{cases} \qquad (5)$$

In [15], optimization of the nonlinear problem is considered to be a very time consuming. It is therefore proposed to reduce the optimization problem to the optimization of piecewise linear functions, which can be optimally solved in polynomial time. Minimize network latency expends more resources than the linear problem. In some cases, it may cause unwanted results in the evaluation of network performance.

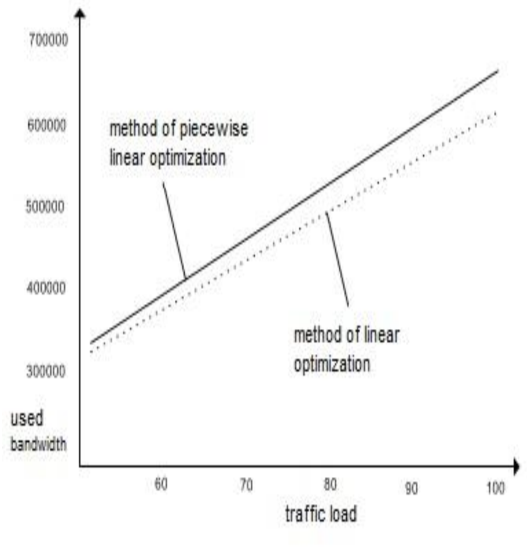

Fig. 1. The use of bandwidth
When traffic is sent on a linear optimization

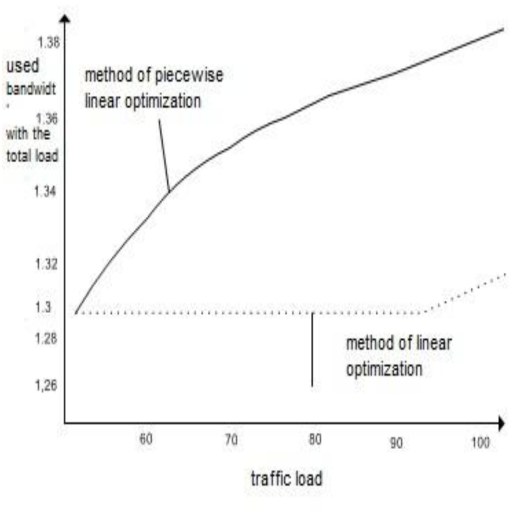

Fig. 2. The used bandwidth,
divided by the total load of overlooking,
as a function of traffic load.

For example, compare Figure 1 and Figure 2, the utilization of bandwidth when the traffic is transmitted over a linear optimization (objective function (4), Figure 1) with the value equal to one channel and minimizing network latency (the objective function (5), Figure 2 .) For linear optimization chosen the shortest path, if enough bandwidth. When the traffic load exceeds 90%, the shortest path are overloaded, and used long routes. When the average delay is minimized, except for short routes and long routes are used for load balancing, which can be seen with high bandwidth utilization. For example, selected a network of 10 nodes and 52 channels and 72 IE-pairs.

It is also well-known algorithms for solving the problem of load distribution algorithms are steepest descent, conjugate gradient, gradient projection method.
In [16] described an algorithm for the problem of the gradient projection of load distribution channel. To ensure the required level of QoS solution to the problem must be reduced to that of the gradient projection algorithm B This algorithm calculates the cost of C_p for each route $p \in \mathrm{P}$, which is the sum of the values of channels around the route :

$$C_p = \sum_{l \in p} C_l(y_l)\,. \qquad (6)$$

An arbitrary chosen initial length of this route

$$C_p^{'} = \sum_{l \in p} C_l^{'}(y_l)\,, \qquad (7)$$

Where $C_l^{'}(y_l)$ an initial arbitrary value of the cost function $C_l(y_l)$ and y_l download traffic channel .

Let $\overline{p}_k$ denotes the path *k*, which has the smallest initial length.
In algorithm iteratively changes the traffic load on the gradient projection algorithm, the objective function. Consequently, at each iteration *t + 1* , distribution of traffic on route , will be updated as follows:

$$x_p(t+1) = \max\left[x_p(t) + \alpha_H H_p^{-1}(C_{\overline{p}_k}^{'} - C_p^{'}),\ 0\right] \quad \forall p \in \mathrm{P}_k,\ p \neq \overline{p}_k$$

, (8)
where H is an approximation of the Hessian matrix,

$$H_p = \sum_{l \in p \cup \overline{p}_k,\, l \notin p \cap \overline{p}_k} C_l^{''}(y_l)\,, \qquad (9)$$

where the scale factor, a $C_l^{''}(y_l)$ second arbitrary value of the cost function $C_l(y_l)$ with respect to load the channel.

In order to prove that the proposed solution is acceptable, ie, amount of routing traffic load is equal to the requirements of traffic d_k, additional traffic for each flow is distributed along the route with the lowest initial length, i

$$x_{\overline{p}_k}(t+1) = d_k - \sum_{p \in P_k,\, p \neq \overline{p}_k} x_p(t+1)\,. \qquad (10)$$

As a result, the algorithm can be written as follows:

1. by (6) the cost is calculated C_p for each route $p \in \mathrm{P}$, which is the sum of the values of channels along the route
2. chosen $C_l^{'}(y_l)$ an initial arbitrary value of the cost of loading $C_l^{'}(y_l)$ and y_l induced traffic channel l
3. by (7) is calculated initial arbitrary length of this route.
4. by (9) is calculated approximation of the Hessian matrix
5. Set the scale factor
6. select $C_l^{''}(y_l)$ second arbitrary value of the cost function $C_l(y_l)$ relative channel loading y_l .

7. perform the required number of iterations t +1, each of which is the distribution of traffic on route p, will be updated (8).
8. In order to prove that the solution is admissible, ie amount of routing traffic load is equal to the requirements of traffic d_k, additional traffic for each flow is distributed along the route with the lowest initial length (10).

4. SIMULATION OF TIME COMPLEXITY

Algorithm of the gradient projection

The total time complexity of the gradient projection algorithm can be obtained as the iteration complexity and time complexity at each iteration. To obtain the results of simulations are used:

n the number of nodes

$T_I(n)$ number of iterations

F(n) estimator $h_{min} N_{max} / \varepsilon^2$
R ratio $T_I(n)/F(n)$

Table 2 presents the numerical values needed for plotting

Table 2. T1 (n) compared to F (n)

n	*Tf(n)*	F(n)	*R*
169	36	9.21E4-7	3.91E-07
289	66	4.70E+8	1.41E-07
256	96	8.22E+8	1.17E-07
324	83	8.70E+8	0.95E-07
361	69	1.13E+9	0.61E-07
225	78	1.75E+9	0.46B-07
441	83	1.82E+9	0.46&-07
400	121	2.71E+9	0.45E-07

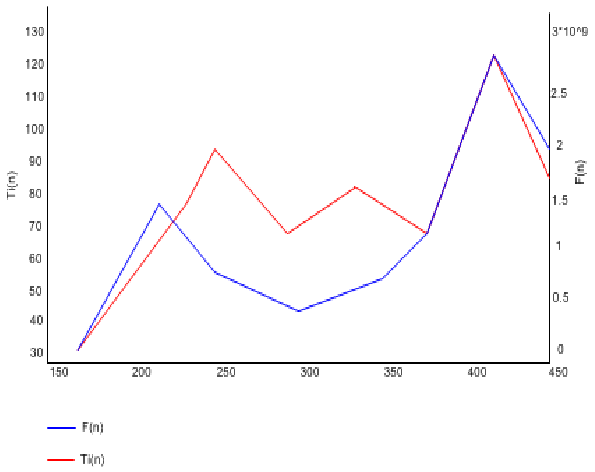

Fig. 3. T1 (n) compared to F (n)

5. Conclusions

From Table 2 and Figure 3 shows that the ratio remains relatively unchanged. This result shows that the theoretically derived limit of TI (n), determines the number of iterations when the number of nodes in the network. Also, the trend of development is predictable, and most are expected to analyze the asymptotic complexity.

The advantages of the gradient projection algorithm is compared with other algorithms [17]:

1. Scale the diagonal of the second derivative;
2. The best algorithm for the optimal quasi-static routing;
3. The deterministic synchronous environment, it converges faster than the shortest path method (used in many networks), and also faster (in terms of computational cost) than the Newton method of planning when starting a far from optimal routing.

-The problem of constructing the traffic and formulate the classical problem of load balancing and optimal routing.

-The algorithm of the gradient projection, and its applicability to the optimization of load in the mobile network.

-The problem of minimizing the cost of flow.

6. References

[1] F. Yu, V. Wong, V. Leung, A New QoS Provisioning Method for Adaptive Multimedia in Cellular Wireless networks, INFOCOM 2004.

[2] Y. Xiao, H. Li, S. Choi, Protection ND Guarantee for Voice and Video Traffic in IEEE 802.11e Wireless LANs, INFOCOM 2004.

[3] W. Sheikh, B. Shafig, R. Paul, A. Ghafoor,Provision of Multimedia Service in a Mobile Ad Hoc Network, Computer Society, IEEE 2004.

[4] G. Yang, D. Shen, V. Li, UEP for Video Transmission in Space-Time Coded OFDM Systems, INFOCOM 2004.

[5] H. Zlatokrilov, H.Levy, Packet Dispersion and Quality of Voice over IP Applications in IP networks, INFOCOM 2004.

[6] A. Kumar, T. Malhotra, A Multi-Signaling Protocol Architecture for Voice over IP Terminal, INFOCOM 2004.

[7] S. Das, E.Lee, K. Basu, N. Kakani, S. Sen, Performance Optimization of VoIP Calls over Wireless Links Using H.323 Protocol, INFOCOM 2002.

[8] M. Mowbray, G. Karlsson, T. Kohler,Capacity reservation for multimedia traffic, Swedish Institute of Computer Science, 1997.

[9] X. Ding, K. Roy, A Novel Bitstream Level Joint Channel Error Concealment Scheme for Realtime Video over Wireless Networks, INFOCOM 2004.

[10] M. Chen, A. Zakhor, Rate Control for Streaming Video over Wireless, INFOCOM 2004.

[11] Таненбаум Э., Компьютерные сети, 4-е издание, Питер 2003.-992с.

[12] D. Awduche, A. Chiu, A. Elwalid, I. Widjaja, X. Xiao, Overview and Principles of Internet Traffic Engineering, May 2002, IETF RFC 3272

http://library.pantek.com/Miscellaneous/RFC%20Archives/rfc3272.html

[13] D. Bertsekas, R. Gallager, Data Networks, Prentice Hall, Englewood Cliffs., N.J., 2nd edition, 1992.

[14] D. Bertsekas, Network Optimization: Continous and Discrete Models, Athena Scientific, 1998.

[15] B. Fortz, M.Thorup, Internet Traffic Engineering by Optimizing OSPF Weights, in Proceeding of IEEE INFOCOM, March 2000.

[16] J. Tsitsiklis, D. Bertsekas, Distributed Asynchronous Optimal Routing in Data Networks, IEEE Transactions on Automatic Control, Vol. AC-31, No.4, April 1986.

[17] D. Bertsekas, Network Optimization: Continous and Discrete Models, Athena Scientific, 1998.