ELSEVIER

Physica A 308 (2002) 533–544

PHYSICA A

www.elsevier.com/locate/physa

Kinetic model of network traffic

I. Antoniou[a,b,*], V.V. Ivanov[a,c], Yu.L. Kalinovsky[c]

[a] *International Solvay Institutes for Physics and Chemistry, CP-231, ULB, Campus plaine Bd. du Triomphe, 1050, Brussels, Belgium*

[b] *Department of Mathematics, Aristoteles University of Thessaloniki, 54006 Thessaloniki, Greece*

[c] *Laboratory of Information Technologies, Joint Institute for Nuclear Research, 141980, Dubna, Russia*

Received 19 November 2001

Abstract

We present the first results on the application of the Prigogine–Herman kinetic approach (Kinetic Theory of Vehicular Traffic, American Elsevier Publishing Company, Inc., New York, 1971) to the network traffic. We discuss the solution of the kinetic equation for homogeneous time-independent situations and for the desired speed distribution function, obtained from traffic measurements analysis. For the log-normal desired speed distribution function the solution clearly shows two modes corresponding to *individual flow patterns* (low-concentration mode) and to *collective flow patterns* (traffic jam mode). For low-concentration situations we found almost linear dependence of the information flow versus the concentration and that the higher the average speed the lower the concentration at which the optimum flow takes place. When approaching the critical concentration there are no essential differences in the flow for different desired average speeds, whereas for the individual flow regions there are dramatic differences. © 2002 Elsevier Science B.V. All rights reserved.

PACS: 05.45+b; 05.20.Dd

Keywords: Kinetic; Internet; Flow; Traffic; Modelling; Network

1. Introduction

Within the global framework of the information society fast and reliable data exchange between local and wide-area computer networks is a priority issue. In this connection, a major challenge for the emerging high-speed integrated-service

* Corresponding author. International Solvary Institute for Physics and Chemistry, CP 231, ULB campus plaine Bd. du Triomphe, 1050, Brussels, Belgium. Tel.: +32-2-650-55-33; fax: +32-2-650-50-28.

E-mail address: iantonio@vub.ac.be (I. Antoniou).

PII: S0378-4371(02)00585-X

communication networks is to elaborate a model that can realistically capture the basic features of network traffic. Such model may serve as the basis for the development of methods and tools for quality assessment, providing more efficient control and management of information flows in the Internet [1,2].

Prigogine and Herman developed (see Ref. [3] and references therein) a model of vehicular traffic dynamics based on the principles of statistical physics and kinetic theory. This model has proved to be quite successful [4].

In the present work, we modify the Prigogine–Herman's kinetic theory to be applicable for network traffic. We describe in Section 2 the data acquisition system for traffic measurements realized on a standard IBM PC. In Section 3 we present the basic concepts of the kinetic equation and discuss its modification related to the network traffic. In Section 4 we discuss the form of the desired speed distribution function related to the network traffic. In Section 5 we analyze the solutions of the kinetic equation for the chosen desired distribution function.

2. Data acquisition system

Our study is based on the traffic measurements obtained at the input of the Dubna University [5] local-area network (LAN), which includes approximately 200–250 interconnected computers. The performance of the data acquisition system is based on the realization of an open-mode driver [6]; see Fig. 1.

In standard conditions the network adapter of a computer is in a mode of detecting a carrying signal (main harmonic 4–6 MHz). After appearing in the cable bits of the package preamble, the network adapter comes to a mode of 1 bit and 1 byte synchronization with the transmitter and starts receiving first bytes of the package heading. As soon as one succeeds in extracting the MAC address of the shot receiver from the first bytes taken by the adapter, the network adapter compares it to its own. In the case of negative result of the comparison, the network adapter ceases to record the shot's bytes into its internal buffer and cleans its contents and then waits until the next package appears.

In order to provide conditions for receiving and analysis of all the packages transmitted over the network, it is necessary to move the adapter devices to a free mode when all possible shots are recorded in the buffer. This operation is executed through the instructions of the NDIS driver.

The free-mode driver records the accepted packages in the preliminary capture buffer and displays the flag of receiving the package. Then the receiving package module is activated and analysis of the margin of the package's type is carried out to extract TCP/IP packages from the whole stream.

After identification, it is possible to separate and delete the data block as well as to record the headers to the SQL-server database. The recording is performed together with the time data with a frequency of up to 10 kHz. Although the recording is performed with buffering, the mode of saving the packages' headers requires enormous server's

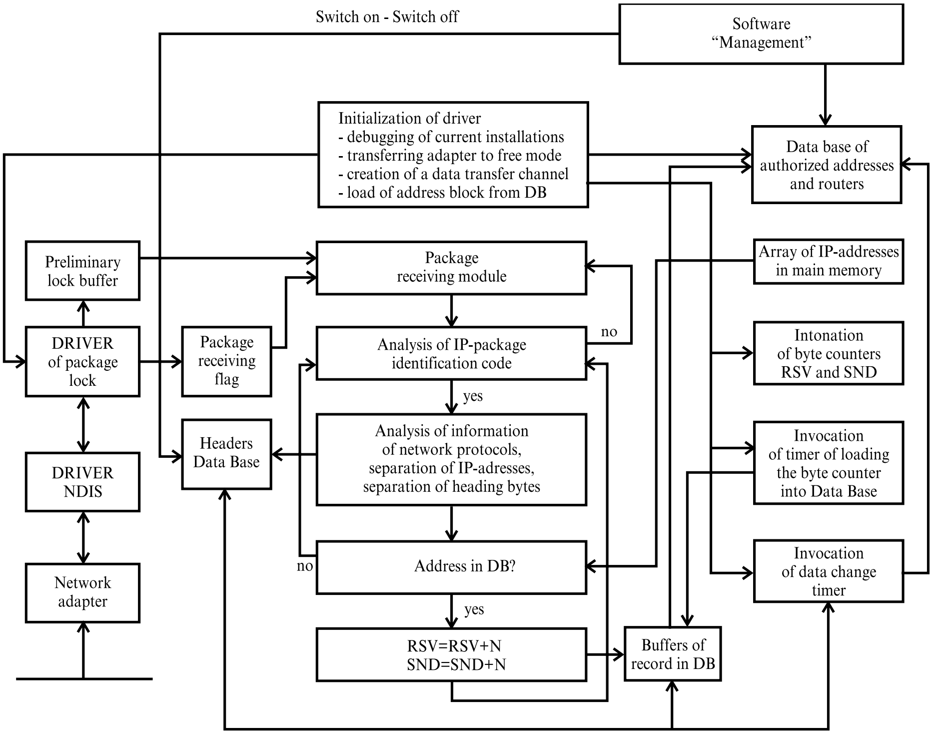

Fig. 1. Scheme of a data acquisition system.

resources, as in this case there is a permanent procedure of recording with small portions to the hard disk. That is why this mode is switched on if required at the management system's instruction.

The system also provides control over the external traffic of the LAN on the basis of controlling the records in the router table. Initial information on the legal IP addresses is saved in the database of the LAN computers from which data on legal addresses are loaded into the main memory array. The users which do not participate in forming the external traffic are not taken into account when calculating the number of transferred and received bytes. In order to decrease the number of sessions of recording the information on the external traffic in the database, a timer of load out of the buffer and a timer of changing a current date have been introduced into the system.

The recorded traffic data correspond approximately to 20 h (1,600,000 records with a frequency of up to 10 kHz, which corresponds to 1 ms bin size) of measurements. The part of this series corresponding approximately to 1 h of measurements and aggregated with different bin sizes is presented in Fig. 2.

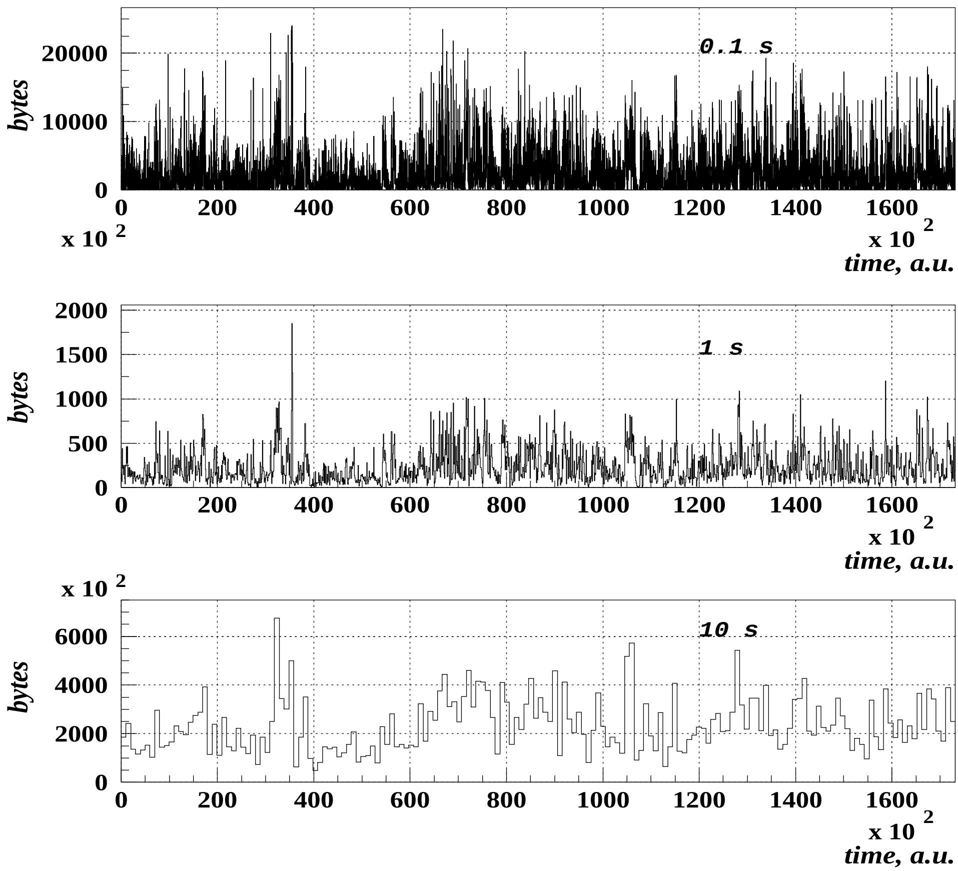

Fig. 2. Traffic measurements aggregated with different bin sizes: 0.1, 1 and 10 s.

3. Basic concepts of the kinetic model of network traffic

Following the basic concepts of the Prigogine–Herman kinetic model of vehicular traffic [3], we assume that for the network traffic as well, the velocity distribution function $f(x,v,t)$ satisfies the following kinetic equation:

$$\frac{\partial f(x,v,t)}{\partial t} + v\,\frac{\partial f(x,v,t)}{\partial v} = -\frac{f(x,v,t) - f_0(x,v,t)}{T} + c(\bar{v}-v)(1-P)f(x,v,t)\,. \tag{1}$$

Here, f_0 is the "desired" velocity distribution function [3], x and v are the position and velocity of the information packages ("vehicles"), $\bar{v}$ is the average velocity, c is the concentration of information packages, P is the probability of "passing" in the sense of increase of information flow, and T is the relaxation time.

For the network traffic in particular we assume the following:

- the space dependence x is omitted because the passing of packets is a homogeneous space-independent process,
- the velocity v is the amount of information passing during the time interval corresponding to the chosen aggregation window, and
- the time t is discrete with the unit being equal to the size of the aggregation window (for example 1 s).

The amount of information packages dN within the velocity interval between v and $v + dv$ at time t is given by

$$dN = f(v,t)\,dv\,.$$

Once the velocity distribution function $f(v,t)$ is known, we may derive other quantities involved in (1), such as the concentration

$$c(t) = \int_0^\infty dv\, f(v,t)\,, \tag{2}$$

the average velocity

$$\bar{v}(t) = \frac{1}{c(t)} \int_0^\infty v\, dv\, f(v,t)\,, \tag{3}$$

and the information flow

$$q(t) = c(t)\bar{v}(t) = \int_0^\infty v\, dv\, f(v,t)\,. \tag{4}$$

To find the distribution function $f(v,t)$, we will study first the stationary solution $f(v)$, which satisfies the following equation:

$$\frac{f(v) - f_0(v)}{T} = c(\bar{v} - v)(1 - P)f(v) \tag{5}$$

or

$$f(v) = \frac{f_0(v)}{1 - cT(\bar{v} - v)(1 - P)}\,. \tag{6}$$

The quantity $f(v)$ describes the situation in which there is a steady state between slowing down of information transfer caused by the *interaction processes* and the speeding up corresponding to more effective passing of information through network.

In order to perform calculations, we need to know how the relaxation time T and the probability P of passing change with concentration. Following Ref. [3] we suppose that the relaxation time T is of the form

$$T = \frac{\tau(1 - P)}{P}\,, \tag{7}$$

where τ is an arbitrary parameter. We also assume [3] that the probability P of passing has the form

$$P = \begin{cases} 1 - c/c_p = 1 - \eta, & c < c_p , \\ 0, & c > c_p , \end{cases} \tag{8}$$

where c_p is the limiting or jam concentration when the information packets can no longer pass and η is the normalized concentration. Note that for small concentration c, the relaxation time T is also small, whereas for large concentrations the relaxation time is also very large.

Let us go back to the stationary solution of the kinetic equation. The denominator can be rewritten as follows:

$$1 - cT\bar{v}(1 - P) + cTv(1 - P) .$$

In the case

$$1 - cT\bar{v}(1 - P) > 0 , \tag{9}$$

f is given by

$$f = \frac{f_0}{1 + cT(v - \bar{v})(1 - P)} . \tag{10}$$

This solution reduces to the desired speed distribution function f_0 in the limit $c \to 0$. If $1 - cT\bar{v}(1 - P) < 0$, we have no solution, because the distribution function cannot be negative.

In the case

$$1 - cT\bar{v}(1 - P) = 0 , \tag{11}$$

f is given by

$$f = \frac{f_0}{cTv(1 - P)} . \tag{12}$$

The important feature of Eq. (12) is that it admits the deterministic solution

$$f = \alpha c\delta(v) , \tag{13}$$

where α is an arbitrary constant and $\delta(v)$ is the Dirac delta function.

As a result, we have the following general form of the stationary solution:

$$f = \frac{f_0}{1 + cT(v - \bar{v})(1 - P)} + \alpha c\delta(v) , \tag{14}$$

where α is unknown. However, for case (9) the singularity occurs for a negative value of v and we have, therefore, to take $\alpha = 0$. On the contrary, if we could have $1 - cT\bar{v}(1 - P) < 0$, then the singularity would occur for positive values of v, which is impossible.

Therefore, we have only two possibilities given by Eqs. (10) and (12). The solution given by Eq. (10) corresponds to what may be called the *individual flow pattern* and is related in a simple way to the ideal or desired speed distribution function. The second solution, given by Eq. (12), corresponds to what may be called the *collective flow pattern*. This solution is characterized by the occurrence of a singularity at the origin. However, the critical concentration at which the individual flow becomes collective does depend on the desired speed distribution function.

4. Choosing the desired speed distribution function

In Ref. [7] we demonstrated that due to the aggregation (with the bin size around 1s and larger) of the traffic measurements recorded at high frequency (such as each arriving packet is recorded independently), the shape of the packet size distribution is not changing with further increase of the aggregation bin; see Fig. 3.

The corresponding distributions are approximated with high accuracy by the log-normal distribution [8]

$$f(x) = \frac{A}{\sqrt{2\pi}\sigma}\,\frac{1}{x}\exp\left[-\frac{1}{2\sigma^2}(\log x-\mu)^2\right] , \tag{15}$$

where x is the variable, σ and μ are the parameters of log-normal distribution and A is the normalizing multiplier.

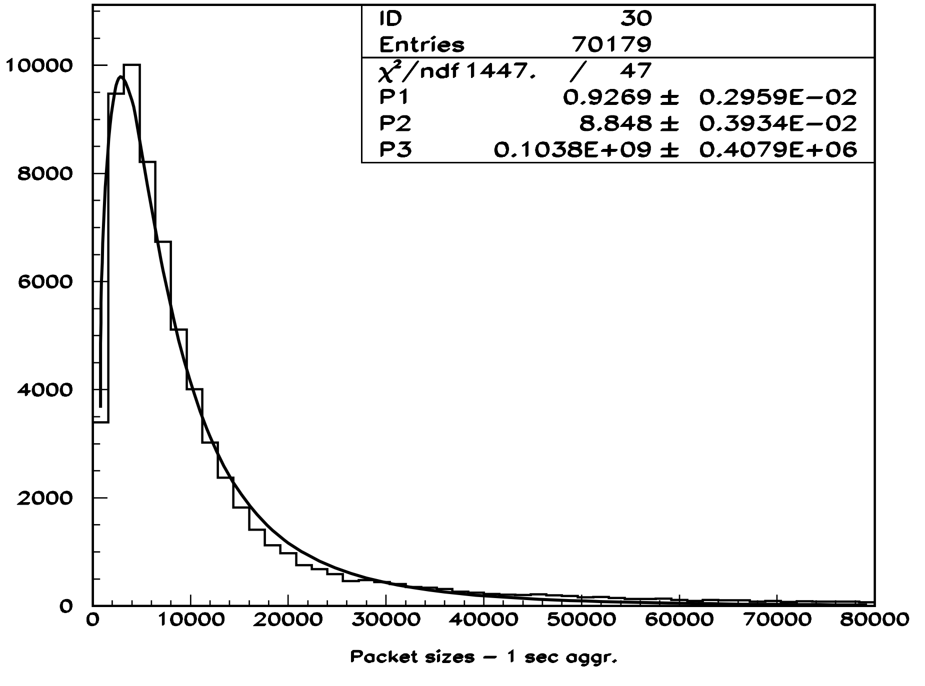

Fig. 3. Packet size distribution for traffic measurements aggregated with bin size 1 s: fitting curve corresponds to function (15).

Table 1
Results of fitting of the packet size distributions aggregated with different bin sizes by function (15)

Bin (s)	σ	μ	$A, \times 10^8$	v	χ^2/v	Critical level
1	0.9269 ± 0.003	8.85 ± 0.004	1.04 ± 0.005	47	30.8	0.966
2	0.8957 ± 0.004	9.62 ± 0.004	1.25 ± 0.007	47	21	> 0.999
3	0.9058 ± 0.005	10.06 ± 0.007	1.14 ± 0.008	47	14.2	> 0.999
4	0.8909 ± 0.006	10.40 ± 0.007	1.15 ± 0.009	47	14.2	> 0.999
5	0.8881 ± 0.006	10.62 ± 0.008	1.19 ± 0.01	47	7.4	> 0.999
7	0.7927 ± 0.008	11.03 ± 0.007	0.72 ± 0.008	47	15	> 0.999
10	0.8428 ± 0.007	11.36 ± 0.011	1.33 ± 0.016	46	4.9	> 0.999

In Table 1 we present the results of fitting the packet size distributions aggregated with different bin sizes with the log-normal distribution (15). The fitting was realized with the help of the MINUIT package [9] in the frame of the well-known Physical Analysis Workstation (PAW, for details see Ref. [10]). The MINUIT package is conceived as a tool to find the minimum value of a multi-parameter function and analyze the shape of the function around the minimum [9].

One can see that the fitting curves corresponding to the log-normal distribution approximate experimental distributions with a very high significance level (see below) and with a reliable accuracy on all regions of the analyzed distributions.

Here *Critical level* $= 1 - \alpha$, where α is the value of the significance level corresponding to the testing null-hypothesis H_0 (the log-normal distribution, in our case): see, for instance, Ref. [8].

From the point of view of further usage of the log-normal distributions presented above for different levels of aggregation, it is convenient to make the following change of variables $y = x/A$. In this case, instead of Eq. (15), we get

$$\hat{f}(y) = \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{y} \exp\left[-\frac{1}{2\sigma^2} (\log y - \mu')^2 \right] , \tag{16}$$

where $\mu' = \mu - \log A$, and parameter σ has the same value, as in Eq. (15); see also Table 1.

5. Analysis of solutions of the kinetic equation

The above-obtained packet size distribution function (16) can be considered as the desired velocity distribution function f_0.

Let us first analyze the case when concentrations are much smaller than the critical concentration c_{crit} (see Eq. (11)) given by

$$c_{crit} = [T\bar{v}(1 - P)]^{-1} \tag{17}$$

at which the transition to the collective flow occurs. Then we may obtain from (10), the first approximation to the desired speed distribution function

$$f = f_0[1 - Tc(1 - P)(v - \bar{v}^0)] \tag{18}$$

or

$$f = f_0[1 - \gamma(v - \bar{v}^0)] \tag{19}$$

with

$$\gamma = Tc(1 - P) . \tag{20}$$

Note that the value of $\bar{v}^0$ is obtained by using the desired speed distribution function (16).

Using expressions (7) and (8) we may rewrite (20) in the form

$$\gamma = c_P \tau \frac{\eta^3}{1 - \eta} . \tag{21}$$

One can see from (21) that for low-concentration situations, the value γ varies approximately as the third power of the concentration.

Using (19) we may obtain the form of the flow at low concentration

$$q = c[\bar{v}^0 - \gamma(\bar{v^2}^0 - (\bar{v}^0)^2)] . \tag{22}$$

This result shows that the first-order deviations from the linear portion of the flow curve are determined by the dispersion of the desired speed distribution function. For higher concentrations higher statistical moments contribute to this effect.

As for higher concentrations it appears impossible to express the distribution function f analytically, numerical analysis becomes necessary. The results of numerical computations are presented in Figs. 4 and 5.

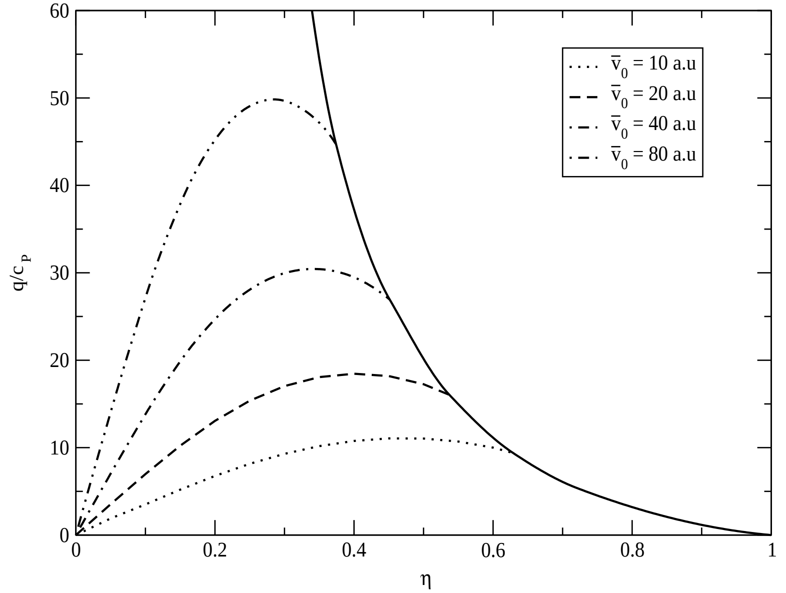

Fig. 4. Normalized flow q/c_p versus normalized concentration, $\eta = c/c_p$ for different average desired speeds $\bar{v}_0$ and $\tau c_p = 0.1$ a.u. for the log-normal distribution with $\sigma = 0.93$ and $\mu' = 0.83$.

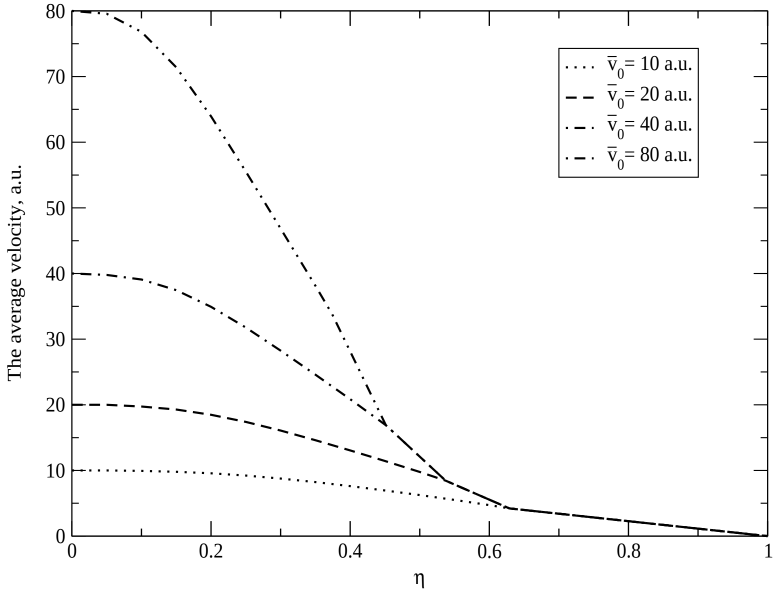

Fig. 5. The average velocity as a function of $\eta = c/c_p$ for different average desired speeds $\bar{v}_0$ and $\tau c_p = 0.1$ a.u. for the log-normal distribution with $\sigma = 0.93$ and $\mu' = 0.83$.

Fig. 4 shows the dependence of the normalized flow q/c_p against the normalized concentration η for the desired speed distribution function (16). The family of curves (dashed and dashed–point curves) are for the case $c_p\tau = 0.1$ a.u. and an average desired speed $\bar{v}^0$ ranging from 10 to 80 a.u. These curves demonstrate the individual flows for various values of $\bar{v}^0$. The linear dependence of the curves for low concentrations is clearly seen. We also see that the higher the average speed, the lower the concentration at which the optimum flow takes place. It must be noted also that for $\eta \sim 1$ there are no essential differences in the flow for different average desired speeds, whereas for lower η there are dramatic differences. The solid curve represents the collective flow curve.

It is quite interesting that the critical values of the normalized flow do not depend on the aggregation window. This may imply that the phase lines which have been obtained with the help of different values σ and μ at a given aggregation window (Table 1) have the same form.

In Fig. 5 we present the dependence of the average speed versus the normalized concentration for different average desired speeds.

Both Figs. 4 and 5 look very similar to what has been obtained by Prigogine and Herman [3] for the vehicular traffic.

6. Conclusion

We present here first results on the application of the kinetic approach to the network traffic. We obtained the solutions of the kinetic equation for homogeneous time-independent situations and for the chosen desired speed distribution function (16), as in Ref. [7] from traffic measurements analysis. For the log-normal form (16) of the desired function and for a given probability P of passing and given c_p and τ, the dependence of the normalized flow q/c_p versus the normalized concentration η clearly shows two modes corresponding to *individual flow patterns* (low-concentration mode) and to *collective flow patterns* (traffic jam mode). For low-concentration situations, we found that the normalized flow depends linearly on η and that for higher average speed, the concentration at which optimum flow takes place is lower. When approaching the critical concentration, there are no essential differences in the flow for various average desired speeds, whereas for lower η (corresponding to the *individual flow* region) there are dramatic differences.

The first results, Fig. 4, obtained for the chosen desired speed distribution function and the simplest forms of the probability P of passing and the relaxation time T demonstrated interesting behavior of information flows, which can be useful from the practical point of view. However, in order to establish a more close relation between the predictions given by the kinetic equation and real network traffic, more detailed study is needed, including the analysis of the relaxation time T, the form of probability P, etc.

Acknowledgements

We thank Yu. Kryukov for the preparation of traffic data. We are grateful to Prof. I. Prigogine and Prof. V.G. Kadyshevsky for encouragement and support.

This work has been done in the frame of the study contract no. 16263-2000-06 F1SC ISP BE between the International Solvay Institutes for Physics and Chemistry (Brussels, BELGIUM) and the Joint Research Center of the European Commission (Ispra, Italy).

References

[1] Tsunyi Tuan, Kihong Park, Multiple Time Scale Congestion Control for Self-Similar Network Traffic, Network Systems Lab, Department of Computer Sciences, Purdue University, West Lafayette, IN 47907, USA, Preprint submitted to Elsevier Preprint.

[2] S. Ata, M. Murata, H. Miyahara, Analysis of Network Traffic and Its Application to Design of High-Speed Routers, IEICE Trans. Inf. Syst. E83-D(5) (2000).

[3] I. Prigogine, R. Herman, Kinetic Theory of Vehicular Traffic, American Elsevier Publishing Company, Inc., New York, 1971.

[4] D.E. Wolf, M. Schreckenberg, A. Bachem (Eds.), Workshop on “Traffic and Granular Flow”, HLRZ, Forshungszentrum Jülich, Germany, 9–11 October 1995, Prologue by I. Prigogine, R. Herman; The New York Times, 25 November 1997, p. B11–B12.

[5] The State University “Dubna”: http://www.uni-dubna.ru.

[6] P.V. Vasiliev, V.V. Ivanov, V.V. Korenkov, Yu.A. Kryukov, S.I. Kuptsov, System for Acquisition, Analysis and Control of Network Traffic for the JINR Local Network Segment: the “Dubna” University Example, JINR Communications, D11-2001-266 Dubna, 2001.

[7] I. Antoniou, V.V. Ivanov, Valery V. Ivanov, P.V. Zrelov, On the lognormal distribution of network traffic, Physica D, submitted.

[8] W.T. Eadie, D. Dryard, F.E. James, M. Roos, B. Sadoulet, Statistical Methods in Experimental Physics, North-Holland Publishing Co., Amsterdam, London, 1971.
[9] F. James, MINUIT—Function Minimization and Error Analysis, CERN Program Library Long Write-up D506, 1998.
[10] R. Brun, O. Couet, C. Vandoni, P. Zanarini, PAW—Physics Analysis Workstation, CERN Program Library Q121, 1989.