Optimization of Hierarchical System for Data Acquisition

Vít NOVOTNÝ

Dept. of Telecommunications, FEEC, Brno University of Technology, Brno, Czech Republic

novotnyv@feec.vutbr.cz

Abstract. *Television broadcasting over IP networks (IP-TV) is one of a number of network applications that are except of media distribution also interested in data acquisition from group of information resources of variable size. IP-TV uses Real-time Transport Protocol (RTP) protocol for media streaming and RTP Control Protocol (RTCP) protocol for session quality feedback. Other applications, for example sensor networks, have data acquisition as the main task. Current solutions have mostly problem with scalability - how to collect and process information from large amount of end nodes quickly and effectively? The article deals with optimization of hierarchical system of data acquisition. Problem is mathematically described, delay minima are searched and results are proved by simulations.*

Keywords

Data acquisition, multicast, SSM, multimedia, RTP/ RTCP, sensor networks.

1. Introduction

Many network applications use plain centralized model with one data centre where all pieces of information from end nodes are directly sent, collected, then processed and available for later evaluation. There is no problem with data acquisition and with data processing provided the number of data sources is fairly low and data flows are weak and low frequent. When these conditions are not fulfilled either the center itself or data links to the center can be overloaded or allowed data transmission frequency is very low. When the data acquisition is auxiliary procedure of the service, the available bandwidth for such procedure is strictly limited and the situation becomes even worse. This is the case of applications like IP-TV where the main task of the service is the multimedia streaming using RTP protocol and the multicast transmission and the session quality parameter collection using RTCP protocol is an optional though useful supplementary service [1], [2], [3]. The transmission capacity of RTCP is limited for 5% of total service bandwidth and it causes large delays in sending RTCP (feedback) data from each receiver for large-scale media streaming services based on Source-Specific Multicast (SSM), [8]. Similar problem arises also with other applications focused on data acquisition in the case of large-scale systems.

There are also other models of data collection and processing in the data network environment. In addition to the centralized model mentioned above, there are hierarchical and distributed models available. The hierarchical system is created by a tree of servers with local data and using links among them the requested information can be found and transported through the tree hierarchy. The third model is distributed model where the pieces of information are distributed among equivalent data centers and using one sophisticated directory services system with one account requested information can be obtained.

2. Hierarchical Data Acquisition System

2.1 Tree Architecture Design for Data Acquisition

To combat the problem with scalability the hierarchical system for data acquisition was proposed in [6], [7] and modified in [8]. In addition to the data center and data sources such tree contains special nodes called summarization nodes, see Fig. 1.

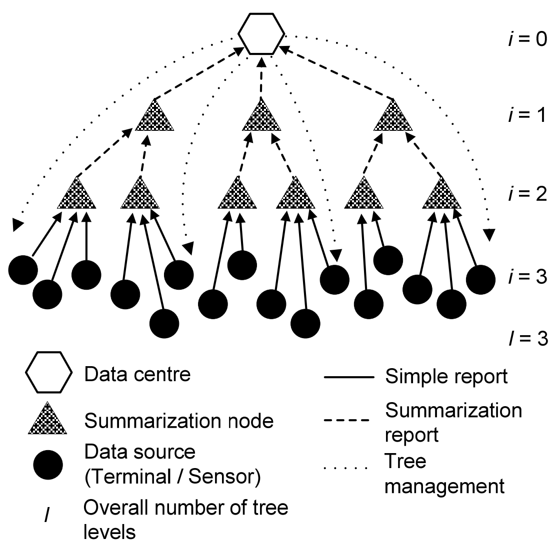

Fig. 1. Tree design for large-scale data acquisition.

Pieces of data are periodically sent from data sources (terminals or sensors) to an assigned summarization node. The summarization node aggregates data from a group of terminals of the size n_B and again periodically sends to an assigned summarization node at the higher level. The summarization nodes are organized into groups of size n_S.

The data aggregation depends on the kind of network application. This change from flat to hierarchical system requires a number of additions and brings other problems that should be addressed.

First it must be specified which kind of information must be transported from end nodes without aggregation and which data can be processed in summarization nodes and transported in the form of histograms.

Optionally in addition to the summarization process (in such a case the detailed information about particular end node is lost) the summarization nodes can store detailed information obtained from terminals or lower level summarization nodes for some time period to allow the data center to get detailed information about particular terminal or group of terminals when necessary.

The other problem is how to manage the tree when the number of terminals rises or declines and how to keep it in a balanced form. In addition to this the problem of limited number of required summarization nodes should be addressed. At the beginning of our research we considered that the summarization nodes are only terminals with special functionality [8]. It was found that there would be lot of overhead with the management of such tree especially when the tree is variable in a large extent, i.e. the terminals will enter and leave the session frequently; this is the case of multimedia streaming sessions. Also this functionality would require additional power and energy from the terminals and that is unwanted issue especially in the case of wireless terminals (sensors) with limited computational power and energy. Therefore in later research ([9], [10], [11], [12]) the summarization nodes are considered as special nodes (or software modules) that are managed by the service provider. Such summarization nodes have higher computational power, larger storage capacity for temporary data, fixed location and always available. The last but one feature is very important when the tree structure is established according to the location of terminals.

Tree establishment according to the terminal location is very useful because it optimizes link lengths between connected nodes in the tree hierarchy, so that it speeds up message delivery, saves the network resources and overall energy consumption (mainly important in wireless networks). This task is quite difficult to solve due to Internet complexity and due to variability of transport conditions. Several methods how to find the terminal location in the Internet have been studied, new protocol TTP (Transmission Tree Protocol) and results have been published, [10], [11] and [12].

As the large number of terminals (end nodes) is divided into many smaller groups the bandwidth restriction is not the problem and the message transmission period of the terminals remains fairly low even if the overall number of terminals rises. Especially this is the case of multimedia multicast sessions which can vary substantially in size from several hundreds up to millions. The overall delay between the time instant when data is generated (or measured) in the terminal (sensor) and the time instant when the data is received in the data center consists of particular transmission delays between transmission instants of adjacent layers in the tree.

When the tree consists of I layers, i.e. (I-1) summarization layers and one terminal layer, a formula for the overall delay T_{RT} between data generation (measurement) and its reception in the data processing center can be derived:

$$T_{RT} = \tau_{MT} + \sum_{i=1}^{I-1} \tau_i , \tag{1}$$

provided the transport delay through the network is neglected. Variable τ_{MT} is the delay between measurement (data generation) and transmission instants and τ_i is the delay between summarized message transmission instants at linked summarization nodes in adjacent layers. The worst case for the delay will be when all summarization nodes at all levels of the tree and also the terminals (sensors) are synchronized, i.e. all of them transmit messages at the same time instants. Provided that the transmission periods will be the same through whole the tree the formula (1) changes into the form

$$T_{RTW} = T_{RR} + T_{\Sigma RT} = T_{RR} + \left(I-1\right) T_{\Sigma R} \tag{2}$$

where T_{RR} is the transmission period of the group of terminals (it depends on the number n_B of terminals in the group, message length and the allocated bandwidth, [8]), $T_{\Sigma RT}$ is the maximum overall delay through all layers of summarization nodes, I is the number of levels in the tree (it depends on the total number of terminals, on the number of terminals in the group n_B and on the number of summarization nodes in the group n_S) and $T_{\Sigma R}$ is the message transmission period of a summarization node group (it depends on the number n_S of summarization nodes in the group, summarization message length and the allocated bandwidth, [8]).

3. Tree Optimization

When a service provider intents to implement a service based on the tree architecture described above in real situation, some initial conditions have to be considered before tree architecture implementation: bandwidth BW_A (or maximum data flow) allocated for the data acquisition (it will be allocated both for a group of terminals and summarization nodes; bandwidth is expected to be the same for both groups in most cases, but generally it can be different, i.e. BW_{AR} and $BW_{A\Sigma}$), expected number of data sources (terminals) n_T, maximum acceptable period (or delay) of data collection T_{Rmax}, length PL_{RR} of plain messages generated by the terminals, length of summarization packets length $PL_{\Sigma\mathrm{R}}$ generated by summarization nodes, minimum periods of message transmission in a group of terminals T_{RRmin} ($T_{\mathrm{RR}} \geq T_{\mathrm{RRmin}}$) and in a group of summarization nodes $T_{\Sigma\mathrm{Rmin}}$ ($T_{\Sigma\mathrm{R}} \geq T_{\Sigma\mathrm{Rmin}}$). Additional constraint can be also maximum overall number of summarization nodes N_{STmax} that are available. The goal is to find such a tree which meets all of these conditions and restrictions, possibly with minimum costs.

Equation (2) shows how to calculate the longest overall delay T_{RTW} (and also the maximum time period of data acquisition) between data generation (measurement) in terminals (data sources) and its reception in the data processing center. It can be worked out in more detailed form:

$$\begin{aligned} T_{\mathrm{RTW}} &= T_{\mathrm{RR}} + (I-1)T_{\Sigma\mathrm{R}} = \tau_{\mathrm{R}} n_{\mathrm{B}} + (I-1)\tau_{\Sigma} n_{\mathrm{S}} = \\ &= (BW_{\mathrm{A}})^{-1}\left[PL_{\mathrm{RR}} n_{\mathrm{B}} + (I-1) PL_{\Sigma\mathrm{R}} n_{\mathrm{S}}\right], \end{aligned} \tag{3}$$

where τ_{R} is the portion of period for a message sent by one terminal, τ_{Σ} is the portion of period for a message sent by one summarization node, n_{B} is the number of terminals in one group of terminals, n_{S} is the number of nodes in the group of summarization nodes, I is the number of levels in the tree, PL_{RR} is the message length generated by terminals and $PL_{\Sigma\mathrm{R}}$ is the message length generated by summarization nodes.

The number of levels with summarization nodes in the tree, i.e. the value (I-1), can be calculated from the condition

$$n_{\mathrm{S}}^{(I\text{-}2)} < \frac{n_{\mathrm{T}}}{n_{\mathrm{B}}} \leq n_{\mathrm{S}}^{(I\text{-}1)} \,. \tag{4}$$

Then

$$(I\text{-}1) \geq \log_{n_{\mathrm{S}}}\left(\frac{n_{\mathrm{T}}}{n_{\mathrm{B}}}\right). \tag{5}$$

As I is an integer number the nearest higher integer will be

$$(I\text{-}1) = \log_{n_{\mathrm{S}}}\left(\frac{n_{\mathrm{T}}}{n_{\mathrm{B}}}\eta_{\mathrm{I}}\right); \quad \eta_{\mathrm{I}} \in \langle 1, n_{\mathrm{S}}) \,. \tag{6}$$

Then (3) changes into form

$$\begin{aligned} T_{\mathrm{RTW}} &= \tau_{\mathrm{R}} n_{\mathrm{B}} + (I-1)\tau_{\Sigma} n_{\mathrm{S}} = \\ &= \tau_{\mathrm{R}} n_{\mathrm{B}} + (\tau_{\Sigma} n_{\mathrm{S}})\left[\log_{n_{\mathrm{S}}}\left(\frac{n_{\mathrm{T}}}{n_{\mathrm{B}}}\eta_{\mathrm{I}}\right)\right]. \end{aligned} \tag{7}$$

In the case of particular service the parameters τ_{R} and τ_{Σ} are specified and fixed. When the total number of terminals (end nodes) n_{T} is known it can be seen from (7) that n_{B} and n_{S} are the key parameters in message transmission delay optimization process. When the constraints T_{RRmin} and $T_{\Sigma\mathrm{Rmin}}$ are considered, we can calculate limits n_{Bmin} and n_{Smin} from the following formulae

$$\begin{aligned} n_{\mathrm{Bmin}} &= \frac{BW_{\mathrm{A}}}{PL_{\mathrm{RR}}} T_{\mathrm{RRmin}} + \eta_{\mathrm{Bmin}} = \frac{T_{\mathrm{RRmin}}}{\tau_{\mathrm{R}}} + \eta_{\mathrm{Bmin}}, \\ &\eta_{\mathrm{Bmin}} \in \langle 0, 1), \\ n_{\mathrm{Smin}} &= \frac{BW_{\mathrm{A}}}{PL_{\Sigma\mathrm{R}}} T_{\Sigma\mathrm{Rmin}} + \eta_{\mathrm{Smin}} = \frac{T_{\Sigma\mathrm{Rmin}}}{\tau_{\Sigma}} + \eta_{\mathrm{Smin}}, \\ &\eta_{\mathrm{Smin}} \in \langle 0, 1). \end{aligned} \tag{8}$$

Parameters η_{Bmin} and η_{Smin} round the quantities n_{Bmin} and n_{Smin} to the nearest higher integer to fulfil at least the minimum transmission periods T_{RRmin} and $T_{\Sigma\mathrm{Rmin}}$. When the constraints T_{RRmin} and $T_{\Sigma\mathrm{Rmin}}$ are not specified n_{Bmin} and n_{Smin} are set to 1 and minimum transmission periods are set to τ_{R} and τ_{Σ}.

To ensure proper message transmission each node within the tree hierarchy has to know:

- quantities to be measured / summarized and sent,
- message structure and length,
- link to the node at higher level,
- current number of nodes in its group,
- maximum bandwidth allocated to the service,
- minimum transmission period.

The parameter limits n_{Bmin} and n_{Smin} divide the plane [n_{B}, n_{S}], i.e. definition domain of (7), into four regions as shown in Fig. 2.

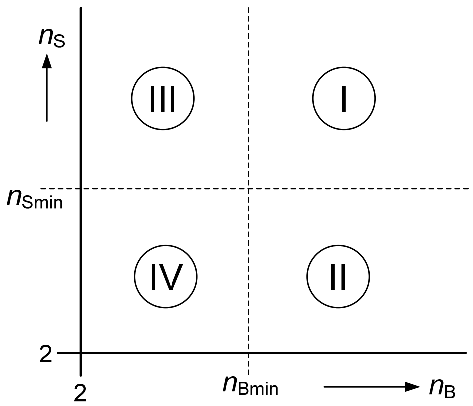

Fig. 2. Regions for tree feedback report delay optimization.

Region I): $\quad n_{\mathrm{B}} \geq n_{\mathrm{Bmin}} \wedge n_{\mathrm{S}} \geq n_{\mathrm{Smin}}$ (9)

Region II): $\quad n_{\mathrm{B}} \geq n_{\mathrm{Bmin}} \wedge n_{\mathrm{S}} < n_{\mathrm{Smin}}$ (10)

Region III): $\quad n_{\mathrm{B}} < n_{\mathrm{Bmin}} \wedge n_{\mathrm{S}} \geq n_{\mathrm{Smin}}$ (11)

Region IV): $\quad n_{\mathrm{B}} < n_{\mathrm{Bmin}} \wedge n_{\mathrm{S}} < n_{\mathrm{Smin}}$ (12)

Regions II, III, IV exist provided the values of n_{Bmin} and / or n_{Smin} are three and larger. Formula (7) will change for each region:

Region I): $$T_{RTW} = \tau_R n_B + (\tau_\Sigma n_S)\left[\log_{n_S}\left(\frac{n_T}{n_B}\eta_I\right)\right], \quad (13)$$

Region II): $$T_{RTW} = \tau_R n_B + T_{\Sigma R\min}\left[\log_{n_S}\left(\frac{n_T}{n_B}\eta_I\right)\right], \quad (14)$$

Region III) $$T_{RTW} = T_{RR\min} + (\tau_\Sigma n_S)\left[\log_{n_S}\left(\frac{n_T}{n_B}\eta_I\right)\right], \quad (15)$$

Region IV) $$T_{RTW} = T_{RR\min} + T_{\Sigma R\min}\left[\log_{n_S}\left(\frac{n_T}{n_B}\eta_I\right)\right]. \quad (16)$$

Each region should be inspected to find the best tree configuration from the total message transport delay point of view.

The functions (13), (14), (15) and (16) are discontinuous due to the integer value of expression in square brackets of (13), (14), (15) and (16), as it can be seen from Fig. 3 (delay was calculated for the example from [8] where $n_T = 10^5$, $PL_{RR} = 736$ bits, $PL_{\Sigma R} = 11296$ bits and $BW_{AR} = BW_{A\Sigma} = BW_A = 37.5$ kbps).

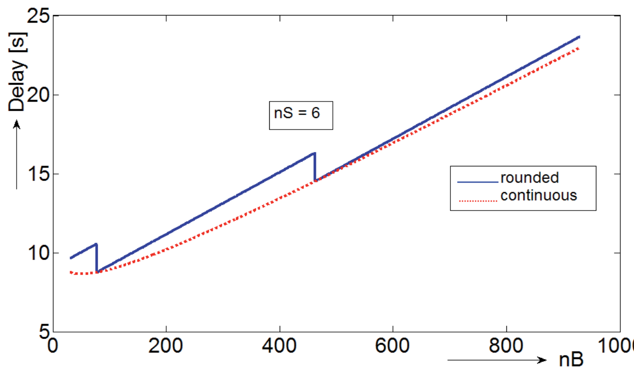

Fig. 3. Course of the worst-case total delay according to (13).

When this discontinuous function (13) is replaced by continuous one (without correction parameter η_I),

$$T_{RTW} = \tau_R n_B + (\tau_\Sigma n_S)\left[\log_{n_S}\left(\frac{n_T}{n_B}\right)\right], \quad (17)$$

the worst-case total delay values obtained from the optimization process with continuous function are quite close to and always better than when discontinuous function is considered (due to the fact that $\eta_I \geq 1$), see Fig. 3. This was also proved by modeling in Matlab, see Fig. 4.

It can be seen that in the case of discontinuous characteristics the course of the function is complex and it is very difficult to find the minimum. Therefore it was decided to realize the optimization process with continuous function according to (17) and after the minimum of continuous function is found, the process of minimum searching process in the discontinuous function is limited to the restricted area.

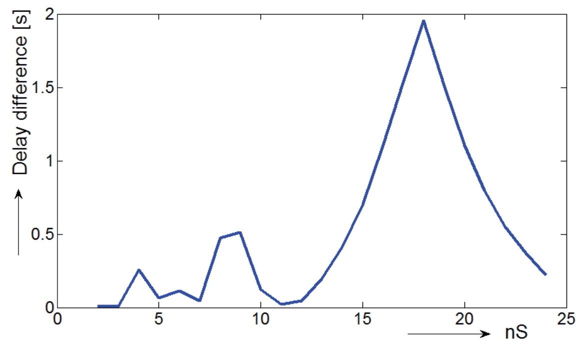

Fig. 4. Differences of minimum worst-case total delay values between discontinuous (13) and continuous functions as the function of n_S.

3.1 Region I Optimization

The continuous form of the total worst-case delay function in region I is described by (17). The goal of optimization is to find its global extreme (minimum) in this region. Global extreme can be located either in local extremes of the function or at the boundary of definition domain. The function is continuous in the whole region and smooth, therefore the first and also second derivatives can be calculated and stationary points of the function can be found:

$$\frac{\partial T_{RTW}}{\partial n_B} = \tau_R - \tau_\Sigma \frac{n_S}{n_B \ln n_S}, \quad (18)$$

$$\frac{\partial T_{RTW}}{\partial n_S} = \tau_\Sigma \frac{\ln\left(\frac{n_T}{n_B}\right)}{\ln^2 n_S}(\ln n_S - 1). \quad (19)$$

Stationary points are the candidates for local extremes and they can be calculated from the conditions that the first derivatives (18) and (19) are put equal to zero and the results are:

$$n_{Ss1} = \mathrm{e}\ (\text{i.e. } 2.71828...) + \eta_S = 3 \quad (20)$$

and when non-rounded n_{Ss1} is used for n_{Bs1} calculations

$$n_{Bs1} = \frac{\tau_\Sigma}{\tau_R}\frac{n_{Ss1}}{\ln n_{Ss1}} = \frac{\tau_\Sigma}{\tau_R}\mathrm{e} + \eta_B = \frac{PL_{\Sigma R}}{PL_{RR}}\mathrm{e} + \eta_B, \quad (21)$$
$$\eta_B \in \langle -0.5; +0.5 \rangle.$$

Unfortunately the result for optimum n_S is very small. To meet the restrictions for region I specified by (9) and to include the stationary point the conditions are:

$$n_{S\min} \leq 3, \qquad n_{B\min} \leq \frac{\tau_\Sigma}{\tau_R}\mathrm{e} + \eta_B. \quad (22)$$

In majority cases $\tau_\Sigma >> \tau_R$ like in the example in [8] where packet lengths of the receiver report and summarization report were $PL_{RR} = 736$ bits and $PL_{\Sigma R} = 11296$ bits. Then n_{Bs} will be

$$n_{Bs1} = \frac{\tau_\Sigma}{\tau_R} e + \eta_B = \frac{PL_{\Sigma R}}{PL_{RR}} e + \eta_B = \frac{11296}{736} e + \eta_B = 42, \quad (23)$$

and therefore $$n_{Bmin} \le 42 \quad (24)$$

as the restriction (9) specifies. When we select $n_{Smin} = 2$ and consider a reasonable requirement that $T_{RRmin} = T_{\Sigma Rmin}$ then according to (8) the parameters n_{Smin} and n_{Bmin} are related by the formula

$$\begin{aligned} n_{Bmin} &= \frac{\tau_\Sigma}{\tau_R} n_{Smin} + \eta_{Bmin} = \frac{PL_{\Sigma R}}{PL_{RR}} n_{Smin} + \eta_{Bmin} = \\ &= \frac{11296}{736} 2 + \eta_{Bmin} = 31. \end{aligned} \quad (25)$$

This shows that requirement (24) is met. In such a case and when the example from [8] is considered where PL_{RR} = 736 bits, $PL_{\Sigma R}$ = 11296 bits and BW_A = 37.5 kbps we obtain the following result for parameters T_{RRmin} and $T_{\Sigma Rmin}$:

$$\begin{aligned} T_{RRmin} &= T_{\Sigma Rmin} = n_{Bmin}\tau_R = n_{Smin}\tau_\Sigma = n_{Bmin} \frac{PL_{RR}}{BW_A} = \\ &= n_{Smin} \frac{PL_{\Sigma R}}{BW_A} = 2 \frac{11296}{37500}\ \mathrm{s} = 602.45\ \mathrm{ms}. \end{aligned} \quad (26)$$

Such situation is only partially acceptable, because of quite short message transmission period and of large number of required summarization nodes in the case of large systems.

When $n_{Smin} = 3$ or more precisely, when $T_{RRmin} > e\tau_\Sigma$ then $n_{Bmin} > 42$ and the condition (24) is not fulfilled.

To prove, whether the local minimum was eventually found, it is necessary to check sufficient conditions for the existence of local minimum

$$\begin{aligned} D_1 &= \left. \frac{\partial^2 T_{RTW}}{\partial n_B^2} \right|_{\substack{n_B = n_{Bs} \\ n_S = n_{Ss}}} \ge 0, \\ D_2 &= \left| \begin{matrix} \frac{\partial^2 T_{RTW}}{\partial n_B^2} & \frac{\partial^2 T_{RTW}}{\partial n_B \partial n_S} \\ \frac{\partial^2 T_{RTW}}{\partial n_B \partial n_S} & \frac{\partial^2 T_{RTW}}{\partial n_S^2} \end{matrix} \right|_{\substack{n_B = n_{Bs} \\ n_S = n_{Ss}}} > 0. \end{aligned} \quad (27)$$

and to calculate the second partial derivatives:

$$\frac{\partial^2 T_{RTW}}{\partial n_B^2} = \tau_\Sigma \frac{n_S}{n_B^2 \ln n_S}, \quad (28)$$

$$\frac{\partial^2 T_{RTW}}{\partial n_S^2} = \tau_\Sigma \ln\left(\frac{n_T}{n_B}\right) \frac{2 - \ln n_S}{n_S \ln^3 n_S}, \quad (29)$$

$$\frac{\partial^2 T_{RTW}}{\partial n_B \partial n_S} = \frac{\partial^2 T_{RTW}}{\partial n_S \partial n_B} = -\frac{\tau_\Sigma}{n_B} \frac{\ln n_S - 1}{\ln^2 n_S}. \quad (30)$$

When the results (20) and (21) are used in (27) we get:

$$\begin{aligned} D_1 &= \frac{\tau_R^2}{\tau_\Sigma e}, \\ D_2 &= \frac{\tau_R^2}{e^2} \left[\ln\left(n_T \frac{\tau_R}{\tau_\Sigma} \right) - 1 \right]. \end{aligned} \quad (31)$$

The condition $D_1 > 0$ is always met and the condition D_2 will be fulfilled when

$$n_T \frac{\tau_R}{\tau_\Sigma} > e. \quad (32)$$

Again in the example of RTCP presented in [8] the length of receiver report PL_{RR} was 736 bits and the length of summarization report $PL_{\Sigma R}$ was 11296 bits. In the case when the same link bandwidths are assigned both to terminals and summarization nodes (32) has the form

$$\begin{aligned} & n_T \frac{\tau_R}{\tau_\Sigma} = n_T \frac{PL_{RR}}{PL_{\Sigma R}} = n_T \frac{736}{11296} \approx 0.065 n_T > e, \\ & n_T > 42. \end{aligned} \quad (33)$$

This condition is quite easy to meet.

Provided the conditions (22) and (32) are met (20) and (21) specifies the local minimum of the region I whose value is

$$\begin{aligned} & T_{RTW1}\left(n_{Bs1} = \frac{\tau_\Sigma}{\tau_R} e + \eta_B, n_{Ss1} = 3 \right) = \\ & = \tau_\Sigma \left[e + 3 \log_3 \left(\frac{n_T}{\frac{\tau_\Sigma}{\tau_R} e + \eta_B} \right) \right] + \tau_R \eta_B. \end{aligned} \quad (34)$$

When parameter $n_{Smin} \ge 3$ which will be obvious case, the course of minima changes into the trajectory shown in Fig. 5 and therefore the absolute minimum of the region I will be reached for the smallest n_S, i.e. n_{Smin} and for n_{Bmin}. Hence the point (n_{Bmin}, n_{Smin}) is the best choice.

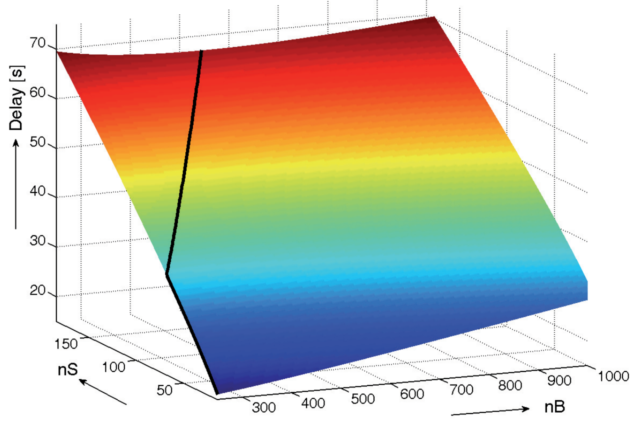

Fig. 5. Total delay versus n_B and n_S according to (13) with minima localization for $n_{Smin} > 3$.

As the global extreme is searched, it is necessary to inspect also boundaries of the region I that are specified by expressions

$$n_{\mathrm{B}} \geq n_{\mathrm{Bmin}} \wedge n_{\mathrm{S}} = n_{\mathrm{Smin}} \tag{35}$$

and

$$n_{\mathrm{B}} = n_{\mathrm{Bmin}} \wedge n_{\mathrm{S}} \geq n_{\mathrm{Smin}} . \tag{36}$$

Condition (35) changes formula (13) to the form

$$T_{\mathrm{RTW}} = \tau_{\mathrm{R}} n_{\mathrm{B}} + T_{\Sigma\mathrm{Rmin}} \left[\log_{n_{\mathrm{Smin}}} \left(\frac{n_{\mathrm{T}}}{n_{\mathrm{B}}} \right) \right]. \tag{37}$$

Its derivative has the form:

$$\frac{\partial T_{\mathrm{RTW}}}{\partial n_{\mathrm{B}}} = \tau_{\mathrm{R}} - \frac{T_{\Sigma\mathrm{Rmin}}}{n_{\mathrm{B}} \ln n_{\mathrm{Smin}}} \tag{38}$$

and the stationary point has the coordinates:

$$\begin{aligned} n_{\mathrm{Bs2}} &= \frac{T_{\Sigma\mathrm{Rmin}}}{\tau_{\mathrm{R}} \ln n_{\mathrm{Smin}}} + \eta_{\mathrm{B}} = \frac{\tau_{\Sigma} n_{\mathrm{Smin}}}{\tau_{\mathrm{R}} \ln n_{\mathrm{Smin}}} + \eta_{\mathrm{B}} = \\ &= \frac{T_{\Sigma\mathrm{Rmin}}}{\tau_{\mathrm{R}} \left(\ln T_{\Sigma\mathrm{Rmin}} - \ln \tau_{\Sigma} \right)} + \eta_{\mathrm{B}}, \\ n_{\mathrm{Ss2}} &= n_{\mathrm{Smin}} = \frac{T_{\Sigma\mathrm{Rmin}}}{\tau_{\Sigma}} + \eta_{\mathrm{Smin}} . \end{aligned} \tag{39}$$

The second derivative

$$\frac{\partial^2 T_{\mathrm{RTW}}}{\partial n_{\mathrm{B}}^2} = \frac{T_{\Sigma\mathrm{Rmin}}}{n_{\mathrm{B}}^2 \ln n_{\mathrm{Smin}}} \tag{40}$$

is positive at the stationary point, i.e. the local minimum was found.

As the expression (35) specifies two restrictions, also the following condition has to be met:

$$\begin{aligned} & n_{\mathrm{Bs2}} \geq n_{\mathrm{Bmin}}, \\ \text{i.e. } & \frac{T_{\Sigma\mathrm{Rmin}}}{\tau_{\mathrm{R}} \left(\ln T_{\Sigma\mathrm{Rmin}} - \ln \tau_{\Sigma} \right)} \geq \frac{T_{\mathrm{RRmin}}}{\tau_{\mathrm{R}}}, \\ & \frac{T_{\Sigma\mathrm{Rmin}}}{T_{\mathrm{RRmin}}} \geq \ln n_{\mathrm{Smin}} . \end{aligned} \tag{41}$$

In the case when $T_{\mathrm{RRmin}} = T_{\Sigma\mathrm{Rmin}}$ this restriction results in the requirement

$$n_{\mathrm{Smin}} = 2 . \tag{42}$$

Then

$$\begin{aligned} & n_{\mathrm{Ss2}} = n_{\mathrm{Smin}} = 2, \\ & n_{\mathrm{Bs2}} = \frac{T_{\Sigma\mathrm{Rmin}}}{\tau_{\mathrm{R}} \ln n_{\mathrm{Smin}}} + \eta_{\mathrm{B}} = \frac{2\tau_{\Sigma}}{\tau_{\mathrm{R}} \ln 2} + \eta_{\mathrm{B}} = \frac{2}{\ln 2} \frac{PL_{\Sigma\mathrm{R}}}{PL_{\mathrm{RR}}} + \eta_{\mathrm{B}}, \\ & \eta_{\mathrm{B}} \in \left(-0.5; +0.5 \right\rangle, \end{aligned} \tag{43}$$

provided the same bandwidths are allocated both for group of terminals and group of summarization nodes and when expression (8) is used. Then the local minimum of the worst-case total delay T_{RTW} will be:

$$\begin{aligned} & T_{\mathrm{RTW2}} \left(n_{\mathrm{Bs2}} = \frac{2}{\ln 2} \frac{PL_{\Sigma\mathrm{R}}}{PL_{\mathrm{RR}}} + \eta_{\mathrm{B}}, n_{\mathrm{Ss2}} = 2 \right) = \\ & = \tau_{\mathrm{R}} \left\{ \frac{2}{\ln(2)} \frac{\tau_{\Sigma}}{\tau_{\mathrm{R}}} \left[1 + \ln \left(\frac{n_{\mathrm{T}}}{\frac{2}{\ln(2)} \frac{\tau_{\Sigma}}{\tau_{\mathrm{R}}} + \eta_{\mathrm{B}}} \right) \right] + \eta_{\mathrm{B}} \right\}. \end{aligned} \tag{44}$$

In the presented example when PL_{RR} = 736 bits and $PL_{\Sigma\mathrm{R}}$ = 11296 bits expressions in (43) result into the following point coordinates

$$\begin{aligned} & n_{\mathrm{Ss2}} = n_{\mathrm{Smin}} = 2, \\ & n_{\mathrm{Bs2}} = \frac{2}{\ln 2} \frac{PL_{\Sigma\mathrm{R}}}{PL_{\mathrm{RR}}} = \frac{2}{\ln 2} \frac{11296}{736} = 44, \\ & n_{\mathrm{Bmin}} \leq 44. \end{aligned} \tag{45}$$

The last condition of (45) is met according to (25). In other cases, i.e. when $n_{\mathrm{Smin}} > 2$, the point (n_{Bmin}, n_{Smin}) provides the minimum of delay.

The second part of the region I boundary is specified by (36) and

$$T_{\mathrm{RTW}} \left(n_{\mathrm{Bmin}} \right) = T_{\mathrm{RR\,min}} + \left(\tau_{\Sigma} n_{\mathrm{S}} \right) \left[\log_{n_{\mathrm{S}}} \left(\frac{n_{\mathrm{T}}}{n_{\mathrm{Bmin}}} \right) \right]. \tag{46}$$

Its derivative according to the parameter n_{S} is

$$\frac{\partial T_{\mathrm{RTW}} \left(n_{\mathrm{Bmin}} \right)}{\partial n_{\mathrm{S}}} = \tau_{\Sigma} \ln \left(\frac{n_{\mathrm{T}}}{n_{\mathrm{Bmin}}} \right) \frac{\ln n_{\mathrm{S}} - 1}{\ln^2 n_{\mathrm{S}}}, \tag{47}$$

which yields the same result like (20), i.e.

$$n_{\mathrm{Ss3}} = \mathrm{e}\ (2.71828...) + \eta_{\mathrm{S}} = 3 . \tag{48}$$

The second derivative at this point proves that the point (n_{Bmin}, n_{Ss3}) is the local minimum at the second part of the boundary of region I, provided

$$n_{\mathrm{Smin}} \leq 3 \tag{49}$$

as it is demanded by the second part of (36). This third candidate for global minimum will have the coordinates

$$\begin{aligned} & n_{\mathrm{Ss3}} = 3, \\ & n_{\mathrm{Bs3}} = n_{\mathrm{Bmin}}, \ \text{ no other restrictions for it.} \end{aligned} \tag{50}$$

The result delay will be

$$\begin{aligned} & T_{\mathrm{RTW3}} \left(n_{\mathrm{Bs3}} = n_{\mathrm{Bmin}}, n_{\mathrm{Ss3}} = 3 \right) = \tau_{\mathrm{R}} n_{\mathrm{Bmin}} + \left(3\tau_{\Sigma} \right) \left[\ln \left(\frac{n_{\mathrm{T}}}{n_{\mathrm{Bmin}}} \right) \right] = \\ & = T_{\mathrm{RR\,min}} + \left(3\tau_{\Sigma} \right) \left[\ln \left(\frac{n_{\mathrm{T}}}{\frac{T_{\mathrm{RR\,min}}}{\tau_{\mathrm{R}}} + \eta_{\mathrm{B}}} \right) \right]. \end{aligned} \tag{51}$$

As it is mentioned above (50) does not put any restriction on n_{Bmin} but when (51) is analyzed, the minimum of (51) is reached when

$$n_{\mathrm{Bmin}} = n_{\mathrm{Bs3}} = \frac{\tau_{\Sigma}}{\tau_{\mathrm{R}}}\mathrm{e} + \eta_{\mathrm{B}} = \frac{PL_{\Sigma\mathrm{R}}}{PL_{\mathrm{RR}}}\mathrm{e} + \eta_{\mathrm{B}}, \quad \eta_{\mathrm{B}} \in \left(-0.5; +0.5\right\rangle. \tag{52}$$

In this case this third candidate coincides with the first candidate specified by (20) and (21). Again when n_{Bmin} is larger then specified by (52), then minimum lays in the point (n_{Bmin}, n_{Smin}).

The point (n_{Bmin}, n_{Smin}) is the last part of the region I and it was already mentioned several times as the best candidate for minimum delay and therefore it is worth inspecting it. The total delay at this point has the following value

$$T_{\mathrm{RTW4}}\left(n_{\mathrm{Bmin}}, n_{\mathrm{Smin}}\right) = \tau_{\mathrm{R}} n_{\mathrm{Bmin}} + \tau_{\Sigma} n_{\mathrm{Smin}} \left[\log_{n_{\mathrm{Smin}}}\left(\frac{n_{\mathrm{T}}}{n_{\mathrm{Bmin}}}\right)\right] = T_{\mathrm{RRmin}} + \eta_{\mathrm{TR}} + \frac{\left(T_{\Sigma\mathrm{Rmin}} + \eta_{\mathrm{T}\Sigma}\right)}{\ln\left(\frac{T_{\Sigma\mathrm{Rmin}}}{\tau_{\Sigma}} + \eta_{\mathrm{S}}\right)} \ln\left(\frac{n_{\mathrm{T}}}{\frac{T_{\mathrm{RRmin}}}{\tau_{\mathrm{R}}} + \eta_{\mathrm{B}}}\right), \tag{53}$$

where

$$\eta_{\mathrm{TR}} = \eta_{\mathrm{B}}\tau_{\mathrm{R}}, \quad \eta_{\mathrm{T}\Sigma} = \eta_{\mathrm{S}}\tau_{\Sigma}. \tag{54}$$

Provided that there are no restrictions for n_{Bmin} and n_{Smin}, i.e. T_{RRmin} and $T_{\Sigma\mathrm{Rmin}}$ are so small (or not specified at all) that following values of n_{Bmin} and n_{Smin} are allowed:

$$n_{\mathrm{Bmin}} = \frac{\tau_{\Sigma}}{\tau_{\mathrm{R}}}\mathrm{e} + \eta_{\mathrm{B}}, \quad \eta_{\mathrm{B}} \in \left(-0.5; +0.5\right\rangle \tag{55}$$

and

$$n_{\mathrm{Smin}} = 2, \tag{56}$$

then (53) changes into

$$T_{\mathrm{RTW4}}\left(n_{\mathrm{Bmin}} = \frac{\tau_{\Sigma}}{\tau_{\mathrm{R}}}\mathrm{e} + \eta_{\mathrm{B}}, n_{\mathrm{Smin}} = 2\right) = \tau_{\mathrm{R}}\left\{\frac{\tau_{\Sigma}}{\tau_{\mathrm{R}}}\left[\mathrm{e} + 2\log_2\left(\frac{n_{\mathrm{T}}}{\frac{\tau_{\Sigma}}{\tau_{\mathrm{R}}}\mathrm{e} + \eta_{\mathrm{B}}}\right)\right] + \eta_{\mathrm{B}}\right\}. \tag{57}$$

When we compare visually the candidates for minimum and expressions for the worst-case total delay we find them quite close to each other. In the case when $T_{\Sigma\mathrm{Rmin}}$ is so small (or not specified at all) so that $n_{\mathrm{Smin}} = 2$ or $T_{\Sigma\mathrm{Rmin}} < \mathrm{e}\tau_{\Sigma}$ and $T_{\mathrm{RRmin}} < \mathrm{e}\tau_{\Sigma}$ then we can examine all candidates for delay minimum in the region I. As it was already mentioned for $n_{\mathrm{Smin}} > 3$ the point (n_{Smin}, n_{Bmin}) is the only candidate for minimum in region I. To compare them we use the following input parameters PL_{RR} = 736 bits and $PL_{\Sigma\mathrm{R}}$ = 11296 bits, $n_{\mathrm{T}} = 2.2 \cdot 10^6$ terminals, $BW_{\mathrm{AR}} = BW_{\mathrm{A}\Sigma} = BW_{\mathrm{A}}$ = 37.5 kbps and when $T_{\mathrm{RRmin}} = T_{\Sigma\mathrm{Rmin}}$. The obtained results are shown in Tab. 1.

			region I		
Tsrmin (s)	nSmin	nBmin	nSs	nBs	Trtwl (s)
0.45	2	24	2	31	10.32
			2	42	10.27
			2	45	10.27
			3	42	**9.76**
			2	24	10.40
0.75	3	39	3	42	**9.76**
			3	44	9.76
			3	42	**9.76**
			3	39	9.76
1.05	4	54	4	54	10.29
1.36	5	70	5	70	11.06
1.66	6	85	6	85	11.92
1.96	7	100	7	100	12.80

Tab. 1. Delay values in candidate points in region I.

It can be seen that the results are very similar. As the global minimum is searched more detailed results are presented in Section 3.5.

3.2 Optimization in Region II

Region II is specified by

$$n_{\mathrm{B}} \geq n_{\mathrm{Bmin}} \wedge n_{\mathrm{S}} < n_{\mathrm{Smin}}, \tag{58}$$

and by

$$T_{\mathrm{RTWII}} = \tau_{\mathrm{R}} n_{\mathrm{B}} + T_{\Sigma\mathrm{R\,min}} \left[\log_{n_{\mathrm{S}}}\left(\frac{n_{\mathrm{T}}}{n_{\mathrm{B}}}\right)\right]. \tag{59}$$

The function (59) is again continuous in the whole region II and also smooth, therefore the first and also second derivatives can be calculated:

$$\frac{\partial T_{\mathrm{RTWII}}}{\partial n_{\mathrm{B}}} = \tau_{\mathrm{R}} - \frac{T_{\Sigma\mathrm{Rmin}}}{n_{\mathrm{B}} \ln n_{\mathrm{S}}}, \tag{60}$$

$$\frac{\partial T_{\mathrm{RTWII}}}{\partial n_{\mathrm{S}}} = -T_{\Sigma\mathrm{Rmin}} \ln\left(\frac{n_{\mathrm{T}}}{n_{\mathrm{B}}}\right)\frac{1}{n_{\mathrm{S}} \ln^2 n_{\mathrm{S}}}. \tag{61}$$

When we put the derivative (60) equal to zero we obtain

$$n_{\mathrm{Bs}} = \frac{T_{\Sigma\mathrm{Rmin}}}{\tau_{\mathrm{R}} \ln n_{\mathrm{S}}} = \frac{\tau_{\Sigma} n_{\mathrm{Smin}}}{\tau_{\mathrm{R}} \ln n_{\mathrm{S}}}. \tag{62}$$

The second derivative of (60) is positive at the point (62) but the condition (58) has to be met. Therefore

$$\frac{T_{\Sigma\mathrm{Rmin}}}{\tau_{\mathrm{R}} \ln n_{\mathrm{S}}} \geq n_{\mathrm{Bmin}}, \quad \frac{T_{\Sigma\mathrm{Rmin}}}{\tau_{\mathrm{R}} \ln n_{\mathrm{S}}} \geq \frac{T_{\mathrm{RRmin}}}{\tau_{\mathrm{R}}}, \quad n_{\mathrm{S}} \leq \mathrm{e}^{\left(T_{\Sigma\mathrm{Rmin}}/T_{\mathrm{RRmin}}\right)}. \tag{63}$$

As the derivative (61) is negative in the whole region II the larger n_S is selected the lower value of T_{RTW} is obtained. Therefore the nearest lower integer should be selected:

$$n_{Ssmax} = e^{\left[\left(T_{\Sigma Rmin}/T_{RRmin}\right)\eta_S\right]}; \quad \eta_S \in (0.5,1). \tag{64}$$

Then

$$n_{Bs} = \frac{T_{\Sigma Rmin}}{\tau_R \ln n_{Ssmax}} + \eta_B = \frac{T_{\Sigma Rmin}}{\tau_R \frac{T_{\Sigma Rmin}}{T_{RRmin}}\eta_S} + \eta_B = \frac{T_{RRmin}}{\tau_R \eta_S} + \eta_B = \frac{n_{Bmin}}{\eta_S} + \eta_B. \tag{65}$$

When (65) is inserted into (59) we get

$$T_{RTW} = \frac{T_{RR\,min}}{\eta_S}\left[1+\ln\left(\frac{n_T\eta_S}{n_{Bmin}+\eta_B\eta_S}\right)\right] + \tau_R\eta_B = \tau_R\left\{\frac{n_{Bmin}}{\eta_S}\left[1+\ln\left(\frac{n_T\eta_S}{n_{Bmin}+\eta_B\eta_S}\right)\right] + \eta_B\right\}. \tag{66}$$

By analysis of (66) it was found that the function has positive derivative according to the parameter n_{Bmin}

$$\frac{\partial T_{RTW}}{\partial n_{Bmin}} = \frac{\tau_R}{\eta_S}\left[\ln\left(\frac{n_T\eta_S}{n_{Bmin}+\eta_B\eta_S}\right) + \frac{\eta_B\eta_S}{n_{Bmin}+\eta_B\eta_S}\right] \tag{67}$$

in the whole region II for all $n_{Bmin} \in \langle 2, n_T \rangle$ so that n_{Bmin} should be selected as low as possible. In the case when $T_{RRmin} = T_{\Sigma Rmin}$ the restriction (63) results in the requirement

$$n_{Ssmax} = 2\,. \tag{68}$$

Therefore to degrease T_{RTW}, n_{Smin} should be selected as low as possible, i.e.

$$n_{Smin} = 3\,. \tag{69}$$

Then

$$n_{Bs} = \frac{T_{\Sigma Rmin}}{\tau_R \ln n_{Ssmax}} + \eta_B = \frac{T_{\Sigma Rmin}}{\tau_R \ln 2} + \eta_B = \frac{3}{\ln 2}\frac{\tau_\Sigma}{\tau_R} + \eta_B = \frac{3}{\ln 2}\frac{PL_{\Sigma R}}{PL_{RR}} + \eta_B; \quad \eta_B \in (-0.5;+0.5\rangle, \tag{70}$$

provided the same bandwidths are allocated both for group of terminals and group of summarization nodes and when expression (8) is used. Then

$$T_{RTWII}\left(n_{BsII} = \frac{3}{\ln 2}\frac{PL_{\Sigma R}}{PL_{RR}} + \eta_B, n_{SsII} = 2\right) = \frac{PL_{RR}}{BW_A}\left\{\frac{3}{\ln 2}\frac{PL_{\Sigma R}}{PL_{RR}}\left[1+\ln\left(\frac{n_T}{\frac{3}{\ln 2}\frac{PL_{\Sigma R}}{PL_{RR}} + \eta_B}\right)\right] + \eta_B\right\}. \tag{71}$$

In other cases, i.e. when $n_{Smin} > 2$, the point (n_{Bmin}, n_{Smin}-1) provides the minimum of delay.

3.3 Optimization in Region III

Region III is specified by expressions

$$n_B < n_{Bmin} \wedge n_S \geq n_{Smin}\,, \tag{72}$$

and

$$T_{RTWIII} = T_{RR\,min} + (\tau_\Sigma n_S)\left[\log_{n_S}\left(\frac{n_T}{n_B}\right)\right]. \tag{73}$$

Its derivative according to the parameter n_S is

$$\frac{\partial T_{RTW}}{\partial n_S} = \tau_\Sigma \ln\left(\frac{n_T}{n_B}\right)\frac{\ln n_S - 1}{\ln^2 n_S}, \tag{74}$$

which when laid to zero brings the result

$$n_{Ss} = e\,(2.71828...) + \eta_S = 3\,. \tag{75}$$

The derivative according to the parameter n_B is

$$\frac{\partial T_{RTW}}{\partial n_B} = -\frac{\tau_\Sigma n_S}{n_B \ln n_S}, \tag{76}$$

which is negative in the whole region III therefore n_B should be as large as possible. The constraints (72) specify the values for n_{Smin} and n_B:

$$\begin{aligned} &n_{Smin} \leq 3, \\ &T_{RRmin} = T_{\Sigma Rmin} \leq 3\tau_\Sigma \\ &n_B = n_{Bmin} - 1 = \frac{\max(T_{RRmin})}{\tau_R} + \eta_{Bmin} - 1 = 3\frac{\tau_\Sigma}{\tau_R} + \eta_{Bmin} - 1. \end{aligned} \tag{77}$$

Then

$$T_{RTWIII} = T_{RR\,min} + 3\tau_\Sigma\left[\log_3\left(\frac{n_T}{n_{Bmin}-1}\right)\right] = 3\tau_\Sigma\left[1+\log_3\left(\frac{n_T}{3\frac{\tau_\Sigma}{\tau_R} + \eta_{Bmin} - 1}\right)\right] \tag{78}$$

When n_{Smin} is selected > 3, which is highly probable in real systems, then to minimize total delay n_S should be selected as small as possible, i.e.

$$\begin{aligned} &n_S = n_{SminZ+} = \frac{T_{\Sigma Rmin}}{\tau_\Sigma} + \eta_{Smin} = n_{Smin}, \quad \eta_{Smin} \in \langle 0,1), \\ &n_B = n_{BminZ-} = \frac{T_{RRmin}}{\tau_R} + \eta_{Bmin} - 1 = n_{Bmin} - 1, \quad \eta_{Bmin} \in \langle 0,1). \end{aligned} \tag{79}$$

Hence the point (n_{Bmin}-1, n_{Smin}) provides the minimum of delay whose value is

$$T_{RTWIII} = T_{RR\,min} + n_{Smin}\tau_\Sigma\left[\log_{n_{Smin}}\left(\frac{n_T}{n_{Bmin}-1}\right)\right]. \tag{80}$$

3.4 Optimization in Region IV

Region IV is specified by

$$n_{\mathrm{B}} < n_{\mathrm{Bmin}} \wedge n_{\mathrm{S}} < n_{\mathrm{Smin}} \tag{81}$$

and by the expression for the worst-case total delay

$$T_{\mathrm{RTWIV}} = T_{\mathrm{RR\,min}} + T_{\Sigma\mathrm{R\,min}} \left[\log_{n_{\mathrm{S}}} \left(\frac{n_{\mathrm{T}}}{n_{\mathrm{B}}} \right) \right]. \tag{82}$$

The derivatives according to both variables n_{S} and n_{B} are negative in the whole region IV

$$\frac{\partial T_{\mathrm{RTWIV}}}{\partial n_{\mathrm{S}}} = -\frac{T_{\Sigma\mathrm{Rmin}}}{n_{\mathrm{S}} \ln^2 n_{\mathrm{S}}} \ln\left(\frac{n_{\mathrm{T}}}{n_{\mathrm{B}}} \right), \tag{83}$$

$$\frac{\partial T_{\mathrm{RTWIV}}}{\partial n_{\mathrm{B}}} = -\frac{T_{\Sigma\mathrm{Rmin}}}{n_{\mathrm{B}} \ln n_{\mathrm{S}}}, \tag{84}$$

therefore the minimum of the delay in this region will lay in the point

$$\begin{aligned} n_{\mathrm{S}} &= n_{\mathrm{SminZ\text{-}}} = \frac{T_{\Sigma\mathrm{Rmin}}}{\tau_{\Sigma}} - \eta_{\mathrm{Smin}} = n_{\mathrm{Smin}} - 1, \quad \eta_{\mathrm{Smin}} \in \left(0,1\right\rangle, \\ n_{\mathrm{B}} &= n_{\mathrm{BminZ\text{-}}} = \frac{T_{\mathrm{RRmin}}}{\tau_{\mathrm{R}}} - \eta_{\mathrm{Bmin}} = n_{\mathrm{Bmin}} - 1, \quad \eta_{\mathrm{Bmin}} \in \left(0,1\right\rangle. \end{aligned} \tag{85}$$

Then the delay will be

$$T_{\mathrm{RTWIV}} = T_{\mathrm{RR\,min}} + T_{\Sigma\mathrm{R\,min}} \left[\log_{(n_{\mathrm{Smin}}-1)} \left(\frac{n_{\mathrm{T}}}{n_{\mathrm{Bmin}} - 1} \right) \right]. \tag{86}$$

It could seem that the result of (86) is always larger than the result at the point (n_{Bmin}, n_{Smin}) in the region I) but some times it can be even smaller at the point (n_{Bmin}-1, n_{Smin}-1) and this can happen due to the rounding processes for n_{Bmin} and n_{Smin} as shown in (87).

$$\begin{aligned} \Delta_{\mathrm{I\text{-}IV}} &= T_{\mathrm{RTWI}}\left(n_{\mathrm{Bmin}}, n_{\mathrm{Smin}}\right) - T_{\mathrm{RTWIV}}\left(n_{\mathrm{Bmin}} - 1, n_{\mathrm{Smin}} - 1\right) = \\ &= \tau_{\mathrm{R}} n_{\mathrm{Bmin}} + \tau_{\Sigma} n_{\mathrm{Smin}} \left[\log_{n_{\mathrm{Smin}}} \left(\frac{n_{\mathrm{T}}}{n_{\mathrm{Bmin}}} \right) \right] - T_{\mathrm{RR\,min}} - \\ &\quad - T_{\Sigma\mathrm{R\,min}} \left[\log_{(n_{\mathrm{Smin}}-1)} \left(\frac{n_{\mathrm{T}}}{n_{\mathrm{Bmin}} - 1} \right) \right] = \\ &= \eta_{\mathrm{TR}} + \eta_{\mathrm{T}\Sigma} \left[\log_{n_{\mathrm{Smin}}} \left(\frac{n_{\mathrm{T}}}{n_{\mathrm{Bmin}}} \right) \right] - \\ &\quad - T_{\Sigma\mathrm{R\,min}} \left[\log_{(n_{\mathrm{Smin}}-1)} \left(\frac{n_{\mathrm{T}}}{n_{\mathrm{Bmin}} - 1} \right) - \log_{n_{\mathrm{Smin}}} \left(\frac{n_{\mathrm{T}}}{n_{\mathrm{Bmin}}} \right) \right]. \end{aligned} \tag{87}$$

3.5 Evaluation of Complete n_{S}-n_{B} Plane

Matlab mathematical tool was used to evaluate theoretical presumptions. Each candidate point ($n_{\mathrm{B}si}$, $n_{\mathrm{S}si}$) in all regions that was mathematically derived and expressed in the text above was numerically evaluated. Input parameters for calculations were: PL_{RR} = 736 bits, $PL_{\Sigma\mathrm{R}}$ = 11296 bits, n_{T} = 2.2·10^6 terminals, BW_{AR} = $BW_{\mathrm{A}\Sigma}$ = BW_{A} = 37.5 kbps and T_{RRmin} = $T_{\Sigma\mathrm{Rmin}}$. The results in numerical form for the most interesting part of n_{S}-n_{B} plane are shown in Tab. 2.

Symbol “x” in region II and IV in the case of n_{Smin} = 2 means that these regions do not exist. The last row in the table shows that the global minimum may be located also in a region different from region I although the differences in delay are minimal.

			region I			region II			region III			region IV		
$T_{\Sigma\mathrm{Rmin}}$ (s)	n_{Smin}	n_{Bmin}	n_{Ss}	n_{Bs}	T_{RTWI} (s)	n_{Ss}	n_{Bs}	T_{RTWII} (s)	n_{Ss}	n_{Bs}	T_{RTWIII} (s)	n_{Ss}	n_{Bs}	T_{RTWIV} (s)
0.45	2	24	2	31	10.32	x	x	x	2	23	10.42	x	x	x
			2	42	10.27				3	23	9.89			
			2	45	10.27									
			3	42	**9.76**									
			2	24	10.40									
0.75	3	39	3	42	**9.76**	2	55	12.59	3	38	9.77	2	38	12.67
			3	44	9.76	2	39	12.65						
			3	42	**9.76**									
			3	39	9.76									
1.05	4	54	4	54	**10.29**	3	54	11.25	4	53	10.30	3	53	11.26
1.36	5	70	5	70	**11.06**	4	70	11.50	5	69	11.06	4	69	11.50
1.66	6	85	6	85	**11.92**	5	85	12.13	6	84	11.92	5	84	12.13
1.96	7	100	7	100	12.80	6	100	12.89	7	99	**12.80**	6	99	12.90

Tab. 2. Delay values in the derived candidate points for searching of global minimum of the total delay.

Graphical form for result presentation was used for longer sequence of $T_{\Sigma\mathrm{Rmin}}$ (the name Tsrmin was used in figures).

Fig. 6 shows the course of minimum delay as the function of $T_{\Sigma\mathrm{Rmin}}$ (or n_{Smin} which is proportional to $T_{\Sigma\mathrm{Rmin}}$). This graph can be used to determine the largest allowed transmission period when the maximum total delay T_{RTmax} is specified. For example when T_{RTmax} = 60 seconds, then max($T_{\Sigma\mathrm{Rmin}}$) = 20 seconds.

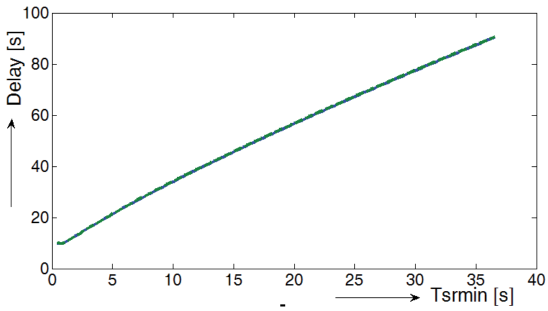

Fig. 6. Worst-case total delay as the function of minimal transmission period $T_{\Sigma\mathrm{Rmin}}$.

From Fig. 7 and Fig. 8 we can determine n_B (1050) and n_S (67). Fig. 9 shows the required number of tree levels (3) and Fig. 10 enables to find the required number of summarization nodes (2103). It can be seen that the numbers of the required tree levels and mainly of the summarization nodes increase rapidly when $T_{\Sigma\mathrm{Rmin}}$ decreases, which is due to small values of the parameters n_{Bopt} and n_{Sopt}.

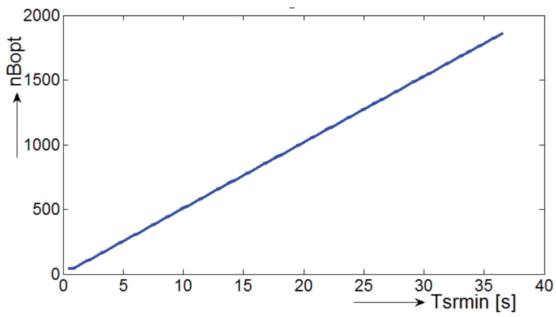

Fig. 7. Optimal value of n_B as the function of $T_{\Sigma\mathrm{Rmin}}$.

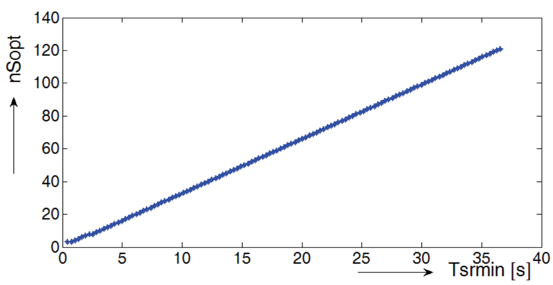

Fig. 8. Optimal value of n_S as the function of $T_{\Sigma\mathrm{Rmin}}$.

As it was mentioned above global delay minimum locations vary with $T_{\Sigma\mathrm{Rmin}}$ as it can be also seen from Fig. 11 where differences n_{Smin}-n_{Sopt} (nSd) and n_{Bmin}-n_{Bopt} (nBd) are depicted and appropriate regions written.

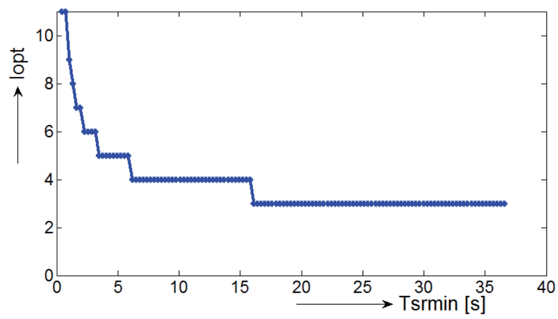

Fig. 9. Required number of tree levels I_{opt}.

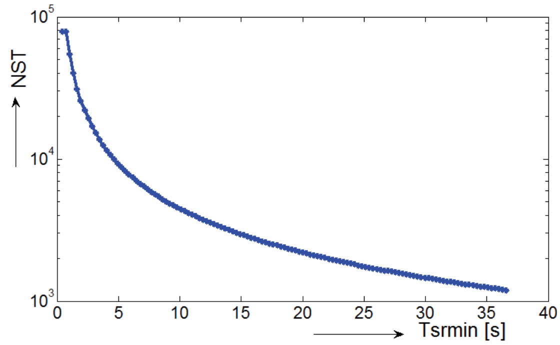

Fig. 10. Required number of summarization nodes N_{ST}.

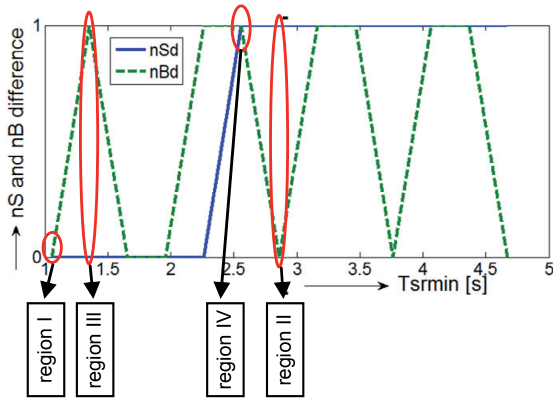

Fig. 11. Locations of global delay minima.

Fig. 12 shows the courses of important tree quantities as the functions of the total number of terminals n_T. In this case the minimum transmission period was constant: $T_{\Sigma\mathrm{Rmin}}$ = 5s. It can be seen that the optimal values of n_B and n_S remain also constant n_{Bopt} = 255 and n_{Sopt} = 17.

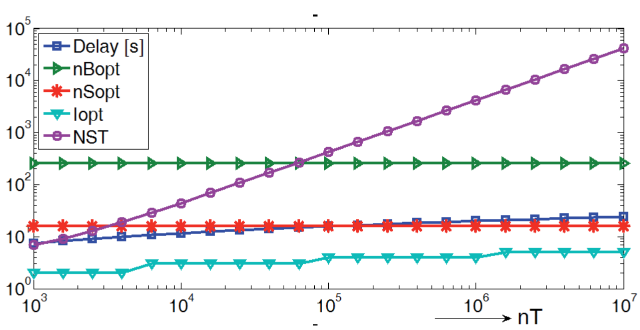

Fig. 12. Tree quantities as the function of the total number of terminals n_T

These values are just equal to region I boundary values n_{Bmin} and n_{Smin}. The number of tree levels I_{opt} gradually raises with n_T increase and the delay increase is quite mild, while the required number of summarization nodes escalates quite rapidly.

4. Conclusion

This article addressed the problem of optimization of hierarchical architecture for data acquisition in network environment. The process of tree design was presented and optimization problems were presented and partially solved. It was proved that in majority cases the optimum or almost optimum architecture of the tree will be reached when n_{Bopt} and n_{Sopt} are put equal to parameters n_{Bmin} and n_{Smin} that are obtained from the input parameters T_{RRmin} and $T_{\Sigma Rmin}$. Provided that the number of required summarization nodes exceeded the number of summarization nodes available then it is necessary to calculate the minimum number of terminals in one group to meet the number of available summarization nodes. This step unfortunately increases the total data acquisition delay so that it is necessary to check whether the maximum acceptable delay was exceeded or not. When it does happen it is possible to try to increase the number of summarization nodes in one group slightly, which may cause reduction both of delay and of the total number of summarization nodes.

Acknowledgements

This article was written with the support of the Ministry of Education of the Czech Republic, fund FRVS, project No. 2954/2010.

References

[1] SCHULZRINNE, H., CASNER, S., FREDERICK, R., JACOBSON, V. *RTP: A Transport Protocol for Real-Time Applications*. Internet Draft, IETF RFC3550, 2003.

[2] HOLBROOK, H., CAIN, B. S*ource-Specific Multicast for IP*. Internet Draft, IETF, 2004.

[3] BHATTACHARYYA, S. *An Overview of Source-Specific Multicast (SSM)*. Request for Comments 3569, IETF, 2003.

[4] CHESTERFIELD, J., SCHOOLER, E. M. An extensible RTCP control framework for large multimedia distributions. In *The second IEEE International Symposium on Network Computing and Applications*, 2003, p.351,ISBN: 0-7695-1938-5.

[5] CHESTERFIELD, J., OTT, J., SCHOOLER, E. M. *RTCP Extensions for Single-Source Multicast Sessions with Unicast Feedback*. Internet draft, IETF draft-ietf-avt-rtcpssm-13.txt, 2007

[6] EL-MARAKBY, R., HUTCHISON, D. Scalability improvement of the Real-Time Control Protocol (RTCP) leading to management facilities in the Internet. In *ISCC - Third IEEE Symposium on Computers & Communications*, 1998, p. 125.

[7] CASTRO, M., DRUSCHEL, P., KERMARREC, A., ROWSTRON, A. A large-scale and decentralized application-level multicast infrastructure. *IEEE Journal on Selected Areas in Communications*, 2002, vol. 20, no. 8, p. 1489-1499.

[8] NOVOTNÝ, V., KOMOSNÝ, D. Optimization of large-scale RTCP feedback reporting in fixed and mobile networks. In *Proceedings of International Scientific Conference on Wireless and Mobile Communications ICWMC2007.* Guadeloupe, March 2007, p. 1 – 6, ISBN: 0-7695-2796-5.

[9] NOVOTNÝ, V., KOMOSNÝ, D. Large-scale RTCP feedback optimization. *Journal of Networks*, 2008, no. 3, p. 1-10. ISSN: 1796-2056.

[10] KOMOSNÝ, D., BURGET, R., NOVOTNÝ, V. Tree transmission protocol for feedback distribution in IPTV systems. In *Proceedings of the Seventh IASTED International Conference on Communication Systems and Networks*. Palma de Mallorca (Spain), 2008, p. 1-7. ISBN: 978-0-88986-758-1.

[11] BURGET, R., KOMOSNÝ, D., NOVOTNÝ, V. Integration of host position prediction into hierarchical aggregation. In *Seventh International Conference on Networking ICN 2008*. 2008, p. 740 to 744, ISBN: 978-0-7695-3106-9.

[12] NOVOTNÝ, V., KOMOSNÝ, D., KATHIRAVELU, G., BURGET, R. Large-scale RTCP feedback theory and implementation. In *Proceeding of the 3rd Mosharaka International Conference on Communications, Computers and Applications*. Amman (Jordan), 2009, p. 1-6, ISBN 978-9957-486-07-5.

About Authors ...

Vít NOVOTNÝ was born in 1969. He received his M.Sc. in 1992, his Ph.D. degree in 2001, both from the Brno University of Technology in “Electronics and Communication Technologies”. In 2005 he became an assistant professor at the same university, again in the area of “Electronics and Communication Technologies”. In the past he did the research in the areas of non-filtering applications of switched capacitors and of the current and voltage conveyors. Current professional interests are mobile and integrated packet data networks, their services and terminal equipment. Now he works with the Dept. of Telecommunications, Brno University of Technology.