OPTIMIZATION OF TRANSPORT MOBILE COMMUNICATION NETWORK

Ihor HVOSTYK 1, Lubov TOKAR 1, Yana KRASNOZHENIUK 1,*,
Al-Vandavi Saif Ahmed ISKANDAR 1

1 V.V. Popovskyy Department of Infocommunication Engineering, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

ABSTRACT

The mechanism of Ethernet signal transmission over the air is considered. The advantage of this transmission method is the minimum delay with high-performance backhaul. It is shown that monitoring delay using brute-force methods at the port or software level does not reveal delay problems affecting certain traffic classes. For deep performance delays, it is necessary to take measurements that cover the full service path at the demarcation points and at the key nodes between them. The relevance of methods for delays compensation in infocommunication networks is confirmed and shown defined by direct influence of a delay on quality of service. The complex decision of delays suppression on the basis of an optimum estimation of a regulator mode is offered.

Keywords: Transport network, Delays in transport network

1. INTRODUCTION

Ethernet mobile services and implementation of mobile systems based on packet data transmission will provide mobile operators with scalable and more economical solutions for processing both the growing number of mobile devices connected to their networks and the volume of traffic.

One of the main reasons for the wider use of Carrier Ethernet in backhaul wireless applications is the ability to use a variety of physical infrastructures to deliver Carrier Ethernet to a base station. One of the physical mechanisms of signal transmission is Ethernet over a radio channel, which is characterized by a minimum delay and a reliable, high-performance backhaul for 3G and 4G wireless networks. It is known, however, that real-time data transfer, transactional applications, high-speed roaming and multimedia streaming are sensitive to delays. Increasing the delay by only a few milliseconds can lead to voice distortion, disruption of the application, which will affect significant losses of operators.

2. PURPOSE AND RESEARCH OBJECTIVES

When transmitting multimedia information, special attention is paid to the delay requirements and variations of this delay (jitter). Studies have shown that the delay value \( \tau_\Sigma \) is affected by two main factors: \( \tau_T \) is the delay in various network elements and \( \tau_R \) is the delay associated with physical restrictions on the communication line and it is cumulative. In accordance with Rec. ITU-TU.1541, the delay \( \tau_R \) should be taken into account as a linear parameter of \( \tau_R = 5 \, \mu s/km \). Therefore, the overall delay is

\[
\tau_\Sigma = R \tau_R + \tau_T
\]

where \( R \) is the length of the backbone.

*Corresponding Author: yana.krasnozheniuk@nure.ua
Received: 14.10.2019 Published: 16.12.2019
At the same time, the delay requirements on the diagonal of 2500 km should not exceed $\tau_{\Sigma} = 100$ ms.

The delay $\tau_{\Sigma}$ is a random process with the dispersion $\sigma^2_{\tau}$ interpreted as the jitter span. The presence of jitter directly affects the quality of transmission of real-time traffic (VoIP, video) and manifests itself in the form of a synchronization failure in corresponding overlays of frames. Moreover, the presence of jitter generates additional noise in the form of the appearance of harmonic components with a frequency equal to half the speed of the synchronous stream [1]. These components further affect the playback quality.

The presence of a jitter compensation buffer makes it possible to facilitate the situation and minimize jitter by the level of mean square error (MSE) $\sigma^2_{\tau} = 10 - 15$ ms, which is often sufficient for high-quality multimedia reception. At the same time, in recent years, applications have appeared that require gigabit access speeds and shorter delays of $\tau_{\Sigma} < 10$ ms (for example, for medical services), as well as jitter reduction [2]. Under these conditions, an integrated and faster response is required with adaptation according to the “predictor-corrector” principle.

In this paper, instead of the usual stochastic approximation (SA) procedures, it is proposed to use Kalman-Bucy (KB) filters with estimation of the complex weight coefficient of the corrector, which allows for simultaneous delay compensation and jitter correction.

3. MATERIALS AND METHODS FOR ANALYZING SUPPRESSION OF DELAYS BASED ON OPTIMAL ESTIMATION OF REGULATOR MODE

Obviously, the total delay (1), as a random process, can be represented as a sum:

$$\tau_{\Sigma}(k) = \tau_{\Sigma0}(k) + \Delta\tau_{\Sigma}(k), \quad k = 1, 2, ..., \quad (2)$$

where $\tau_{\Sigma0}(k)$ is a constant or slowly changing average delay value,

$\Delta\tau_{\Sigma}(k)$ is a rapidly changing variation with dispersion $\sigma^2_{\tau}$.

The delay correction (2) at the output of the buffer compensator can be implemented in two ways:

1. Using additive corrective action

$$\tau_{\text{out}}(k) = \tau_{\Sigma}(k) \pm \Delta\tau_{\text{cor}}(k), \quad (3)$$

where for the difference $(\tau_{\Sigma}(k) - \Delta\tau_{\text{cor}}(k))$ a minimum of the mean square deviation is provided.

2. Using the multiplier corrector

$$\tau_{\text{out}}(k) = w(k)\tau_{\Sigma}(k), \quad (4)$$

where $w(k)$ is the complex weighting factor (CWF) represents a regulator.

As a delay indicator, in both the first and second cases, the value of the residual can be used, i.e. the phase difference of the input and reference signals:

$$\chi(k) = \tau_{\Sigma}(k) - t_0(k), \quad (5)$$

where $t_0(k)$ is the phase of the reference generator.
By the value of the residual, it is possible to find the optimal value of the CWF, which in the first method provides minimization (5) by correcting the phase of the reference generator, and in the second method by correcting the phase of the input signal. The relevant recursive Widrow algorithm for finding an estimate of the optimal CMF value \( \hat{w}(k) \) is not suitable, since it has optimal properties for estimating random variables and slowly changing processes [1]. This means that such an algorithm in this situation will be constantly in transition mode and, accordingly, give large errors.

For the rapidly changing situation, the KB procedure should be used, which is optimal for evaluating random processes, the measurement speed of which can be represented by the value of the correlation interval \( \alpha \), where \( R = \sigma^2 \exp\left(-\alpha|\tau|\right) \) is the form of the correlation function.

The optimal procedure for estimating CMF according to KB is:

\[
\hat{w}(k) = F(k/k - 1)\hat{w}(k - 1) + K(k)[\tau_\Sigma(k - 1) - \hat{w}(k - 1)t_0(k)], t_0(k)
\]

where \( F(k/k - 1) = \exp\{-\alpha\Delta t\} \), \( \Delta t = t_k - t_{k-1} \) is a sampling step.

The algorithm corresponding to procedure (6), using the multiplicative correction method (4), is shown in Fig. 1.

![Buffer control algorithm with multiplicative delay correction](image)

The principle of the algorithm is as follows. The phase mismatch signal of the output \( \tau_{out} \) and the reference \( \tau_0(k) \) signal, i.e. the residual \( \chi(k) \) is formed at the output of the adder \( \Sigma_2 \). This signal is then used to form a CWF \( \hat{w}(k) \). Next, this signal is fed to the input of the multiplier \( m_2 \), the second input of which receives the signal of the current phase of the input stream \( \tau(k) \cdot K(k) \), where \( K(k) \) is the scaling factor. At the output of the multiplier \( m_1 \), a signal is generated proportional to the phase difference of the signals at its inputs and, after the integrating delay loop, affects the multiplier \( m_1 \), creating a complex weighing impact \( \hat{w}(k) \) in the form of a CMF. The impact \( \hat{w}(k) \) forms such a phase of the output signal \( \tau_{out} \), at which the residual \( \chi \) is minimized in the mean-square sense. The desired compensation effect \( \tau_\Sigma(k) \) defined by expression (2) is thereby achieved.

An important property of the procedure used here, in contrast to the similar SA algorithm, is that the multiplier \( K(k) = f(\sigma^2_{ps}) \), which is determined depending on the change in the posterior variance of the estimate \( \sigma^2_{ps} \) and the parameter \( F(k/k - 1) \) that is the probability of the process \( \tau_\Sigma \) transitioning from state \( k - 1 \) to \( k \), provide the required speed of the KB algorithm to change the delay.
Figure 2 shows the buffer control algorithm with additive delay correction. The input 1 of the adder \( \sum_1 \) receives an information stream, at input 2 receives a signal of the reference generator \( t_0(k) \). This algorithm, in contrast to the algorithm in Fig. 1, functions in such a way that the phase of the reference generator is corrected for the phase of the input stream. In some cases, this solution is more technologically advanced.

A variant of the WCF estimation algorithm that implements an additive corrective impact (Figure 2) is characterized by the fact that the required delay compensation effect is achieved by summing (3).

![Additive delay correction buffer control algorithm](image)

Compensation occurs due to the correction of the phase of the reference generator \( t_0(k) \), which, accordingly, provides the correction of the output signal \( t_{out}(k) \). Experience in the practical use of KB algorithms (6) shows that the time spent by the filter on the transient mode is calculated in several sampling steps \( \Delta t \in (r...z) \), while for the SA algorithm this period is 50-200 steps. Together with the speed, the proposed algorithms achieve high accuracy and minimal a posteriori estimation error.

4. DISCUSSION OF RESEARCH RESULTS

Although Ethernet offers easy access to bandwidth on demand, the bandwidth of the dialing service often has little effect on delay if the link was originally correctly defined. Let us consider a cellular communication BS connected to an access platform using an optical communication line 100FX, with a CIR (fixed information transmission rate, forcibly limiting rate) of 20 Mbit/s (Figure 3). If the BS traffic never exceeds 20 Mbit/s, then increasing the throughput to 30, 50 or even 100 Mbit/s will not affect the delay. Although the capacity can be increased, each packet is still tied to the physical communication speed agreed between the ports (for example, 100 Mbit/s for 100FX media) [1]. This means that the delay is the same as long as the physical connection remains the same.

![Network architecture of the transport network](image)
As bandwidth increases, data loads faster, but each individual packet still moves at the same rate. Instead of increasing CIR, applications should connect at a faster rate to reduce delay, for example, changing from 100FX to the GbE interface would reduce delay by a factor of ten. In many cases, this requires replacement or modernization of network equipment, which is an expensive and time-consuming solution.

Under the overload, an increase in bandwidth will reduce the delay, but only if the bandwidth increases from end to end. For example, if Ethernet-over-SONET/SDH (EoS) is used for transport, increasing the last mile carrier will have little effect on overall delay if packets are still transported through the same TDM “container” (for example, DS3/E3). This is similar to the Ethernet data rate effect: the rate of traffic entering the network does not affect the delay if the TDM/core preparation remains unchanged.

Similarly, in all packet networks (for example, MPLS core), if the end-to-end network does not have sufficient capacity, the packets will be overloaded in the core instead of the access network, simply pushing out the problem elsewhere. When this happens, an increase in access bandwidth can lead to even longer delays, since already occupied NES cores add more traffic to their queues and load processing [4].

The reasons for the delay are complex and non-deterministic. To maintain overall QoS, delay and jitter should be constantly monitored based on each service, application, SLA or VLAN. Monitoring of delay using brute-force methods at the port level or software (for example, ping) does not reveal delay problems that affect certain classes of traffic, and cannot isolate whether a delay occurs at the IP or Ethernet level (Figure 4).

Likewise, end-to-end delay monitoring is not sufficient for SLA reporting and troubleshooting of delay problems in networks where traffic is often asymmetric, therefore one-way delay and jitter measurements are required. In order to provide a complete picture of the delay performance, measurements should also cover the complete service path, from end to end (demarcation point and at key nodes between them). This allows the operator to isolate delay problems for access, core networks or, more importantly, customer networks.

Delay measurements require sufficient accuracy to detect minor delay changes in order to proactively determine delay drift, which can lead to QoS problems and SLA violations. Accuracy must be at least one order of magnitude greater than the target threshold in order to be useful (for example, submilliseconds if the SLA determines the maximum delay of 10 ms). Otherwise a measurement error will mask the delay problem, generate false alerts, or even result in to negative delays in client SLA reports.
Similarly, measurements require sufficient details to detect short-term delay problems that may indicate micro-bursts, inefficient traffic management methods, or improperly configured NEs. The ability to measure every second, for example, provides enough information to diagnose intermittent QoS problems.

Thus, to ensure that delay problems are fully visible, measurements should:
- be held for each flow, application, service or VLAN;
- the delay should be measured from end to end to accurately take into account all segments of the network, as well as key intermediate nodes, to isolate problems;
- have a high degree of detail.

Therefore, to reduce the delay with the increase in bandwidth, it is important to determine if there is an overload in a service segment. Since providing additional bandwidth is time consuming and expensive from the point of view of the network, it is better to increase it where the bandwidth will increase the productivity the most.

REFERENCES


