



ΟΙΚΟΝΟΜΙΚΟ  
ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΑΘΗΝΩΝ  
ΒΙΒΛΙΟΘΗΚΗ  
810 39587  
Αρ. 004.6  
ταξ. ΜΑΚ

## ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΔΙΠΛΩΜΑ ΕΙΔΙΚΕΥΣΗΣ (MSc)  
στα ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ

### ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

«Μελέτη & Χαρακτηρισμός της κίνησης του Δικτύου  
Δεδομένων του Οικονομικού Πανεπιστημίου Αθηνών σε  
σχέση με το φαινόμενο της Αυτό-Ομοιότητας»

Μακρυδάκη Αικατερίνη

M3970020

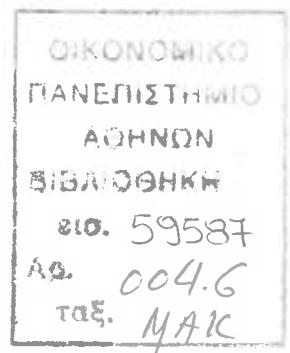
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ  
ΚΑΤΑΛΟΓΟΣ



ΑΘΗΝΑ, ΦΕΒΡΟΥΑΡΙΟΣ 1999

**ΜΕΤΑΠΤΥΧΙΑΚΟ ΔΙΠΛΩΜΑ ΕΙΔΙΚΕΥΣΗΣ (MSc)  
στα ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ**

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**



**«Μελέτη & Χαρακτηρισμός της κίνησης του Δικτύου  
Δεδομένων του Οικονομικού Πανεπιστημίου Αθηνών σε  
σχέση με το φαινόμενο της Αυτό-Ομοιότητας»**

**Μακρυδάκη Αικατερίνη**

**M3970020**



**Επιβλέπων Καθηγητής: Θεόδωρος Αποστολόπουλος**



ΠΟΡΤΑΚΑΣ ΠΑΝΕΠΙΣΤΗΜΟΥ

**ΜΕΛΕΤΗ ΚΑΙ ΧΑΡΑΚΤΗΡΙΣΜΟΣ ΤΗΣ ΚΙΝΗΣΗΣ ΤΟΥ ΔΙΚΤΥΟΥ ΔΕΔΟΜΕΝΩΝ ΤΟΥ  
ΟΙΚΟΝΟΜΙΚΟΥ ΠΑΝΕΠΙΣΤΗΜΙΟΥ ΑΘΗΝΩΝ ΣΕ ΣΧΕΣΗ ΜΕ ΤΟ ΦΑΙΝΟΜΕΝΟ ΤΗΣ  
ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ**

Διπλωματική εργασία

της

Αικατερίνης Ε. Μακρυδάκη

Υποβλήθηκε στα πλαίσια της εκπλήρωσης των απαιτήσεων για την λήψη του Μεταπτυχιακού Διπλώματος Ειδίκευσης (MSc) στα Πληροφοριακά Συστήματα του Οικονομικού Πανεπιστημίου Αθηνών.

Η εκπόνηση αυτής της διπλωματικής εργασίας πραγματοποιήθηκε με την καθοδήγηση του Καθηγητή Θ. Αποστολόπουλου.

Αθήνα, Φεβρουάριος 1999



## **ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ**

<b>ΚΕΦΑΛΑΙΟ 1.....</b>	<b>1</b>
<b>ΕΙΣΑΓΩΓΗ .....</b>	<b>1</b>
<b>ΚΕΦΑΛΑΙΟ 2.....</b>	<b>4</b>
<b>ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ ΔΙΑΧΕΙΡΙΣΗΣ ΥΠΟΛΟΓΙΣΤΩΝ .....</b>	<b>4</b>
1. ΤΟ ΠΡΩΤΟΚΟΛΛΟ SNMP .....	4
2. ΕΝΝΟΙΑ & ΟΡΓΑΝΩΣΗ ΤΩΝ MIBS .....	5
3. Ο ΜΗΧΑΝΙΣΜΟΣ RMON .....	8
<b>ΚΕΦΑΛΑΙΟ 3.....</b>	<b>10</b>
<b>ΠΕΡΙΓΡΑΦΗ ΤΟΥ ΦΑΙΝΟΜΕΝΟΥ ΤΗΣ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ.....</b>	<b>10</b>
1. Η ΕΝΝΟΙΑ ΤΗΣ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ .....	10
2. ΑΥΤΟ-ΟΜΟΙΑ ΔΕΔΟΜΕΝΑ ΚΙΝΗΣΗΣ .....	12
3. ΟΡΙΣΜΟΙ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ .....	16
3.1 Ορισμός αυτό-ομοιότητας για συνεχείς τιμές χρόνου .....	16
3.2 Ορισμός αυτό-ομοιότητας για διακριτές τιμές χρόνου .....	16
3.3 Εξάρτηση μακρού χρόνου ( <i>long-range dependence</i> ) .....	18
3.4 Heavy-tailed κατανομές .....	18
4. ΑΙΤΙΑ ΕΜΦΑΝΙΣΗΣ ΤΟΥ ΦΑΙΝΟΜΕΝΟΥ ΤΗΣ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ .....	20
5. ΠΑΡΑΔΕΙΓΜΑΤΑ ΑΥΤΟ-ΟΜΟΙΩΝ ΔΕΔΟΜΕΝΩΝ ΚΙΝΗΣΗΣ .....	21
5.1 Ethernet δεδομένα κίνησης .....	21
5.2 Κίνηση World Wide Web .....	21
5.3 Κίνηση TCP, FTP και TELNET .....	22
5.4 Video μεταβλητού ρυθμού bits .....	22
6. ΣΥΝΕΠΕΙΕΣ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ ΣΤΗΝ ΑΠΟΔΟΣΗ ΔΙΚΤΥΩΝ ΥΠΟΛΟΓΙΣΤΩΝ .....	23
6.1 Ανάλυση Ethernet & ISDN δεδομένων κίνησης .....	23
6.2 Απατήσεις χωρητικότητας .....	24
7. Η ΕΦΑΡΜΟΣΙΜΟΤΗΤΑ ΤΩΝ ΑΥΤΟ-ΟΜΟΙΩΝ ΜΟΝΤΕΛΩΝ ΚΙΝΗΣΗΣ .....	26
<b>ΚΕΦΑΛΑΙΟ 4.....</b>	<b>27</b>
<b>ΕΚΤΙΜΗΣΗ ΒΑΘΜΟΥ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ.....</b>	<b>27</b>
1. ΜΕΘΟΔΟΙ ΕΚΤΙΜΗΣΗΣ ΑΥΤΟ-ΟΜΟΙΩΝ ΔΕΔΟΜΕΝΩΝ ΚΙΝΗΣΗΣ .....	27
1.1 Μέθοδος Αθροιστικών Χρονολογικών Σειρών .....	27
1.2 R/S Ανάλυση .....	27
2. ΥΛΟΠΟΙΗΣΗ ΕΡΓΑΛΕΙΩΝ ΕΚΤΙΜΗΣΗΣ ΤΟΥ ΒΑΘΜΟΥ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ .....	29

2.1 Περιγραφή της διαδικασίας AggregatedSeries.....	30
2.2 Αποτελέσματα εκτέλεσης της διαδικασίας AggregatedSeries.....	32
2.3 Περιγραφή της διαδικασίας RS_Statistic.....	34
2.4 Αποτελέσματα εκτέλεσης της διαδικασίας RS_Statistic .....	36
2.5 Εύρος πιμών της παραμέτρου H (βαθμός αυτό-ομοιότητας).....	38
3. ΕΛΕΓΧΟΣ ΤΩΝ ΕΡΓΑΛΕΙΩΝ ΕΚΤΙΜΗΣΗ ΤΟΥ ΒΑΘΜΟΥ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ .....	40
4. ΕΠΙΔΡΑΣΗ ΤΟΥ ΕΠΙΠΕΔΟΥ ΑΘΡΟΙΣΗΣ ΣΤΙΣ ΕΚΤΙΜΗΣΕΙΣ ΤΟΥ ΒΑΘΜΟΥ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ.....	42
4.1 Σύγκριση αθροιστικού & πολ/κού τρόπου εξέλιξης των επιπέδων άθροισης.....	46
<b>ΚΕΦΑΛΑΙΟ 5.....</b>	<b>47</b>
<b>ΣΥΛΛΟΓΗ ETHERNET ΚΙΝΗΣΗΣ ΑΠΟ ΤΟ ΔΙΚΤΥΟ ΔΕΔΟΜΕΝΩΝ ΤΟΥ Ο.Π.Α. ....</b>	<b>47</b>
1. ΔΙΚΤΥΟ ΔΕΔΟΜΕΝΩΝ ΤΟΥ Ο.Π.Α.....	47
2. ΤΡΟΠΟΙ ΣΥΛΛΟΓΗΣ ΣΤΟΙΧΕΙΩΝ ΑΠΟ ΤΗΝ RMON MIB V1 (RFC1757).....	48
2.1 Συλλογή στοιχείων με χρήση του Ethernet LanProbe.....	49
2.2 Συλλογή στοιχείων μέσω εγκατάστασης αντιπροσώπων της RMON MIB v1 .....	51
3. ΠΕΡΙΓΡΑΦΗ ΧΡΗΣΙΜΟΠΟΙΟΥΜΕΝΩΝ RMON MIB V1 ΑΝΤΙΚΕΙΜΕΝΩΝ.....	52
4. ΔΙΑΔΙΚΑΣΙΑ ΣΥΛΛΟΓΗΣ ΣΤΟΙΧΕΙΩΝ ΑΠΟ ΤΗΝ RMON MIB V1 .....	53
<b>ΚΕΦΑΛΑΙΟ 6.....</b>	<b>54</b>
<b>ΕΚΤΙΜΗΣΗ ΒΑΘΜΟΥ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ ΕΤΗΕΡΝΕΤ ΚΙΝΗΣΗΣ ΤΟΥ ΔΙΚΤΥΟΥ ΔΕΔΟΜΕΝΩΝ ΤΟΥ Ο.Π.Α.....</b>	<b>54</b>
1. ΕΙΣΑΓΩΓΗ .....	54
2. ΓΕΝΙΚΗ ΣΥΝΟΨΗ ΠΕΙΡΑΜΑΤΩΝ .....	55
3. ΔΕΔΟΜΕΝΑ ΚΙΝΗΣΗΣ ΑΠΟ ΤΟ ΤΟΠΙΚΟ ΔΙΚΤΥΟ ΤΟΥ ΕΡΓΑΣΤΗΡΙΟΥ ΠΟΛΥΜΕΣΩΝ 56	56
3.1 Διεξαγωγή και Αποτελέσματα 1 <sup>ης</sup> Μέτρησης.....	56
3.2 Διεξαγωγή και Αποτελέσματα 2 <sup>ης</sup> Μέτρησης.....	65
3.3 Διεξαγωγή και Αποτελέσματα 3 <sup>ης</sup> Μέτρησης.....	72
4. ΔΕΔΟΜΕΝΑ ΚΙΝΗΣΗΣ ΑΠΟ ΤΟ ΤΟΠΙΚΟ ΔΙΚΤΥΟ ΤΟΥ ΚΕΝΤΡΟΥ ΔΙΑΧΕΙΡΙΣΗΣ ΔΙΚΤΥΩΝ .....	76
5. ΔΕΔΟΜΕΝΑ ΚΙΝΗΣΗΣ ΑΠΟ ΤΟ ΤΟΠΙΚΟ ΔΙΚΤΥΟ HELTRUN .....	83
6. ΔΕΔΟΜΕΝΑ ΚΙΝΗΣΗΣ ΑΠΟ ΤΟ ΤΟΠΙΚΟ ΔΙΚΤΥΟ ΤΟΥ ΚΕΝΤΡΟΥ ΥΠΟΛΟΓΙΣΤΩΝ ...	94
<b>ΚΕΦΑΛΑΙΟ 7.....</b>	<b>98</b>
<b>ΣΥΛΛΟΓΗ &amp; ΕΚΤΙΜΗΣΗ ΒΑΘΜΟΥ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ ΕΞΩΤΕΡΙΚΗΣ ΚΙΝΗΣΗΣ ΤΟΥ ΔΙΚΤΥΟΥ ΔΕΔΟΜΕΝΩΝ ΤΟΥ Ο.Π.Α.....</b>	<b>98</b>
1. ΕΙΣΑΓΩΓΗ .....	98
2. Η ΤΕΧΝΟΛΟΓΙΑ NETFLOW.....	98

3. Η ΕΦΑΡΜΟΓΗ ΣΥΛΛΟΓΗ Σ ΔΕΔΟΜΕΝΩΝ ΚΙΝΗΣΗΣ CFLOWD .....	100
4. Η ΕΦΑΡΜΟΓΗ ΣΥΛΛΟΓΗ Σ ΔΕΔΟΜΕΝΩΝ ΚΙΝΗΣΗΣ FLOWCOLLECTOR.....	106
5. ΔΙΑΔΙΚΑΣΙΑ ΣΥΛΛΟΓΗΣ ΔΕΔΟΜΕΝΩΝ ΕΞΩΤΕΡΙΚΗΣ ΚΙΝΗΣΗΣ ΑΠΟ ΤΟ ΔΙΚΤΥΟ ΔΕΔΟΜΕΝΩΝ ΤΟΥ Ο.Π.Α.....	108
6. ΔΕΔΟΜΕΝΑ ΕΞΩΤΕΡΙΚΗΣ ΚΙΝΗΣΗΣ ΕΙΣΟΔΟΥ ΤΟΥ Ο.Π.Α.....	110
6.1 Διεξαγωγή και Εκτιμήσεις 1 <sup>ης</sup> Μέτρησης.....	110
6.2 Διεξαγωγή και Εκτιμήσεις 2 <sup>ης</sup> Μέτρησης.....	111
6.3 Διεξαγωγή και Εκτιμήσεις 3 <sup>ης</sup> Μέτρησης.....	111
6.4 Σχολιασμός Αποτελεσμάτων Μετρήσεων.....	112
<b>ΚΕΦΑΛΑΙΟ 8.....</b>	<b>115</b>
<b>ΣΥΜΠΕΡΑΣΜΑΤΑ &amp; ΠΡΟΤΑΣΕΙΣ.....</b>	<b>115</b>
1. ΣΥΜΠΕΡΑΣΜΑΤΑ ΕΡΕΥΝΑΣ.....	115
2. ΠΡΟΤΑΣΕΙΣ ΜΕΛΛΟΝΤΙΚΗΣ ΕΡΕΥΝΑΣ.....	116
<b>ΠΑΡΑΡΤΗΜΑ 1.....</b>	<b>117</b>
<b>ΛΕΙΤΟΥΡΓΙΚΗ ΠΕΡΙΓΡΑΦΗ ΤΗΣ ΣΥΣΚΕΥΗΣ HP J3458A FAST ETHERNET LANPROBE .....</b>	<b>117</b>
<b>ΠΑΡΑΡΤΗΜΑ 2.....</b>	<b>121</b>
<b>Ο ΚΩΔΙΚΑΣ ΤΗΣ ΥΛΟΠΟΙΗΣΗΣ ΤΩΝ ΜΕΘΟΔΩΝ ΕΚΤΙΜΗΣΗΣ ΤΟΥ ΒΑΘΜΟΥ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ ΤΩΝ ΑΘΡΟΙΣΤΙΚΩΝ ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ &amp; ΤΗΣ R/S ΣΤΑΤΙΣΤΙΚΗΣ .....</b>	<b>121</b>
<b>ΠΑΡΑΡΤΗΜΑ 3.....</b>	<b>130</b>
<b>Ο ΚΩΔΙΚΑΣ ΤΗΣ ΜΑΚΡΟΕΝΤΟΛΗΣ ΔΗΜΙΟΥΡΓΙΑΣ ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ ΑΠΟ ΤΑ ΔΕΔΟΜΕΝΑ ΚΙΝΗΣΗΣ ΤΗΣ ΕΦΑΡΜΟΓΗΣ CFLOWD... 130</b>	
<b>ΠΑΡΑΡΤΗΜΑ 4.....</b>	<b>133</b>
<b>Ο ΚΩΔΙΚΑΣ ΤΗΣ ΜΑΚΡΟΕΝΤΟΛΗΣ ΔΗΜΙΟΥΡΓΙΑΣ ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ ΑΠΟ ΤΑ ΔΕΔΟΜΕΝΑ ΚΙΝΗΣΗΣ ΤΗΣ ΕΦΑΡΜΟΓΗΣ FLOWCOLLECTOR.....</b>	<b>133</b>
<b>ΒΙΒΛΙΟΓΡΑΦΙΑ.....</b>	<b>136</b>

## ΠΡΟΛΟΓΟΣ

Στην αρχή αυτής της διπλωματικής εργασίας θα ήθελα να ευχαριστήσω όλους όσους με βοήθησαν για να την ολοκληρώσω.

Πρώτα απ' όλα θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή Θεόδωρο Αποστολόπουλο για τη σημαντική καθοδήγησή του και τις γνώσεις που μου προσέφερε κατά τη φοίτησή μου στο Οικονομικό Πανεπιστήμιο Αθηνών.

Επίσης, θα ήθελα να ευχαριστήσω τον Κώστα Κουσουρή, τον Γιώργο Αλεξανδρή, τη Βικτωρία Δασκάλου και τον Αλέξη Ζάβρα για την βοήθεια τους στα τεχνικά προβλήματα που αντιμετώπισα, καθώς και τον συμφοιτητή και φίλο μου Αλέξη Τζωρμπατζάκη για την ηθική συμπαράσταση και συνεργασία του.

Τέλος, πάνω απ' όλους, θέλω να ευχαριστήσω τους γονείς μου και τον Γιάννη για την ουσιαστική –από κάθε πλευρά- υποστήριξη και κατανόησή τους.

Αικατερίνη Ε. Μακρυδάκη  
Φεβρουάριος 1999

## EXECUTIVE SUMMARY

The Computer and Communications Systems Laboratory at the Department of Informatics of the Athens University of Economics and Business (AUEB) conducted research which includes real traffic data collection of the University data network. The goal of the research is the comprehension of the IP traffic features concerning the concept of self-similarity.

### INTRODUCTION

In recent years, a number of studies have demonstrated that the traffic pattern for many network environments is self similar. A phenomenon that is self-similar looks the same or behaves the same when viewed at different degrees of "magnification" or different scales on a dimension. The dimension can be space (length, width) or time. In this study, we are concerned with IP traffic time series that exhibit self-similarity with respect to time. Many relevant analysis provide evidence that the concept of self-similarity must be taken into account in the process of network capacity planning. The consequence of self-similarity in buffer-size is summarized as follows: On high level of network utilization , drastically larger buffers are required for self-similar traffic [NORR94]. Furthermore the analysis in [LELA94] shows that if the input is self-similar, then increased delays and increased buffer size requirements will be experienced in any multiplexing of self-similar streams. This applies to switches, such as ATM, frame relay, and 100BASE-T switches; to WAN routers; to shared-medium LANs, such as Ethernet; and to statistical multiplexers.

### DEFINITION

A self-similar time series has the property that when aggregated (leading to a shorter time series in which each point is the sum of multiple original points) the new series has the same autocorrelation function as the original. That is, given a stationary time series  $X = (X_t; t = 0, 1, 2, \dots)$ , we define the  $m$ -aggregated series  $X^{(m)} = (X_k^{(m)}; k = 1, 2, 3, \dots)$  by summing the original series  $X$  over non overlapping blocks of size  $m$ . Then if  $X$  is self-similar, it has the same autocorrelation function  $r(k) = E[(X_t - m)(X_{t+k} - m)]$  as the series  $X^{(m)}$  for all  $m$ . Note that this means that the series is distributionally self-similar: the distribution of the aggregated series is the same (except for changes in scale) as that of the original.

As a result, self-similar processes show *long-range dependence*. A process with long-range dependence has an autocorrelation function that behaves as  $r(k)^{-\beta}$  as  $k \rightarrow \infty$ , where  $0 < \beta < 1$ . Thus the autocorrelation function of such a process decays hyperbolically (as compared to the exponential decay exhibited by traditional traffic models). Hyperbolic decay is much slower than exponential decay, and since  $\beta < 1$ , the sum of the autocorrelation values of such a series approaches infinity.



Self-similar stochastic processes can be defined with *heavy-tailed* distributions. The distribution of a random variable  $X$  is said to be heavy tailed if  $1 - F(x) = \Pr[X > x] \sim \frac{1}{x^\alpha}$  as  $x \rightarrow \infty$ ,  $0 < \alpha$ . In general, a random variable with a heavy-tailed distribution exhibits a high or even infinite variance. The term *heavy tail* is derived from the tail of the distributions which decays more slowly than exponential. The simplest heavy-tailed distribution is the Pareto distribution.

One of the attractive features of using self-similar models for time series, when appropriate, is that the degree of self-similarity of a series is expressed using only a single parameter. The parameter expresses the speed of decay of the series autocorrelation function. For historical reasons, the parameter used is the *Hurst* parameter  $H=1-\beta/2$ . Thus, for self-similar series,  $\frac{1}{2} < H < 1$ . As  $H \rightarrow 1$ , the degree of self-similarity increases. Thus the fundamental test for self-similarity of a series reduces to the question of whether  $H$  is significantly different from  $\frac{1}{2}$ .

## STATISTICAL TESTS FOR SELF-SIMILARITY

In this study we use two methods to test for self-similarity. These methods are heuristic and should not be used to obtain point estimates of  $H$ , but to get a rough idea whether a given data set is consistent with self-similar features ( $H > 0.5$ ). The first method, the *variance-time plot*, relies on the slowly decaying variance of a self-similar series. The variance of  $X^{(m)}$  is plotted against  $m$  on a log-log plot; a straight line with slope  $\beta$  greater than  $-1$  is indicative of self-similarity. The parameter  $H$  is given by  $H=1-\beta/2$ . The second method, the *R/S plot* uses the fact that for a self-similar data set, the *rescaled range* of *R/S* statistic grows according to a power law with exponent  $H$  as a function of the number of points included ( $m$ ). For a stochastic process  $X = (X_t : t = 0, 1, 2, \dots)$  the rescaled range of  $X$  over a time interval  $m$  is defined

$$\text{as the ratio } R/S: \frac{R}{S} = \frac{\max_{1 \leq j \leq m} \left[ \sum_{k=1}^j (X_k - M(m)) \right] - \min_{1 \leq j \leq m} \left[ \sum_{k=1}^j (X_k - M(m)) \right]}{\sqrt{\frac{1}{m} \sum_{j=1}^m (X_j - M(m))^2}}, \text{ where } M(m) \text{ is}$$

the sample mean over the time period  $m$ :  $M(m) = \frac{1}{m} \sum_{j=1}^m X_j$ . The numerator in this

ratio is a measure of the range of the process and the denominator is the sample standard deviation. For a self-similar process, the ratio has the following characteristic for large  $m$ :  $R/S \sim (m/2)^H$  with  $H > 0.5$ . This can be rewritten as  $\log[R/S] \sim H \log(m) - H \log(2)$ . If we plot  $\log[R/S]$  versus  $m$  on a log-log graph, the result should fit a straight line with slope  $H$ . Again, this analysis has been performed on a number of data sets with results that fit a straight line.

## DATA COLLECTION SCHEME

The LANs traffic monitoring is based on the RMON (Remote Network Monitoring) mechanism. An RMON agent monitors all the packets which are sent or received through the monitoring network, maintaining statistical data for them. The agent could be implemented in a PC connected to the network, in a probe, or in another network device.

The communication between the RMON agent and the management application is accomplished using SNMP (Simple Network Management Protocol). SNMP allows the management information interchange between the management application (client) and the RMON agent (server), and includes operations for this purpose: retrieval and manipulation of management information and reporting extraordinary events.

The Hewlett Packard LANprobe J3458A device is used for the Ethernet LAN traffic data collection. This device includes implementations of the RMON MIB v1 (RFC 1757) and v2 (RFC 2021) agents.

The possible ways to collect Ethernet traffic using the LAN probe in the University data network environment are described as follows:

1. *Connecting LAN probe with ATM-to-Ethernet switch.* In this situation the traffic concerning the communication of the VLAN's computer devices as well as the traffic concerning the communication from every computer device with the ATM link are gathered.
2. *Connecting LAN probe with Ethernet Hub, whose the default router is connected to FDDI-to-Ethernet switch or to FDDI network.* In this situation the traffic the traffic concerning the communication among the LAN's computer devices communication as well as the traffic concerning the communication between each computer device with the default router are gathered.
3. *Connecting LAN probe to router as well as to FDDI-to-Ethernet switch which in turn is connected to FDDI network.* In this situation the traffic concerning the communication between the router and the switch is gathered.

The measurement technology used for external traffic collection follows the client-server model: A data collection service, called NetFlow, was enabled in one of the Cisco router interface which connects the University data network with the outside world. Furthermore, a UNIX application was configured to consume and filter the traffic data received from the NetFlow-enabled Cisco router.



## MEASUREMENTS DESCRIPTION

Totally, 6 measurements were held in 4 LANs of the University data network. The RMON MIB v1 (RFC 1757) objects for which data was collected concern the total number of packets and bytes which sent and received from each network host, as well as the total number of packets and bytes received on the network.

The number of the time series observations was between 2000 and 10000 traces. For most of the time series the observation collection rate was defined per 5 seconds, while for the rest ones per 10 seconds.

The implementation of the *Variance-Time plot* and *R/S Statistic* methods for estimating the degree of self-similarity allowed as to get a rough idea of whether our measured data exhibit self-similarity.

For every measurement held the *Hurst* parameter was estimated for the time series of:

- Input packets and bytes received from the most active hosts.
- Output packets and bytes sent from the most active hosts.
- Total packets and bytes received on the network.

## MAIN FINDINGS

According to the estimated degree of self-similarity of the resulting time series, there is a clear indication for the existence of self-similarity in the Ethernet traffic data. This indication becomes stronger as the network utilization increases. In addition, it was found that as far as the estimation of the degree of self-similarity is concerned, it is careless if the traffic data are measured in packets or in bytes.

As far as the external traffic is concerned, it was estimated that the data were not consistent with self-similarity. However, the degree of self-similarity of the Web traffic data was higher than this of the total traffic.

## ΚΕΦΑΛΑΙΟ 1

### ΕΙΣΑΓΩΓΗ

Στα πλαίσια της παρούσας μελέτης διεξάγεται έρευνα η οποία περιλαμβάνει τη συλλογή πραγματικών δεδομένων κίνησης του δικτύου δεδομένων του Οικονομικού Πανεπιστημίου Αθηνών και αφορά στη κατανόηση των χαρακτηριστικών της κίνησης του δικτύου δεδομένων.

Ο έλεγχος της κίνησης των τοπικών δικτύων βασίζεται στον μηχανισμό RMON (Remote Network Monitoring). Ένας RMON αντιπρόσωπος (agent) καταγράφει όλα τα πακέτα που διακινούνται στο υπό έλεγχο δίκτυο, διατηρώντας στατιστικά στοιχεία. Υλοποιήσεις RMON αντιπροσώπων περιέχονται σε προσωπικούς υπολογιστές οι οποίοι συνδέονται στο δίκτυο ή σε άλλες δικτυακές συσκευές.

Η επικοινωνία μεταξύ των RMON αντιπροσώπων και της εφαρμογής διαχείρισης (management application) διεξάγεται χρησιμοποιώντας το πρωτόκολλο SNMP (Simple Network Management Protocol). Το SNMP επιτρέπει την ανταλλαγή πληροφοριών διαχείρισης μεταξύ της εφαρμογής διαχείρισης και του RMON αντιπροσώπου.

Η συσκευή LANprobe J3458A της Hewlett Packard χρησιμοποιείται για τη συλλογή δεδομένων κίνησης Ethernet τοπικών δικτύων. Αυτή περιλαμβάνει υλοποιήσεις αντιπροσώπων RMON MIB v1 (RFC 1757) και v2 (RFC 2021). Επίσης, για τον ίδιο σκοπό, έχουν εγκατασταθεί υλοποιήσεις RMON v1 αντιπροσώπων σε FDDI-to Etherent μεταγωγείς δικτύων. Για την συλλογή εξωτερικών δεδομένων κίνησης χρησιμοποιούνται στοιχεία από τα log αρχεία των Cisco συσκευών του δικτύου.

Τα σύνολα των δεδομένων κίνησης που συλλέγονται χρησιμοποιούνται για τον εντοπισμό των χαρακτηριστικών της κίνησης. Ιδιαίτερη έμφαση δίνεται στο να διαπιστωθεί εάν η IP κίνηση είναι στατιστικά αυτό-όμοια (self-similar) ή όχι. Το βασικό χαρακτηριστικό των αυτό-όμοιων διαδικασιών είναι ότι μοιάζουν ίδιες σε διαφορετικές κλίμακες (χρόνου ή χώρου). Αυτό το χαρακτηριστικό της ανεξαρτησίας από κλίμακες (αυτό-ομοιότητα) δεν συναντάται στα συμβατικά στοχαστικά μοντέλα κίνησης πακέτων, αλλά ούτε και στη τηλεφωνική κίνηση.

Ο βαθμός της αυτό-ομοιότητας εκφράζεται μέσω της παραμέτρου του Hurst. Η εκτίμηση του βαθμού της αυτό-ομοιότητας προσεγγίζεται από δύο διαφορετικές οπτικές γωνίες: (1) ανάλυση βασισμένη στην R/S-στατιστική, (2) ανάλυση των διακυμάνσεων των αθροιστικών διαδικασιών  $X^{(n)}$ .

Συμπερασματικά, η κατανόηση των χαρακτηριστικών κίνησης των δικτύων δεδομένων αποτελεί τον βασικό στόχο της παρούσας μελέτης. Τέτοιες προσπάθειες τείνουν να υποστηρίζουν τη λήψη αποφάσεων σχετικά με την καταλληλότητα των μοντέλων κίνησης, αλλά και να προσεγγίσουν την επίλυση προβλημάτων που σχετίζονται με τον οικονομικό σχεδιασμό, έλεγχο και απόδοση των μελλοντικών δικτύων.

Η παρούσα διπλωματική εργασία εντάσσεται στο ευρύτερο γνωστικό αντικείμενο της Διαχείρισης Δικτύων Υπολογιστών.

Οι κύριοι στόχοι της διπλωματικής εργασίας είναι:

- Η μελέτη του φαινομένου της αυτό-ομοιότητας σε σχέση με τα δεδομένα κίνησης δικτύων υπολογιστών.
- Η εξαγωγή συμπερασμάτων όσον αφορά το χαρακτηριστικό της αυτό-ομοιότητας πραγματικών δεδομένων κίνησης.

Η διπλωματική εργασία δομείται από τα ακόλουθα κεφάλαια:

### *Κεφάλαιο 1 – Εισαγωγή*

Στο κεφάλαιο αυτό παρατίθεται μια περίληψη της εργασίας, ενώ ορίζονται οι αντικειμενικοί της στόχοι.

### *Κεφάλαιο 2 – Βασικές Έννοιες Διαχείρισης Δικτύων Υπολογιστών*

Στο κεφάλαιο αυτό ορίζονται μερικές από τις πιο βασικές έννοιες διαχείρισης δικτύων υπολογιστών με έμφαση σε αυτές που θα χρησιμεύσουν στην πραγματοποίηση των στόχων της εργασίας.

### *Κεφάλαιο 3 – Περιγραφή του Φαινομένου της Αυτό-Ομοιότητας*

Στο κεφάλαιο αυτό επιχειρείται η παρουσίαση του φαινομένου της αυτό-ομοιότητας, παρατίθενται οι μαθηματικοί ορισμοί και οι ιδιότητες του φαινομένου, ενώ παρουσιάζεται μια προσπάθεια εντοπισμού των επιδράσεων του φαινομένου στην απόδοση (performance) των δικτύων υπολογιστών.

### *Κεφάλαιο 4 – Εκτίμηση Βαθμού Αυτό-Ομοιότητας*

Στο κεφάλαιο αυτό παρουσιάζονται μερικές από τις πιο κύριες μεθόδους εκτίμησης του βαθμού αυτό-ομοιότητας μιας χρονολογικής σειράς, ενώ περιγράφεται η υλοποίηση και τα εξαγόμενα τους.

### *Κεφάλαιο 5 – Συλλογή Ethernet Κίνησης από το Δίκτυο Δεδομένων του Ο.Π.Α.*

Στο κεφάλαιο αυτό περιγράφεται η διαδικασία συλλογής πραγματικών δεδομένων Ethernet κίνησης από τοπικά υποδίκτυα του δικτύου δεδομένων του Πανεπιστημίου.

## Κεφάλαιο 6 – Εκτίμηση Βαθμού Αυτό-Ομοιότητας Ethernet Κίνησης του Δικτύου Δεδομένων του Ο.Π.Α.

Στο κεφάλαιο αυτό παρατίθενται οι εκτιμήσεις του βαθμού αυτό-ομοιότητας των δεδομένων της Ethernet κίνησης που συλλέχθηκαν, ενώ γίνεται μια προσπάθεια χαρακτηρισμού της συμπεριφοράς της κίνησης σε σχέση με το φαινόμενο της αυτό-ομοιότητας.

## Κεφάλαιο 7 – Συλλογή & Εκτίμηση Βαθμού Αυτό-Ομοιότητας Εξωτερικής Κίνησης του Δικτύου Δεδομένων του Ο.Π.Α.

Στο κεφάλαιο αυτό περιγράφονται οι βασικές τεχνολογίες λήψης δεδομένων κίνησης από τον δρομολογητή σύνδεσης του δικτύου δεδομένων του Πανεπιστημίου με το Internet, ενώ παρουσιάζονται τα αποτελέσματα των εκτιμήσεων του βαθμού αυτό-ομοιότητας των μετρήσεων που πραγματοποιήθηκαν.

## Κεφάλαιο 8 – Συμπεράσματα & Προτάσεις.

Στο κεφάλαιο αυτό συνοψίζονται τα συμπεράσματα της έρευνας που πραγματοποιήθηκε και προτείνονται θέματα για περαιτέρω έρευνα.

### Παράρτημα 1 – Λειτουργική Περιγραφή της Συσκευής HP J345A Fast Ethernet Lan Probe.

Στο παράρτημα αυτό παρατίθεται μια συνοπτική περιγραφή των λειτουργιών της συσκευής συλλογής και διατήρησης Ethernet στατιστικών στοιχείων κίνησης.

### Παράρτημα 2 – Ο Κώδικας της Υλοποίησης των Μεθόδων Εκτίμησης του Βαθμού Αυτό-Ομοιότητας των Αθροιστικών Χρονολογικών Σειρών & της R/S Στατιστικής

Στο παράρτημα αυτό παρατίθεται ο κώδικας των εργαλείων εκτίμησης του βαθμού αυτό-ομοιότητας που υλοποιήθηκαν στα πλαίσια της παρούσας εργασίας.

### Παράρτημα 3 – Ο Κώδικας της Μακροεντολής δημιουργίας Χρονολογικών Σειρών από τα Δεδομένα Κίνησης της εφαρμογής CFLWD

Στο παράρτημα αυτό παρατίθεται ο κώδικας της εφαρμογής μετατροπής των στοιχείων κίνησης που παρέχονται μέσω της εφαρμογής cflowd σε χρονολογικά στοιχεία κίνησης.

### Παράρτημα 4 – Ο Κώδικας της Μακροεντολής δημιουργίας Χρονολογικών Σειρών από τα Δεδομένα Κίνησης της εφαρμογής FLOWCOLLECTOR

Στο παράρτημα αυτό παρατίθεται ο κώδικας της εφαρμογής μετατροπής των στοιχείων κίνησης που παρέχονται μέσω της εφαρμογής flowcollector σε χρονολογικά στοιχεία κίνησης.

## Βιβλιογραφία

Στο παράρτημα αυτό παρατίθεται η χρησιμοποιούμενη βιβλιογραφία.

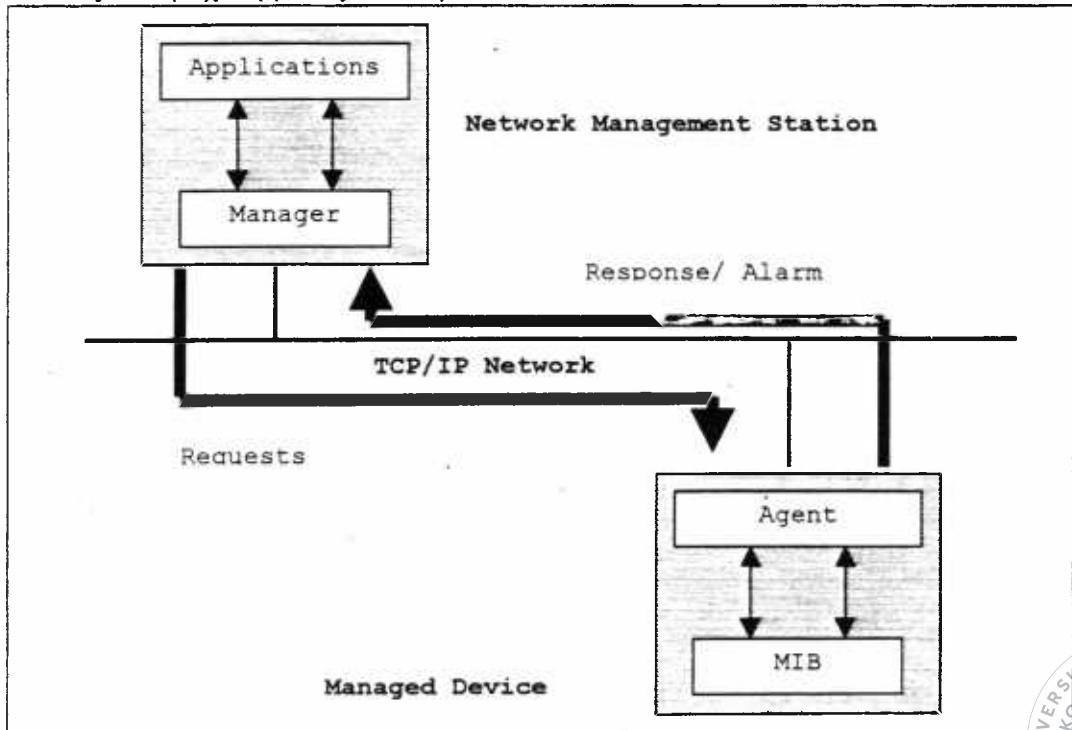


## ΚΕΦΑΛΑΙΟ 2

### ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ ΔΙΑΧΕΙΡΙΣΗΣ ΔΙΚΤΥΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

#### 1. ΤΟ ΠΡΩΤΟΚΟΛΛΟ SNMP (SIMPLE NETWORK MANAGEMENT PROTOCOL)

Ένα σύστημα διαχείρισης δικτύου βασισμένο στο πρωτόκολλο SNMP περιλαμβάνει τρία μέρη: μια εφαρμογή διαχείρισης (*manager*), έναν αντιπρόσωπο (*agent*) και μια βάση δεδομένων (*Management Information Base- MIB*). Το SNMP είναι ένα πρωτόκολλο ανταλλαγής πληροφοριών διαχείρισης μεταξύ της εφαρμογής διαχείρισης και των αντιπροσώπων η οποία πραγματοποιείται μέσω μιας σχέσης πελάτη/ εξυπηρετητή. Η εφαρμογή διαχείρισης αποτελεί τον πελάτη, ενώ ο αντιπρόσωπος ο οποίος εκτελείται σε μια απομακρυσμένη συσκευή είναι ο εξυπηρετητής. Επίσης η βάση δεδομένων η οποία ελέγχεται από τον SNMP αντιπρόσωπο αποτελεί την SNMP MIB. Στο Σχήμα 2.1 απεικονίζεται η σχέση μεταξύ των τριών συστατικών του SNMP.



Σχήμα 2.1 Η σχέση μεταξύ των συστατικών του SNMP

**Εφαρμογή Διαχείρισης (Manager):** Η εφαρμογή διαχείρισης λειτουργεί σε έναν ή περισσότερους υπολογιστές. Μια ή περισσότερες εφαρμογές διαχείρισης μπορούν να ρυθμιστούν (configured) για τη διαχείριση ενός υποδικτύου. Η βασική λειτουργία των εφαρμογών διαχείρισης είναι η μεταφορά αιτήσεων (requests) σε μια ή περισσότερες διαχειρίσιμες συσκευές. Επού, αφού η εφαρμογή διαχείρισης πληροφορηθεί σχετικά με την κατάσταση των τοπικών αντικειμένων της συσκευής, έχει τη δυνατότητα πραγματοποίησης αλλαγών, εφ' όσον διαθέτει την κατάλληλη εξουσιοδότηση. Η αλληλεπίδραση μεταξύ ενός τελικού χρήστη και της εφαρμογής διαχείρισης, επιτυγχάνεται με τη χρήση μιας ή περισσότερων εφαρμογών, οι οποίες μαζί με την εφαρμογή διαχείρισης αλλά και την πλατφόρμα υλικού (hardware platform) αποτελούν το σταθμό διαχείρισης δικτύου (Network Management Station). Σήμερα, σχεδόν όλες οι εφαρμογές κάνονται χρήση των γραφικών διεπαφών χρήστη (Graphic User Interface), παρέχουν ένα φιλικό περιβάλλον το οποίο αλληλεπιδρά με την εφαρμογή διαχείρισης προκειμένου να παραχθούν γραφήματα και χάρτες τα οποία δίνουν μια οπτική περιγραφή των δραστηριοτήτων του δικτύου.

**Αντιπρόσωποι (Agents):** Κάθε διαχειρίσιμη συσκευή περιλαμβάνει λογισμικό το οποίο διερμηνεύει τις SNMP αιτήσεις και απαντάει σε αυτές. Αυτό το λογισμικό αποτελεί τον αντιπρόσωπο. Μια συσκευή προκειμένου να μπορεί να διαχειριστεί άμεσα βάσει του SNMP πρωτοκόλλου, πρέπει να διαθέτει έναν αντιπρόσωπο. Παρ' όλα αυτά, συσκευές οι οποίες είναι ασύμβατες με το SNMP πρωτόκολλο μπορούν να διαχειριστούν, με την προϋπόθεση ότι υποστηρίζουν κάποιο άλλο πρωτόκολλο διαχείρισης. Προκειμένου να επιτευχθεί αυτό, απαιτείται η χρήση ενός πληρεξούσιου αντιπροσώπου (proxy agent). Ο τελευταίος μπορεί να θεωρηθεί ως ένας μετατροπέας πρωτοκόλλων, αφού μεταφράζει τις SNMP αιτήσεις σε αιτήσεις του πρωτοκόλλου διαχείρισης το οποίο υποστηρίζεται από την συσκευή.

Κατά βάση στο SNMP πρωτόκολλο οι αιτήσεις δημιουργούνται από την εφαρμογή διαχείρισης και απαιτούνται από τον αντιπρόσωπο. Παρ' όλα αυτά, ο αντιπρόσωπος έχει την δυνατότητα να στείλει, χωρίς να έχει ερωτηθεί, μια απάντηση. Μια τέτοια απάντηση σηματοδοτεί την ικανοποίηση κάποιας προκαθορισμένης συνθήκης η οποία βάσει των ρυθμίσεων πρέπει να αναφερθεί στην εφαρμογή διαχείρισης. Τέτοιες περιπτώσεις συμβαίνουν όταν ένας κόμβος του δικτύου βρίσκεται σε συμφόρηση, όταν μια γραμμή «πέφτει» και μετά «σηκώνεται», όταν ένας διαχειρίσιμος κόμβος καταρρέει και επανεκκινείται (“crash and reboot”). Όταν συμβαίνουν τέτοια γεγονότα ο αντιπρόσωπος ειδοποιεί για το γεγονός όλους τους σταθμούς διαχείρισης δικτύου που βρίσκονται στη λίστα διάρθρωσής του (configuration list). Η ειδοποίηση αυτή υλοποιείται με ένα μήνυμα του πρωτοκόλλου SNMP που ονομάζεται παγίδα (SNMP trap).

**Βάση Διαχείρισης Πληροφοριών (MIB):** Κάθε διαχειρίσιμη συσκευή χαρακτηρίζεται από τις δικές τις ρυθμίσεις, τη δική της κατάσταση και στατιστικές πληροφορίες. Αυτές οι πληροφορίες προσδιορίζουν τις λειτουργικές δυνατότητες της συσκευής και μπορούν να περιλαμβάνουν ρυθμίσεις του υλικού, τιμές μεταβλητών αποθηκευμένες σε πίνακες, εγγραφές ή πεδία μεταβλητών τοποθετημένες σε αρχεία κλπ. Το σύνολο όλων αυτών των στοιχείων δεδομένων αναφέρονται ως η βάση των

πληροφοριών διαχείρισης της διαχειρίσιμης συσκευής. Κάθε στοιχειώδης μονάδα δεδομένων αναφέρεται ως διαχειρίσιμο αντικείμενο (managed object) και αποτελείται από ένα όνομα, μια ή περισσότερες ιδιότητες, καθώς και από ένα σύνολο πράξεων οι οποίες μπορούν να εκτελεστούν πάνω στο αντικείμενο. Μέσω της βάσης διαχείρισης πληροφοριών ορίζεται το είδος της πληροφορίας η οποία μπορεί να ανακτηθεί από την διαχειρίσιμη συσκευή καθώς και το είδος των ρυθμίσεων της συσκευής οι οποίες μπορούν να ελεγχθούν από το σύστημα διαχείρισης.

Το SNMP είναι ένα ασύγχρονο πρωτόκολλο αίτησης/ απάντησης. Κάθε οντότητα που επικοινωνεί με κάποια άλλη δεν χρειάζεται να περιμένει να έρθει η απάντηση σε προηγούμενη αίτηση πριν προβεί σε επόμενη ερώτηση ή εκτελέσει μια άλλη λειτουργία.

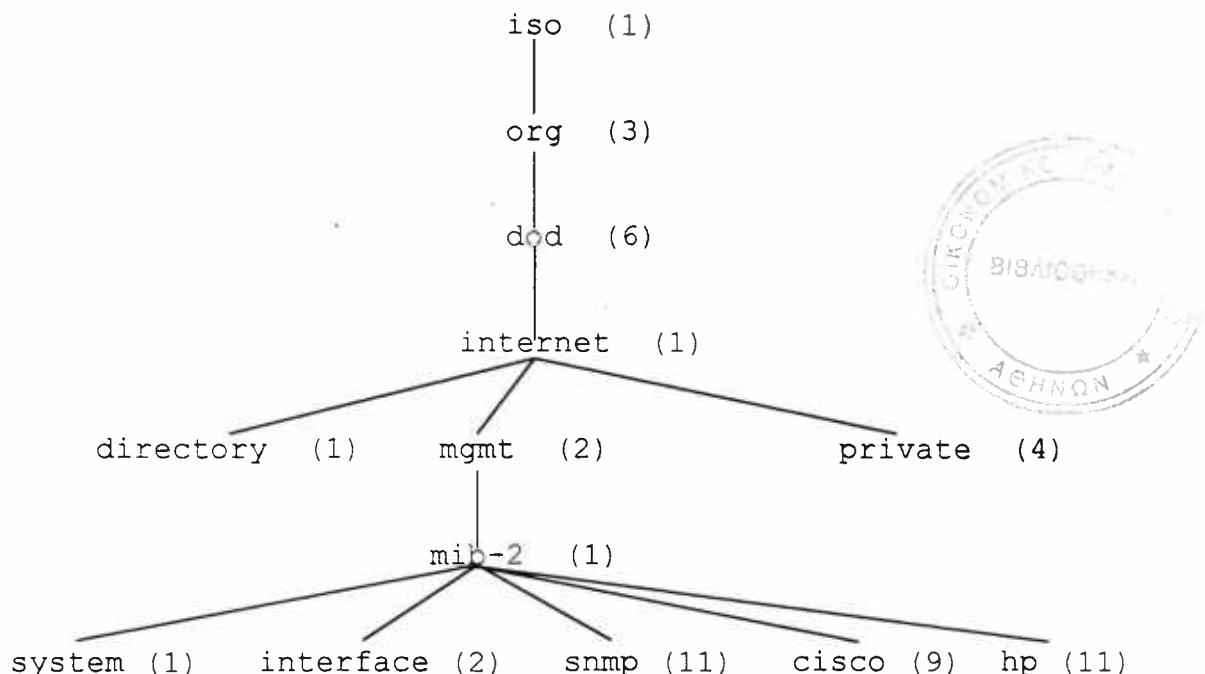
Υπάρχουν τέσσερις περιπτώσεις επικοινωνίας μεταξύ δύο SNMP οντοτήτων ή καλύτερα ενός σταθμού διαχείρισης δικτύου (Network Management Station – NMS) και ενός agent:

- Ο σταθμός διαχείρισης δικτύου επιθυμεί να ανακτήσει μια πληροφορία διαχείρισης από τη MIB του αντιπροσώπου. Σε αυτή την περίπτωση, ο σταθμός διαχείρισης στέλνει ένα μήνυμα *get-request* και ο αντιπρόσωπος απαντά με ένα μήνυμα *get-response*.
- Ο σταθμός διαχείρισης δικτύου θέλει να ανακτήσει ένα μέρος της MIB του agent. Σε αυτή την περίπτωση ο σταθμός διαχείρισης στέλνει ένα μήνυμα *get-next-request* και ο agent απαντά με ένα μήνυμα *get-response*.
- Ο σταθμός διαχείρισης δικτύου θέλει να αποθηκεύσει μια πληροφορία διαχείρισης στη MIB του agent. Σε αυτή την περίπτωση ο σταθμός διαχείρισης στέλνει ένα μήνυμα *set-request* και ο agent απαντά με ένα μήνυμα *get-response*.
- Ο agent θέλει να αναφέρει ένα απρόσμενο γεγονός και στέλνει ένα μήνυμα *trap* στον σταθμό.

## 2. ΕΝΝΟΙΑ & ΟΡΓΑΝΩΣΗ ΤΩΝ MIBS (MANAGEMENT INFORMATION BASES)

Τα περισσότερα δίκτυα περιλαμβάνουν συσκευές διαφορετικών κατασκευαστών. Για να μπορέσει ένας σταθμός διαχείρισης δικτύου να επικοινωνήσει με όλες αυτές τις τόσο διαφορετικές συσκευές, είναι αναγκαίο να τηρείται η πληροφορία που αφορά κάθε συσκευή σε μια ή περισσότερες μεταβλητές. Αυτές οι μεταβλητές ονομάζονται αντικείμενα (*objects*). Σε αντίθεση με τα αντικείμενα ενός αντικειμενοστραφούς συστήματος, τα αντικείμενα αυτά έχουν μόνο κατάσταση (*state*) και καμία μέθοδο (*method*) (εκτός από την ανάγνωση και την εγγραφή των τιμών τους). Όπως έχει ήδη αναφερθεί η συλλογή αυτών των αντικειμένων ονομάζεται Βάση Διαχείρισης Πληροφοριών – MIB (*Management Information Base*).

Τα αντικείμενα των Βάσεων Διαχείρισης Πληροφοριών (MIBs) οργανώνονται σε μια ιεραρχική δομή δέντρου. Κάθε κλάδος του δέντρου έχει ένα μοναδικό όνομα και ένα μοναδικό αριθμητικό αναγνωριστικό (identifier). Οι ενδιάμεσοι κλάδοι του δέντρου ομαδοποιούν σχετιζόμενα MIB αντικείμενα. Τα «φύλλα» του δέντρου αποτελούν τα πραγματικά MIB αντικείμενα. Ένα υποδέντρο χρησιμοποιείται προκειμένου να προσδιοριστεί μια ομάδα από κλάδους και φύλλα τα οποία βρίσκονται κάτω από ένα συγκεκριμένο ενδιάμεσο κλάδο. Στο Σχήμα 2.2 απεικονίζεται η δομή του δέντρου, δίνοντας ιδιαίτερη έμφαση στο Internet υποδέντρο.



Σχήμα 2.2 Δενδρική δομή οργάνωσης των Βάσεων Πληροφοριών Διαχείρισης (MIBs).

Ένα MIB αντικείμενο παίρνει το όνομά του και αναγνωρίζεται μοναδικά ανάλογα με τη θέση όπου βρίσκεται σε δέντρο. Ένα πλήρες αναγνωριστικό αντικειμένου αποτελείται από τα αναγνωριστικά όλων των κλάδων του μονοπατιού από την αρχή του δέντρου μέχρι το φύλλο. Ένα πλήρες αναγνωριστικό αντικειμένου εκφράζεται ξεχωρίζοντας κάθε αναγνωριστικό κλάδου του μονοπατιού με τελείες. Για παράδειγμα το υποδέντρο mib-2 είναι iso.org.dod.internet.mgmt.mib-2 και το ισοδύναμο αριθμητικό αναγνωριστικό είναι 1.3.6.1.2.1. Οι παραπάνω συμβολισμοί των MIB αντικειμένων ακολουθούν τον προτυποποιημένο συμβολισμό ο οποίος έχει οριστεί μέσω της Abstract Syntax Notation One (ASN.1).

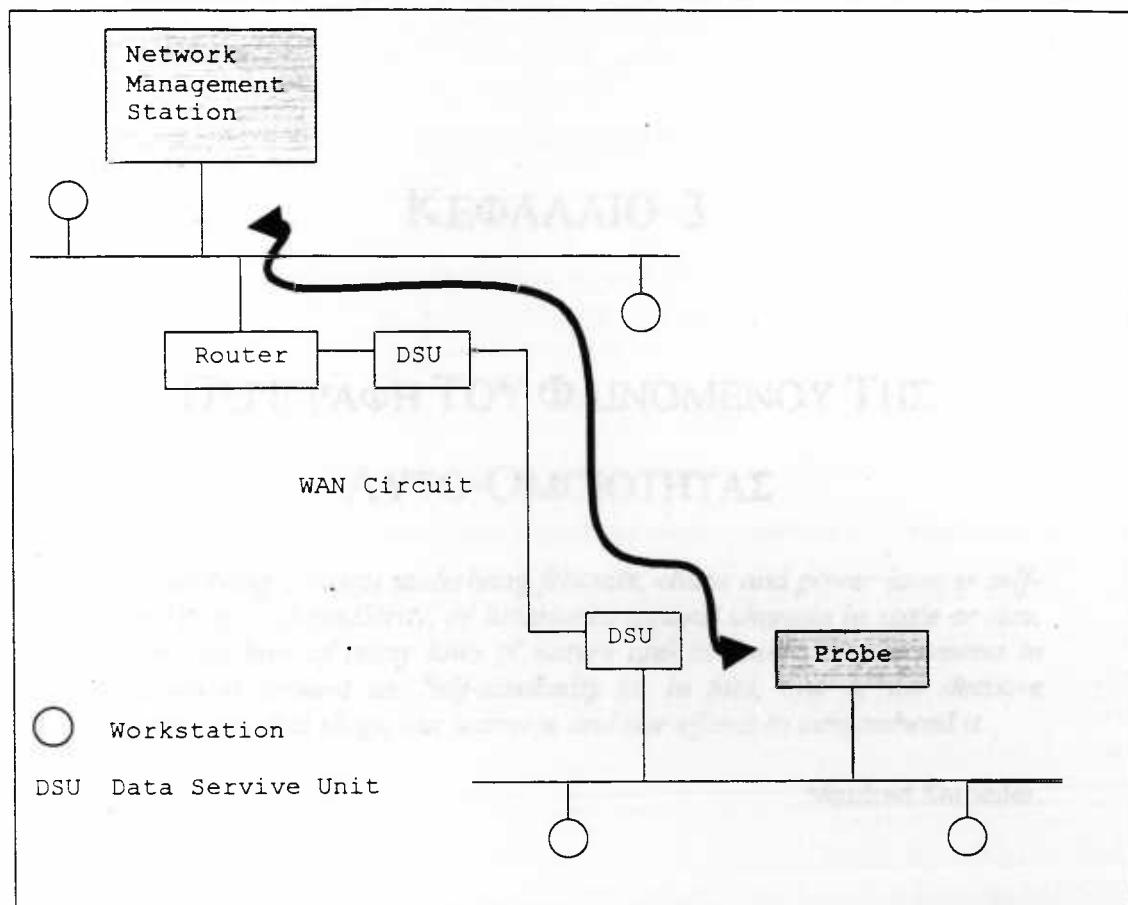
Ένας αντιπρόσωπος ή σύστημα αντιπροσώπου (agent system) μπορεί να έχει ένα ή περισσότερα στιγμιότυπα ενός συγκεκριμένου MIB αντικειμένου. Ένα συγκεκριμένο στιγμιότυπο διαχωρίζεται από άλλα στιγμιότυπα του ίδιου αντικειμένου από ένα αναγνωριστικό στιγμιότυπου (instance ID). Το αναγνωριστικό στιγμιότυπου προστίθεται

στο τέλος του αναγνωριστικού του MIB αντικειμένου. Για παράδειγμα ο συμβολισμός iso.org.dod.internet.mgmt.mib-2.interfaces.ifTable.ifEntry.ifInOctets.3 αναπαριστά την τιμή του αντικειμένου ifInOctets για το στιγμιότυπο 3 (το οποίο σε αυτή την περίπτωση είναι η διεπαφή 3). Για ορισμένα MIB αντικείμενα υπάρχει μόνο ένα στιγμιότυπο τους σε κάθε συσκευή, όπως για παράδειγμα συμβαίνει με το αντικείμενο system.sysDescriptor. Σε αυτές τις περιπτώσεις το αναγνωριστικό του στιγμιότυπου παίρνει την τιμή μηδέν.

### 3. ΜΗΧΑΝΙΣΜΟΣ RMON (REMOTE NETWORK MONITORING)

Η λειτουργία του SNMP πρωτοκόλου (διατύπωση ερωτήσεων και λήψη απαντήσεων) ενώ δεν παρουσιάζει μεγάλο επικοινωνιακό κόστος στην περίπτωση ενός τοπικού δικτύου (LAN), μπορεί να επιβαρύνει σημαντικά ένα δίκτυο ευρείας περιοχής (WAN) όταν ελέγχονται γεωγραφικά απομακρυσμένα δίκτυα. Αναγνωρίζοντας αυτό το πρόβλημα η ομάδα εργασίας ελέγχου απομακρυσμένων δικτύων (Remote Network Monitoring Working Group) ανέπτυξε το πρότυπο διαχείρισης ελέγχου απομακρυσμένων δικτύων (Remote Monitoring -RMON- network management standard).

Ο μηχανισμός RMON αποτελεί μια επέκταση της διαχείρισης δικτύων σε απομακρυσμένα δίκτυα. Σε αυτά, ευφυείς συσκευές γνωστές ως *probes* ή *RMON agents* ελέγχουν την κίνηση των δεδομένων και την οργανώνουν σε πληροφορίες τις οποίες η εφαρμογή διαχείρισης μπορεί εύκολα, χρησιμοποιώντας το πρωτόκολλο SNMP ως μηχανισμό μεταφοράς μεταξύ της εφαρμογής διαχείρισης και του αντιπροσώπου, να προσπελάσει και να διερμηνεύσει. Στο Σχήμα 2.3 απεικονίζεται η σχέση μεταξύ ενός σταθμού διαχείρισης δικτύου ο οποίος βρίσκεται σε ένα τοπικό δίκτυο και ο οποίος χρησιμοποιείται για τη διαχείριση ενός απομακρυσμένου τοπικού δικτύου μέσω ενός RMON agent ή ενός probe. Το απομακρυσμένο probe ελέγχει και οργανώνει τα δεδομένα της κίνησης του απομακρυσμένου δικτύου. Με αυτόν τον τρόπο μειώνεται το ποσό της πληροφορίας το οποίο θα απαιτούνταν (χωρίς τη χρήση RMON agent ή probe) να μεταφερθεί στο σταθμό διαχείρισης δικτύου προκειμένου να αναλυθεί.



Σχήμα 2.3 Επέκταση της διαχειριστικής δικτύου σε απομακρυσμένη περιοχή μέσω RMON probe.

## ΚΕΦΑΛΑΙΟ 3

### ΠΕΡΙΓΡΑΦΗ ΤΟΥ ΦΑΙΝΟΜΕΝΟΥ ΤΗΣ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ

*The unifying concept underlying fractals, chaos and power laws is self-similarity. Self-similarity, or invariance against changes in scale or size, is an attribute of many laws of nature and innumerable phenomena in the world around us. Self-similarity is, in fact, one of the decisive symmetries that shape our universe and our efforts to comprehend it.*

Manfred Shroeder

#### 1. Η ΕΝΝΟΙΑ ΤΗΣ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ

Ας υποτεθεί ότι ελέγχεται η κίνηση μιας frame relay γραμμής ταχύτητας 1 Mbps και ότι σταθερού μεγέθους πλαίσια (frames) των 4000 bits μεταφέρονται, έτσι ώστε ο χρόνος μεταφοράς ενός πλαισίου είναι 4 ms. Παρακάτω παρατίθενται οι χρόνοι άφιξης των πλαισίων, καταγεγραμμένοι από τον παραλήπτη (οι χρόνοι στους οποίους το πρώτο bit κάθε πλαισίου καταφθάνει στον παραλήπτη):

0	8	24	32	72	80	96	104	216	224	240	288
	296	312	320	648	656	672	680	720	728	744	752
	864	872	888	896	936	944	960	968			

Έτσι έχουμε ότι το πρώτο πλαισίο καταφθάνει τη χρονική στιγμή  $t=0\text{ ms}$ , το δεύτερο τη χρονική στιγμή  $t=8\text{ ms}$  κλπ.

Είναι δύσκολο να γίνει διάκριση κάποιου μοτίβου (προτύπου) ή στατιστικών ιδιοτήτων για τα παραπάνω δεδομένα. Παρ' όλα αυτά η κίνηση ομοιάζει με κίνηση κατά ριτές. Μερικοί χρόνοι άφιξης βρίσκονται πολύ κοντά μεταξύ τους, ενώ υπάρχουν και μεγάλα κενά διαστήματα. Το μεγαλύτερο κενό διάστημα είναι 328 ms από τη χρονική στιγμή 320 ως την 648, αλλά υπάρχουν και μικρότερα κενά διαστήματα των 40 ms (ισοδύναμο της μεταφοράς 10 πλαισίων). Ας υποθέσουμε ότι ομαδοποιούμε την κίνηση θεωρώντας



ως ομάδα ένα οποιοδήποτε σύνολο από αφίξεις πλαισίων μεταξύ των οποίων δεν υπάρχει κενό διάστημα μεγαλύτερο από 20 ms (χρόνος μεταφοράς πέντε πλαισίων). Προκύπτει:

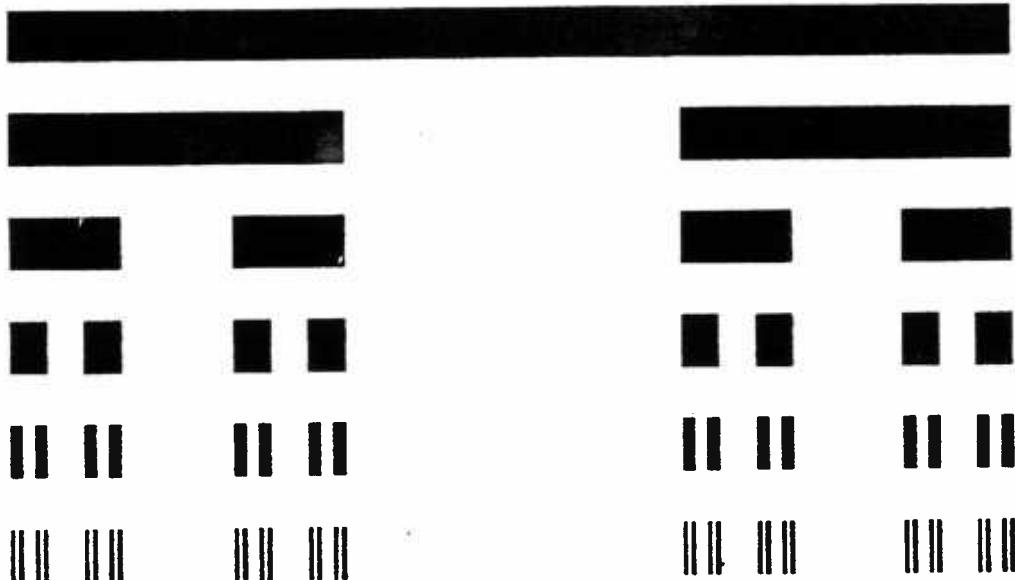
0      72      216      288      648      720      864      936

Τα κενά διαστήματα μεταξύ των ομάδων έχουν άνισο μήκος, όμως είναι ακόμα δύσκολο να εντοπιστεί κάποιο υπόδειγμα (pattern). Επιχειρώντας έναν μεγαλύτερο βαθμό ομαδοποίησης, ορίζουμε μια ομάδα ως ένα οποιοδήποτε σύνολο από πλαίσια μεταξύ των οποίων δεν υπάρχουν κενά διαστήματα μεγαλύτερα των 80 ms (ισοδύναμο της μεταφοράς 20 πλαισίων). Οι τιμές αφίξης που προκύπτουν είναι οι ακόλουθες:

0      216      648      864

Τα κενά διαστήματα είναι: 216, 432, 216. Το υπόδειγμα είναι: ομάδα, κενό διάστημα, ομάδα, μεγαλύτερο κενό διάστημα, ομάδα, κενό διάστημα. Παρατηρώντας την προηγούμενη ομαδοποίηση με τις έξι ομάδες, το υπόδειγμα επαληθεύεται. Οι τέσσερις πρώτοι χρόνοι άφιξης ακόλουθουν το πρότυπο (άφιξη, μικρό κενό, άφιξη, μεγάλο κενό, άφιξη, μικρό κενό, άφιξη). Το ίδιο ισχύει και για τους τέσσερις τελευταίους χρόνους άφιξης. Επιπλέον, για τα αρχικά δεδομένα των 32 αφίξεων, παρατηρούμε ότι ίδιο πρότυπο επαναλαμβάνεται οκτώ φορές. Έτσι, έχουμε ένα πρότυπο το οποίο εμφανίζεται στα αρχικά δεδομένα, αλλά και σε διαφορετικούς βαθμούς ομαδοποίησης. Αυτή είναι η έννοια της αυτό-ομοιότητας.

Ένα φαινόμενο το οποίο έχει το χαρακτηριστικό της αυτό-ομοιότητας μοιάζει το ίδιο ή συμπεριφέρεται με τον ίδιο τρόπο σε διαφορετικές κλίμακες χώρου (μήκος, πλάτος) ή χρόνου. Το Σχήμα 3.1, παρουσιάζει την κατασκευή του συνόλου του Cantor, η οποία ακολουθεί τους εξής κανόνες:



Σχήμα 3.1 Το σύνολο του Cantor μετά από πέντε βήματα αναδρομής

1. Αρχίζουμε με το κλειστό διάστημα  $[0, 1]$ , το οποίο αναπαρίσταται από ένα ευθύγραμμο τμήμα.
2. Μετακινούμε από το κέντρο του ευθύγραμμου τμήματος το  $1/3$  αυτού.
3. Σε κάθε βήμα, μετακινούμε από το κέντρο του ευθύγραμμου τμήματος που προέκυψε από το προηγούμενο βήμα το  $1/3$  αυτού.

Τα βήματα αυτά αποτελούν μια αναδρομική διαδικασία ή οποία μπορεί να οριστεί με περισσότερη ακρίβεια ως εξής: Ας συμβολίσουμε με  $S_i$  το σύνολο του Cantor μετά από  $i$  επαναλήψεις. Έτσι έχουμε:

$$\begin{aligned}S_0 &= [0, 1] \\S_1 &= [0, 1/3] \cup [2/3, 1] \\S_2 &= [0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1]\end{aligned}$$

κλπ. Από το σύνολο του Cantor προκύπτουν δύο κοινές ιδιότητες για όλα τα φαινόμενα που χαρακτηρίζονται από αυτό-ομοιότητα:

1. Ύπαρξη δομής ακόμα και σε πολύ μικρές κλίμακες.
2. Επαναλαμβανόμενες δομές.

Αυτές οι ιδιότητες δεν ισχύουν οπωσδήποτε σε όλα τα πραγματικά φαινόμενα. Μετά από έναν αριθμό επαναλήψεων αλλαγής κλίμακας, η δομή και η ομοιότητα παύουν να ισχύουν. Παρ' όλα αυτά για έναν μεγάλο εύρος από κλίμακες, πολλά φαινόμενα εμφανίζουν αυτό-ομοιότητα.

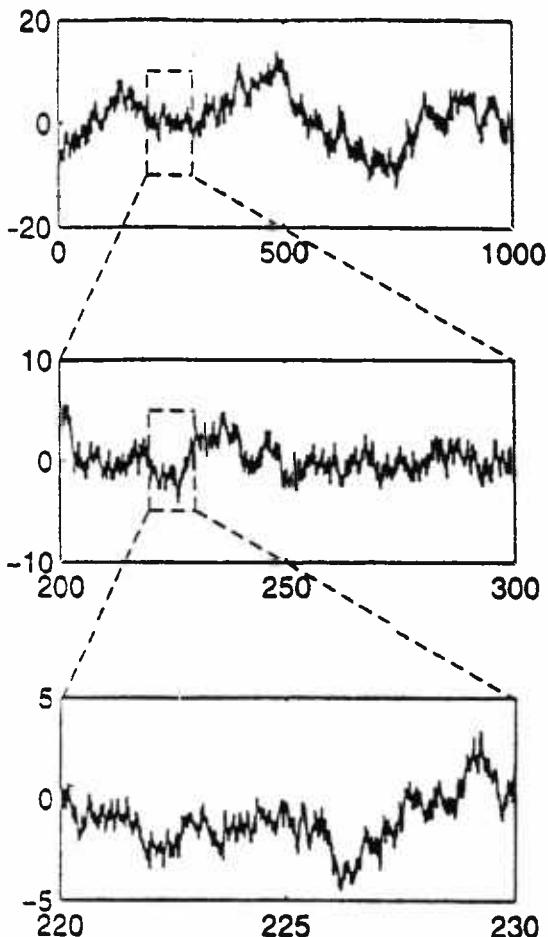
## 2. ΑΥΤΟ-ΟΜΟΙΑ ΔΕΔΟΜΕΝΑ ΚΙΝΗΣΗΣ

Ο τύπος της αυτό-ομοιότητας που περιγράφηκε παραπάνω μπορεί να οριστεί ως **ακριβής αυτό-ομοιότητα** (exact self-similarity): Ένα συγκεκριμένο υπόδειγμα αναπαράγεται επακριβώς σε διαφορετικές κλίμακες. Η ακριβής αυτό-ομοιότητα έχει νόημα μόνο σε ντετερμινιστικά δεδομένα. Τα δεδομένα κίνησης (traffic data) εξετάζονται ως στοχαστικές διαδικασίες και κατά συνέπεια μπορούμε μόνο να μιλάμε για **στατιστική αυτό-ομοιότητα** (statistical self-similar).

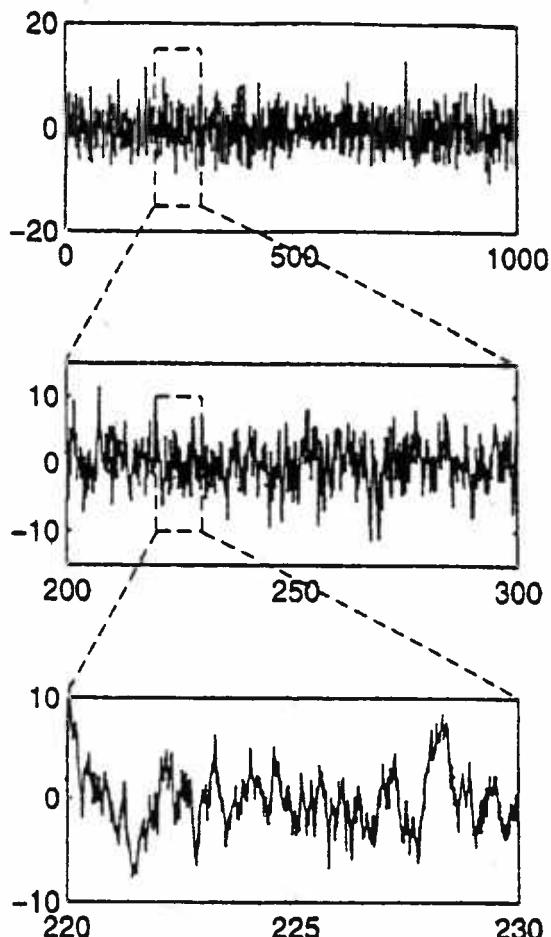
Μία ντετερμινιστική αυτό-όμοια διαδικασία είναι ανεξάρτητη από αλλαγές κλίμακας. Αντίθετα, για μια στοχαστική διαδικασία οι στατιστικές της διαδικασίας είναι ανεξάρτητες των αλλαγών κλίμακας. Αυτό σημαίνει ότι η συνολική συμπεριφορά της διαδικασίας είναι η ίδια σε μικρές και μεγάλες κλίμακες χρόνου.

Στο Σχήμα 3.2 (α), παρουσιάζεται ένα παράδειγμα αυτό-όμοιας στοχαστικής διαδικασίας. Η συνάρτηση δεν αναπάραγεται ακριβώς σε διαφορετικές κλίμακες χρόνου, όμως τα ιστογράμματα που αντιστοιχούν σε διαφορετικές χρονικές κλίμακες, μοιάζουν μεταξύ τους. Στο Σχήμα 3.2 (β), παρουσιάζεται μια τυπική, στάσιμη, τυχαία διαδικασία. Σε αυτή την περίπτωση παρατηρείται ότι σε μικρές κλίμακες χρόνου, η συνάρτηση έχει

διαφορετικά χαρακτηριστικά: παρουσιάζει μεγαλύτερη διακύμανση και γίνεται λιγότερο ομαλή.



Σχήμα 3.2 (α) Στατιστικά αυτό-όμοια στοχαστική διαδικασία.



Σχήμα 3.2 (β) Μη αυτό-όμοια στοχαστική διαδικασία.

Για την καλύτερη κατανόηση της έννοιας των αυτό-όμοιων δεδομένων κίνησης, ακολουθεί μια συνοπτική παρουσίαση μιας σχετικής έρευνας, τα αποτελέσματα της οποίας δημοσιεύτηκαν στο [WILL97]. Οι συγγραφείς του άρθρου στηριζόμενοι στα αποτελέσματα της στατιστικής ανάλυσης ενός μεγάλου όγκου δεδομένων κίνησης, ισχυρίζονται ότι η κίνηση που προέρχεται από τοπικά δίκτυα τύπου Ethernet, απαιτεί μια νέα προσέγγιση μοντελοποίησης και ανάλυσης η οποία θα λαμβάνει υπ'όψη τα αυτό-όμοια χαρακτηριστικά της κίνησης.

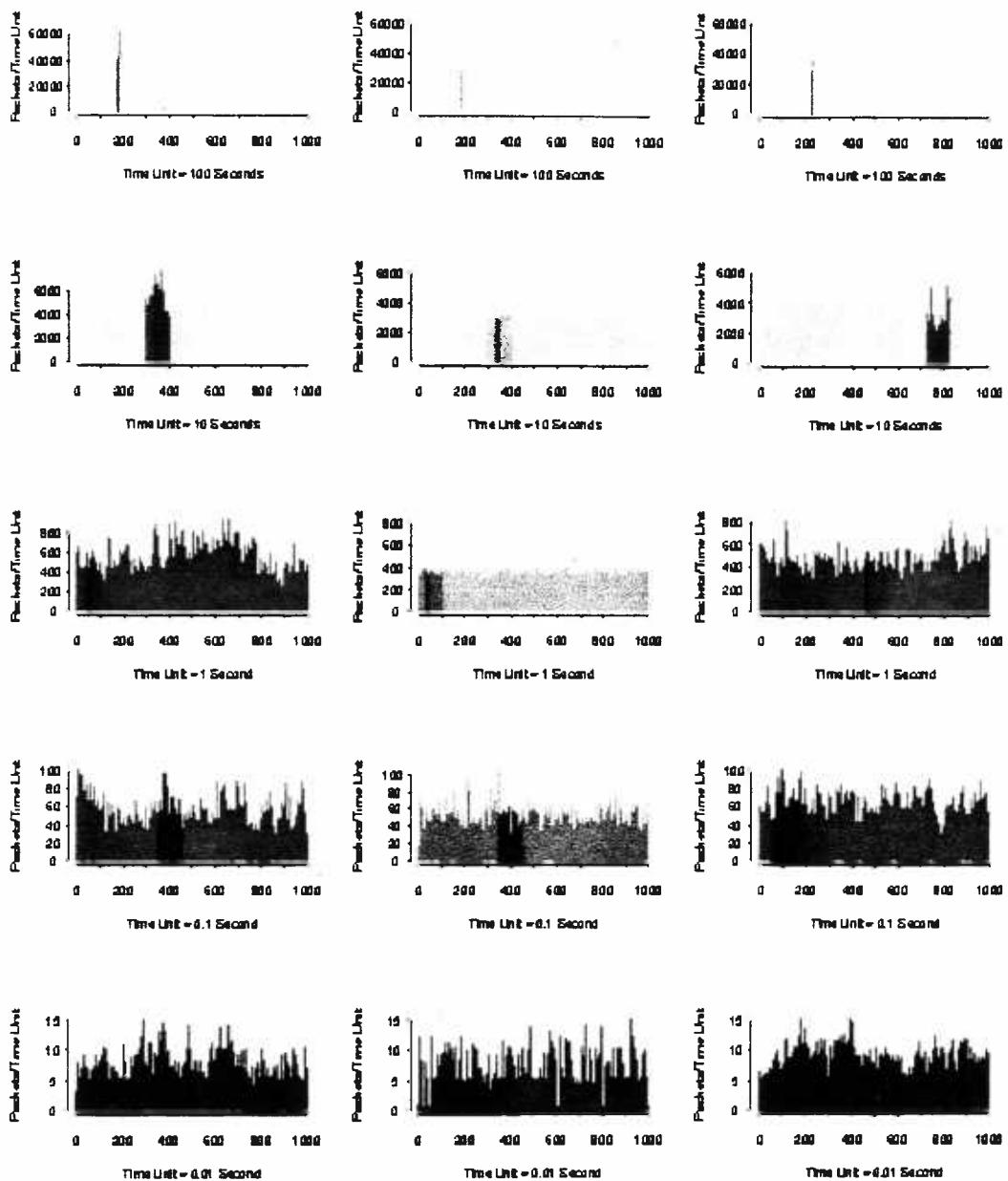
Στην αριστερή στήλη του Σχήματος 3.3, εμφανίζεται μια σειρά από διαγράμματα του αριθμού των πακέτων ανά μονάδα χρόνου. Τα στοιχεία αποτελούν πραγματικά δεδομένα κίνησης τα οποία προέκυψαν από τον επί 27 ώρες συνεχή έλεγχο της κίνησης Ethernet δικτύου. Στο πρώτο διάγραμμα της αριστερής στήλης παρουσιάζεται η κίνηση που καταγράφηκε στο διάστημα των 27 ωρών. Χρησιμοποιώντας ως μονάδα χρόνου τα 100 δευτερόλεπτα το διάγραμμα περιλαμβάνει περίπου 1000 σημεία. Κάθε διάγραμμα που

ακολουθεί προκύπτει από το προηγούμενό του αυξάνοντας την ανάλυση του χρόνου (time resolution) κατά 10 φορές, και εμφανίζοντας μόνο ένα τυχαία επιλεγμένο υποδιάστημα του προηγούμενου γραφήματος (το οποίο υποδεικνύεται με σκούρο χρώμα). Ετσι, το δεύτερο διάγραμμα καλύπτει μια περίοδο 2.7 ωρών, το τρίτο 0.27 ωρών κλπ. Παρατηρώντας τα διαγράμματα καθώς κινούμαστε από το τελευταίο προς το αρχικό διάγραμμα της πρώτης στήλης, προκύπτει ότι κάθε σημείο οποιουδήποτε διαγράμματος, προκύπτει από την άθροιση των αντίστοιχων 10 διαδοχικών σημείων του επόμενου διαγράμματος.

Όλα τα διαγράμματα της πρώτης στήλης, με μια πιθανή εξαίρεση του πρώτου διαγράμματος, μοιάζουν μεταξύ τους όσον αφορά τον τρόπο με τον οποίο κατανέμονται. Αυτό σημαίνει ότι η Ethernet κίνηση έχει την τάση να μοιάζει ίδια σε μεγάλες (ώρες, λεπτά) αλλά και σε μικρές κλίμακες χρόνου (δευτερόλεπτα, χλιοστά του δευτερολέπτου).

Η αντίθετη περίπτωση εμφανίζεται στη μεσαία στήλη του Σχήματος 3.3. Τα διαγράμματα της μεσαίας στήλης δημιουργήθηκαν με τον ίδιο τρόπο που δημιουργήθηκαν και τα διαγράμματα της Ethernet κίνησης της πρώτης στήλης, με τη διαφορά ότι για την παραγωγή των δεδομένων χρησιμοποιήθηκε το μοντέλο του Poisson. Σε υψηλά επίπεδα ανάλυσης χρόνου (χρησιμοποιούμενη μονάδα χρόνου 0.01 δευτερόλεπτα) η κίνηση είναι κατά ριτές. Καθώς όμως τα δεδομένα αθροίζονται σε συνεχώς αυξανόμενες κλίμακες χρόνου, το υπόδειγμα της κίνησης ομαλοποιείται.

Βασισμένοι σε μια πληθώρα στατιστικών ελέγχων, οι συγγραφείς του άρθρου υπολόγισαν ότι τα δεδομένα της Ethernet κίνησης ήταν αυτό-όμοια με τιμή για την παράμετρο Hurst ίση με 0.9. Η δεξιά στήλη του Σχήματος 3.3 παρέχει μια οπτική επιβεβαίωση της στατιστικής ανάλυσης. Αυτά τα διαγράμματα δημιουργήθηκαν χρησιμοποιώντας ένα αυτό-όμοιο μοντέλο κίνησης με  $H = 0.9$  και δείχνουν τα ίδια γενικά χαρακτηριστικά με τα διαγράμματα της πραγματικής Ethernet κίνησης της πρώτης στήλης.



Σχήμα 3.3 Σύγκριση μεταξύ πραγματικών και συνθετικών δεδομένων Ethernet κίνησης [WILL97].

### 3. ΟΡΙΣΜΟΙ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ

#### 3.1 ΟΡΙΣΜΟΣ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ ΓΙΑ ΣΥΝΕΧΕΙΣ ΤΙΜΕΣ ΧΡΟΝΟΥ

Ένας αρκετά διαδεδομένος ορισμός για τις αυτό-όμοιες στοχαστικές διαδικασίες βασίζεται στην άμεση αλλαγή κλίμακας της συνεχούς μεταβλητής χρόνου, ως εξής: Μια στοχαστική διαδικασία  $\mathbf{x}(t)$  είναι στατιστικά αυτό-όμοια με παράμετρο  $H$  ( $0.5 \leq H \leq 1$ ) αν για κάθε πραγματικό  $\alpha > 0$ , η διαδικασία  $\alpha^{-H} \mathbf{x}(\alpha t)$  έχει τις ίδιες στατιστικές ιδιότητες με την  $\mathbf{x}(t)$ . Αυτή η σχέση εκφράζεται από τις ακόλουθες τρεις συνθήκες:

1.  $E[\mathbf{x}(at)] = \frac{E[\mathbf{x}(at)]}{\alpha^H}$  Μέση τιμή
2.  $Var[\mathbf{x}(at)] = \frac{Var[\mathbf{x}(at)]}{\alpha^{2H}}$  Διακύμανση
3.  $R_x(t,s) = \frac{R_x(at,as)}{\alpha^{2H}}$  Αυτοσυσχέτιση

Η παράμετρος  $H$ , γνωστή ως παράμετρος του Hurst, ή παράμετρος της αυτό-ομοιότητας, αποτελεί τον δείκτη της αυτό-ομοιότητας. Ακριβέστερα, η παράμετρος  $H$ , μετράει την διατήρηση (persistence) ενός στατιστικού φαινομένου –π.χ. την ύπαρξη των χαρακτηριστικών της κίνησης κατά ριτές σε όλες τις κλίμακες χρόνου. Όταν  $H = 0.5$ , δεν υπάρχει ένδειξη αυτό-ομοιότητας. Όσο πιο κοντά στην μονάδα βρίσκεται η παράμετρος  $H$ , τόσο μεγαλύτερος είναι ο βαθμός της διατήρησης (persistence) ή του long-range dependence.

#### 3.2 ΟΡΙΣΜΟΣ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ ΓΙΑ ΔΙΑΚΡΙΤΕΣ ΤΙΜΕΣ ΧΡΟΝΟΥ

Για μια στάσιμη χρονολογική σειρά  $\mathbf{x}$ , ορίζουμε την  $m$ -αθροιστική χρονολογική σειρά  $\mathbf{x}^{(m)} = \{\mathbf{x}_k^{(m)}, k = 0, 1, 2, \dots\}$ , αθροίζοντας χωρίς επικάλυψη τη γειτονικές τιμές της αρχικής χρονολογικής σειράς. Αυτό μπορεί να εκφραστεί ως εξής:

$$\mathbf{x}_k^{(m)} = \frac{1}{m} \sum_{i=k-(m-1)}^{km} \mathbf{x}_i$$

Για παράδειγμα η  $\mathbf{x}^{(3)}$ , ορίζεται ως:

$$\mathbf{x}_k^{(3)} = \frac{\mathbf{x}_{3k-2} + \mathbf{x}_{3k-1} + \mathbf{x}_{3k}}{3}$$

Η αθροιστική χρονολογική σειρά αποτελεί μια τεχνική “συμπίεσης” των δεδομένων της αρχικής χρονολογικής σειράς. Μπορούμε να θεωρήσουμε ότι η  $\mathbf{x}^{(1)}$  είναι η μεγαλύτερη δυνατή “μεγέθυνση” ή “ανάλυση” για την χρονολογική σειρά. Η  $\mathbf{x}^{(3)}$  προκύπτει

αθροίζοντας διαδοχικές τριάδες τιμών της  $\mathbf{x}^{(1)}$  και διαιρώντας το άθροισμα δια του 3. Έτσι η  $\mathbf{x}^{(3)}$  είναι η ίδια διαδικασία με την  $\mathbf{x}^{(1)}$ , με ένα παράγοντα “σμίκρυνσης” 3. Αν τα στατιστικά της διαδικασίας (μέσος, διακύμανση, αυτοσυσχέτιση κλπ) διατηρούνται και μετά την συμπίεση, τότε η διαδικασία είναι αυτό-όμοια.

Μια διαδικασία  $\mathbf{x}$  καλείται **ακριβώς αυτό-όμοια** (exactly self-similar) με παράμετρο  $\beta$  ( $0 < \beta < 1$ ) αν για όλα τα  $m = 1, 2, \dots$  έχουμε

$$Var(\mathbf{x}^{(m)}) = \frac{Var(\mathbf{x})}{m^\beta} \quad \text{Διακύμανση}$$

$$R_{\mathbf{x}^{(m)}}(k) = R_{\mathbf{x}}(k) \quad \text{Αυτοσυσχέτιση}$$

Μπορεί να αποδειχθεί ότι η παράμετρος  $\beta$  σχετίζεται με την παράμετρο Hurst ως εξής:  $H = 1 - (\beta/2)$ .

Μια ασθενέστερη συνθήκη είναι η εξής: Μια διαδικασία  $\mathbf{x}$  καλείται **ασυμπτωτικά αυτό-όμοια** (asymptotically self-similar) αν για όλα τα μεγάλα  $k$ , ισχύουν τα εξής:

$$Var(\mathbf{x}^{(m)}) = \frac{Var(\mathbf{x})}{m^\beta} \quad \text{Διακύμανση}$$

$$R_{\mathbf{x}^{(m)}}(k) \rightarrow R_{\mathbf{x}}(k) \quad \text{όταν } m \rightarrow \infty \quad \text{Αυτοσυσχέτιση}$$

Έτσι, με αυτόν τον ορισμό της αυτό-ομοιότητας, η αυτοσυσχέτιση των αθροιστικών διαδικασιών έχουν την ίδια μορφή με την αρχική διαδικασία. Αυτό σημαίνει ότι ο βαθμός της μεταβλητότητας ή της κίνησης κατά ριπές (burstiness) είναι ο ίδιος σε διαφορετικές κλίμακες χρόνου.

Ένα ενδιαφέρον χαρακτηριστικό των προηγουμένων ορισμών είναι ότι η αυτοσυσχέτιση των αθροιστικών χρονολογικών σειρών των αυτό-όμοιων διαδικασιών δεν τείνουν προς το μηδέν όταν το  $m$  τείνει στο άπειρο. Το γεγονός αυτό έρχεται σε αντίθεση με τις στοχαστικές διαδικασίες οι οποίες χρησιμοποιούνται για την μοντελοποίηση πακέτων δεδομένων, για τις οποίες ισχύει η ακόλουθη σχέση:

$$R^{(m)}(\tau) \rightarrow 0 \quad \text{όταν } m \rightarrow \infty$$

Μια συνάρτηση αυτοσυσχέτισης  $R(\tau)$  η οποία είναι ίση με το μηδέν είναι συνεπής με τον άσπρο θόρυβο (white noise). Στο Σχήμα 3.2 ( $\beta$ ) καθώς ο βαθμός άθροισης ( $m$ ) αυξάνει, η διαδικασία όλο και περισσότερο παίρνει τη μορφή άσπρου θορύβου. Σε αντίθεση στο Σχήμα 3.2 ( $\alpha$ ) όλα τα διαγράμματα έχουν παρόμοια μορφή και είναι διαφορετικά του άσπρου θορύβου.

Επίσης πρέπει να σημειωθεί σχετικά με τους παραπάνω ορισμούς, ότι η διακύμανση της  $\mathbf{x}^{(m)}$  μειώνεται βραδύτερα από  $1/m$ , καθώς  $m \rightarrow \infty$ . Ακριβέστερα, μειώνεται με ρυθμό  $1/m^\beta$ .

### 3.3 ΕΞΑΡΤΗΣΗ ΜΑΚΡΟΥ ΧΡΟΝΟΥ (LONG-RANGE DEPENDENCE)

Μια από τις πιο σημαντικές ιδιότητες των αυτό-όμοιων διαδικασιών αποτελεί η εξάρτηση μακρού χρόνου. Η ιδιότητα αυτή ορίζεται σε σχέση με τη συμπεριφορά της αυτοσυνδιακύμανσης (autocovariance)  $C(\tau)$  όταν το  $\tau$  αυξάνει.

Γενικά, μια διαδικασία βραχαίου χρόνου ικανοποιεί τη συνθήκη ότι η αυτοσυνδιακύμανσή της μειώνεται τουλάχιστον με εκθετικό ρυθμό:

$$C(k) \sim a^{|k|} \quad \text{όταν } |k| \rightarrow \infty, \quad 0 < a < 1$$

όπου το σύμβολο  $\sim$  συμβολίζει ότι οι εκφράσεις στο δεξί και αριστερό μέλος είναι ασυμπτωτικά ανάλογες.

Αντίθετα, στις long-range dependent διαδικασίες η αυτοσυνδιακύμανση παρουσιάζει υπερβολική (hyperbolically) μείωση:

$$C(k) \sim |k|^{-\beta} \quad \text{όταν } |k| \rightarrow \infty, \quad 0 < \beta < 1$$

όπου  $\beta$  είναι η ίδια παράμετρος που αναφέρθηκε προηγουμένων και σχετίζεται με την Hurst παράμετρο, με βάση τον τύπο  $H = 1 - (\beta/2)$ . Σε αυτή την περίπτωση  $\sum_k C(k) = \infty$ .

### 3.4 HEAVY-TAILED ΚΑΤΑΝΟΜΕΣ

Οι προηγούμενοι ορισμοί των αυτό-όμοιων δεδομένων κίνησης (αθροιστικές χρονολογικές σειρές, long-range dependence) είναι ισοδύναμοι. Παρ' όλα αυτά, οι αυτό-όμοιες στοχαστικές διαδικασίες μπορούν να οριστούν και με τη βοήθεια των heavy-tailed κατανομών.

Οι heavy-tailed κατανομές χρησιμοποιούνται για την περιγραφή διαδικασιών κίνησης όπως ο χρόνος μεταξύ διαδοχικών αφίξεων πακέτων καθώς και το μήκος των ριπών (κίνηση κατά ριπές).

Η κατανομή μιας τυχαίας μεταβλητής  $X$  είναι heavy tailed εάν:

$$1 - F(x) = \Pr[X > x] \sim \frac{1}{x^\alpha} \quad \text{Όταν } x \rightarrow \infty, 0 < \alpha$$

Γενικά, μια τυχαία μεταβλητή με heavy-tailed κατανομή παρουσιάζει μεγάλη ή ακόμα και άπειρη διακύμανση.

Η πιο απλή heavy-tailed κατανομή είναι η κατανομή Pareto, με παραμέτρους  $k$  και  $a$  ( $k, a > 0$ ), και συναρτήσεις κατανομής και πυκνότητας πιθανότητας:

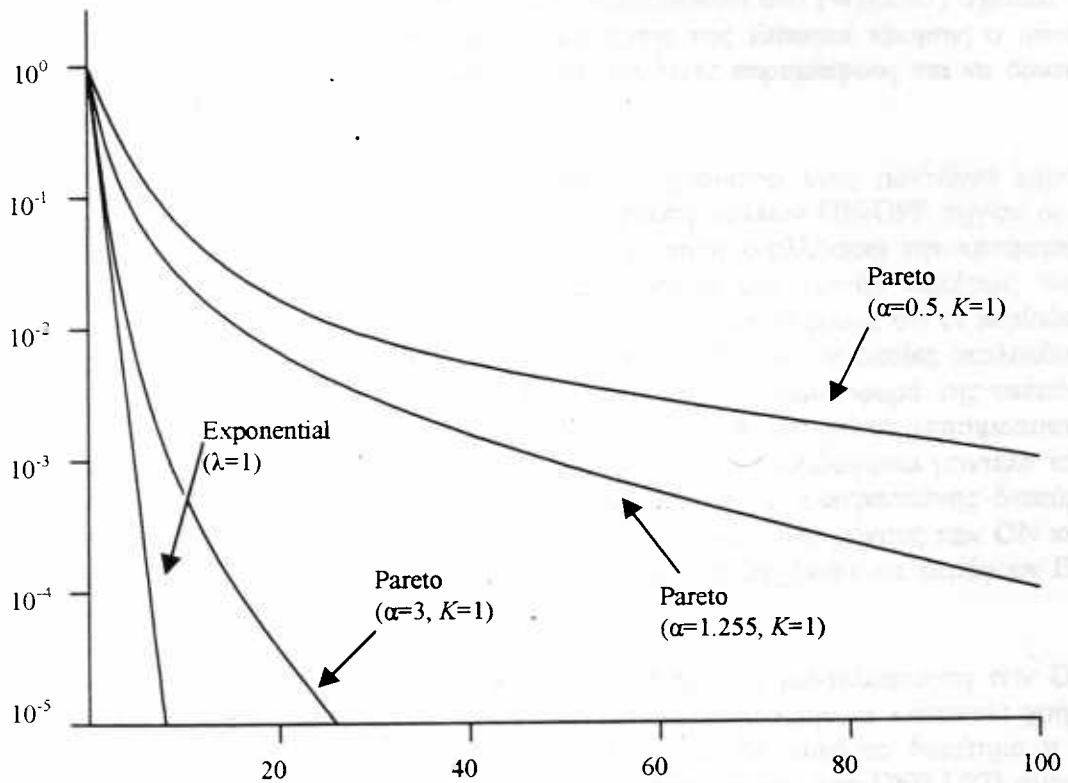
$$f(x) = F(x) = 0 \quad (x \leq k)$$

$$f(x) = \frac{a}{k} \left( \frac{k}{x} \right)^{a+1} \quad F(x) = 1 - \left( \frac{k}{x} \right)^a \quad (x > k, a > 0)$$

καθώς και μέση τιμή

$$E[X] = \frac{\alpha}{\alpha - 1} k \quad (\alpha > 1)$$

Η παράμετρος  $k$  είναι η μικρότερη τιμή που μπορεί να πάρει η τυχαία μεταβλητή. Η παράμετρος  $\alpha$  ορίζει το μέσο και την διακύμανση της τυχαίας μεταβλητής: Αν  $\alpha \leq 2$ , τότε η κατανομή έχει άπειρη διακύμανση, ενώ αν  $\alpha \leq 1$ , τότε έχει πεπερασμένο μέσο και διακύμανση. Συγκρίνοντας την Pareto και την εκθετική κατανομή (Σχήμα 3.4), προκύπτει ότι η ουρά της Pareto κατανομής μειώνεται πολύ πιο αργά από αυτήν της εκθετικής. Αυτό το φαινόμενο οδήγησε στον όρο *heavy tail*.



Σχήμα 3.4 Σύγκριση εκθετικής και Pareto συνάρτησης πυκνότητας πιθανότητας.

#### 4. ΑΙΤΙΑ ΕΜΦΑΝΙΣΗΣ ΤΟΥ ΦΑΙΝΟΜΕΝΟΥ ΤΗΣ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ

Όπως έχει ήδη αναφερθεί, το κύριο χαρακτηριστικό των αυτό-όμοιων διαδικασιών είναι ότι εμφανίζουν τα ίδια χαρακτηριστικά σε διαφορετικές κλίμακες χρόνου. Θα μπορούσε να ειπωθεί ότι η ιδιότητα αυτή υπονοεί την ύπαρξη μιας πολυεπίπεδης ιεραρχίας μηχανισμών (π.χ. ένας για κάθε κλίμακα χρόνου) ο συνδυασμός της επενέργειας των οποίων δημιουργεί το φαινόμενο της αυτό-ομοιότητας.

Το ερώτημα που ανακύπτει αφορά τα αίτια πρόκλησης ενός τέτοιου φαινομένου. Η εγκυρότητα των συμβατικών μοντέλων ουρών αναμονής τα οποία χρησιμοποιούνται για την ανάλυση κίνησης έγκειται στην υπόθεση ότι οι στατιστικές τους είναι ίδιες με τις στατιστικές που προκύπτουν από τον συνδυασμό των δεδομένων ενός αριθμού πηγών (τέτοιοι συνδυασμοί παρατηρούνται σε ένα τοπικό δίκτυο ή στην περίπτωση της πολύπλεξης). Σύμφωνα με την έρευνα που δημοσιεύεται στο [WILL97] σχετικά με την Ethernet κίνηση, υπάρχει τρόπος μοντελοποίησης της Ethernet κίνησης ο οποίος (1) ταιριάζει στα πραγματικά δεδομένα, (2) απαιτεί λίγες παραμέτρους για να οριστεί, (3) είναι διαισθητικά αποδεκτός.

Η προσέγγιση που χρησιμοποιήθηκε είναι η δημιουργία ενός μοντέλου παραγωγής Ethernet κίνησης η οποία βασίζεται στην υπέρθεση πολλών ON/OFF πηγών οι οποίες ακολουθούν την κατανομή Pareto. Κάθε τέτοια πηγή εναλλάσσει την κατάστασή της μεταξύ ON περιόδων, κατά τις οποίες μεταφέρεται μια «ριπή» πακέτων, και OFF περιόδων, κατά τις οποίες δεν υπάρχει μετάδοση. Αν υποθέσουμε ότι οι περίοδοι κάθε πηγής χαρακτηρίζονται από ανεξάρτητες τυχαίες μεταβλητές οι οποίες ακολουθούν την ίδια κατανομή, τότε είναι δυνατό να χαρακτηριστεί η συμπεριφορά της υπέρθεσης ή πολύπλεξης πολλών τέτοιων πηγών. Αυτή η προσέγγιση μπορεί να χρησιμοποιηθεί για την παραγωγή κίνησης η οποία είναι «συμβατή» με τα παραδοσιακά μοντέλα κίνησης. Κάτι τέτοιο απαιτεί την χρησιμοποίηση μιας κατανομής πεπερασμένης διακύμανσης (όπως η εκθετική ή η γεωμετρική) για τον χαρακτηρισμό του μήκους των ON και OFF περιόδων. Όπως έχει αποδειχθεί αυτά τα μοντέλα αποτυγχάνουν να παράγουν Ethernet κίνηση όμοια με την πραγματική.

Αυτό που έκαναν οι συγγραφείς του [WILL97] ήταν η μοντελοποίηση των ON/OFF περιόδων με κατανομές άπειρης διακύμανσης και συγκεκριμένα κάνοντας χρήση της κατανομής Pareto με παράμερο α μεταξύ 1 και 2. Σε αυτό το διάστημα η τυχαία μεταβλητή έχει πεπερασμένο μέσο και άπειρη διακύμανση. Στο [WILL97] αναφέρεται ότι η υπέρθεση πολλών ON/OFF πηγών που ακολουθούν την Pareto κατανομή οδηγεί σε αυτό-όμοια κίνηση με παράμετρο Hurst  $H=(3-\alpha)/2$ . Σημειώνεται ότι για  $1 < \alpha < 2$ , ισχύει  $0.5 < H < 1$  (ύπαρξη αυτό-ομοιότητας).

Στο [WILL97] αναφέρεται ότι μια heavy-tailed κατανομή όπως η Pareto αντικατοπτρίζει την πραγματική συμπεριφορά κάθε μίας από τις Ethernet πηγές. Διαισθητικά η υψηλή ή άπειρη διακύμανση μιας heavy-tailed κατανομής συνεπάγεται υψηλή μεταβλητότητα και συνεπώς μεταβλητότητα σε όλες τις κλίμακες χρόνου. Μια εφαρμογή ή ένας σταθμός εργασίας παράγει κίνηση κατά ριπές με ανενεργές περιόδους ενδιαμέσως. Με μια κατανομή υψηλής διακύμανσης το εύρος των διαστημάτων χρόνου μπορεί να είναι πολύ

μεγάλο με πολλές μικρές ριτές κίνησης, πολλές μεγάλες ριτές κίνησης και μερικές πολύ μεγάλες ριτές κίνησης. Ο Mandelbrot, ο μαθηματικός που επινόησε τον όρο *fractal*, αναφέρεται σε αυτό το χαρακτηριστικό με τον όρο *Noah effect* σε αναφορά με το εδάφιο της Βίβλου (Γένεσις 7:11-12): «Στον εξήκοστο χρόνο ζώής του Noah, τα παράθυρα του ουρανού ἀνοιξαν, και η βροχή στάθηκε πάνω από τη γη σαράντα ημέρες και σαράντα νύχτες.»

## 5. ΠΑΡΑΔΕΙΓΜΑΤΑ ΑΥΤΟ-ΟΜΟΙΩΝ ΔΕΔΟΜΕΝΩΝ ΚΙΝΗΣΗΣ

Τα τελευταία χρόνια εμφανίστηκαν στην βιβλιογραφία μια σειρά από μελέτες σύμφωνα με τις οποίες το πρότυπο (pattern) της κίνησης δικτύων ταιριάζει με αυτό των αυτό-ομοιων διαδικασιών. Στην παραγράφο αυτή γίνεται προσπάθεια μιας σύντομης παρουσίασης αυτών των ερευνών.

### 5.1 ETHERNET ΔΕΔΟΜΕΝΑ ΚΙΝΗΣΗΣ

Στο άρθρο [LEL94] σχετικά με την κίνηση Ethernet τοπικών δικτύων, οι ερευνητές ασχολήθηκαν με τις επιπτώσεις της αυτό-ομοιότητας στην απόδοση δικτύων υπολογιστών. Μια σημαντική ανακάλυψη ήταν πως όσο πιο μεγάλη είναι η κίνηση σε ένα Ethernet δίκτυο, τόσο πιο υψηλή είναι η εκτίμηση της παραμέτρου Hurst  $H$ , ή ισοδύναμα, τόσο μεγαλύτερος είναι ο βαθμός της αυτό-ομοιότητας. Η διαπίστωση αυτή είναι σημαντική ακριβώς επειδή σε συνθήκες μεγάλης κίνησης, τα θέματα απόδοσης παρουσιάζουν μεγαλύτερο ενδιαφέρον.

Ένα εξίσου σημαντικό αποτέλεσμα από την ανάλυση των Ethernet δεδομένων κίνησης ήταν η ανεπάρκεια των συμβατικών μοντέλων αναμονής να προβλέψουν την απόδοση. Για παράδειγμα, μια γνωστή υπόθεση σχετικά με τα δεδομένα κίνησης είναι ότι πολυπλέκοντας ένα μεγάλο αριθμό από ανεξάρτητες πηγές κίνησης προκύπτει μια διαδικασία Poisson. Στο [WILL94] αναφέρεται ότι αυτή η υπόθεση καθώς και η ανάλυση ουρών αναμονής η οποία στηρίζεται σε αυτή, οδήγησε τους πωλητές των ATM μεταγωγέων να κατασκευάσουν τους μεταγωγείς πρώτης γενιάς με μικρούς ενταμιευτές (10-100 κελιά). Σύμφωνα με την ανάλυση στο [LEL94], εάν η διαδικασία εισόδου παρουσιάζει αυτό-ομοιότητα, τότε αναμένονται αυξημένες καθυστερήσεις σε κάθε είδους πολύπλεξη αυτό-όμοιων ροών, έτσι ώστε να απαιτείται αυξημένο μέγεθος ενταμιευτών. Αυτό ισχύει για ATM, frame relay, και 100BASE-T μεταγωγείς, WAN δρομολογητές, τοπικά δίκτυα διαμοιραζόμενου μέσου, καθώς και στατιστικούς πολυπλέκτες.

### 5.2 ΚΙΝΗΣΗ WORLD WIDE WEB

Το άρθρο [CROV96] αφορά μια μελέτη Web κίνησης, η οποία περιλαμβάνει πάνω από μισό εκατομμύριο αιτήσεις στο Web. Τα δεδομένα συλλέχθηκαν από 37 κατάλληλα διαμορφωμένους Web browsers του Πανεπιστημίου της Βοστώνης. Η έρευνα έδειξε ότι

το πρότυπο της κίνησης που δημιουργήθηκε από τους Web browsers ήταν αυτό-όμοιο. Οι αναλυτές θεώρησαν κάθε Web browser ως μια ON/OFF πηγή και βρήκαν ότι η κατανομή Pareto ταιριάζει πολύ καλά στα δεδομένα.

Στα πλαίσια αυτής της έρευνας οι αναλυτές προσπάθησαν να εξηγήσουν την εμφάνιση της αυτό-όμοιας συμπεριφοράς. Εξετάζοντας τα μεγέθη των αρχείων που μεταφέρονταν από τους εξυπηρετητές στους browsers βρήκαν ότι αυτά ακολουθούν μια κατανομή που ομοιάζει με την Pareto. Οι συγγραφείς ερμήνευσαν το γεγονός βάσει του σκεπτικού ότι η Web κίνηση αντικατοπτρίζει μια τυχαία επιλογή για μεταφορά αρχείων. Συγκεκριμένα, αν οι χρήστες επιλέγουν αρχεία ακολουθώντας συνδέσμους, αγνοώντας το μέγεθός τους, τότε αναμένεται ότι το μέγεθος της μεταφοράς αντιπροσωπεύει ένα τυχαίο δείγμα όλης της κατανομής μεγέθους των Web αρχείων. Επιπρόσθετα, οι αναλυτές έδειξαν ότι τα αρχεία που είναι διαθέσιμα μέσω του Web στο Internet έχουν μια heavy-tailed κατανομή μεγέθους. Κάτι τέτοιο είναι διαισθητικά αποδεκτό αφού υπάρχουν πολλά μικρά αρχεία διαθέσιμα μέσω του Web αλλά και πολλά μεγάλα και πολύ μεγάλα (αρχεία πολυμέσων).

### 5.3 ΚΙΝΗΣΗ TCP, FTP ΚΑΙ TELNET

Το άρθρο [PAXS95] αναφέρεται σε μια μελέτη για την TCP κίνηση ευρείας περιοχής, καθώς και στην FTP και TELNET κίνηση η οποία μεταφέρεται μέσω TCP συνδέσεων. Τα συμπεράσματα που προκύπτουν συνοψίζονται ως εξής:

- Τα ευρέως χρησιμοποιούμενα μοντέλα Poisson υποεκτιμούν σε μεγάλο βαθμό την κίνηση κατά ριτές της TCP κίνησης για ένα μεγάλο εύρος κλιμάκων χρόνου.
- Όσον αφορά την TELNET κίνηση οι αφίξεις των αιτήσεων σύνδεσης ακολουθούν την κατανομή Poisson. Όμως, η κατανομή Poisson για τις αφίξεις των πακέτων – δηλαδή ότι οι χρόνοι μεταξύ των αφίξεων ακολουθούν την εκθετική κατανομή – υποεκτιμά σημαντικά το χαρακτηριστικό της κίνησης κατά ριτές.
- Όπως συμβαίνει και με την TELNET κίνηση, οι αφίξεις των FTP συνόδων ταιριάζουν με το μοντέλο Poisson, αλλά η μεταφορά των δεδομένων ακολουθεί έναν λιγότερο ομαλό ρυθμό αφίξεων.

### 5.4 VIDEO ΜΕΤΑΒΛΗΤΟΥ ΡΥΘΜΟΥ BITS

Μια σειρά μελετών έχουν δείξει ότι η κίνηση μεταφοράς ψηφιακού video πάνω από ATM δίκτυα και το Internet είναι αυτό-όμοια. Στο [GARR94] πραγματοποιήθηκε ένα πείραμα μεταφοράς video, χρησιμοποιώντας την ταινία *O πόλεμος των Άστρων*. Η κωδικοποίηση της ταινίας πραγματοποιήθηκε βάσει του JPEG προτύπου. Το αποτέλεσμα ήταν μια ροή δεδομένων η οποία αποτελούνταν από μεταβλητού μήκους πλαίσια. Το μεταβλητό μήκος των πλαισίων οφείλεται στην φύση του αλγορίθμου συμπίεσης/κωδικοποίησης. Λόγω αυτής ακριβώς της μεταβλητότητας των πλαισίων προκύπτει η έννοια της στοχαστικής διαδικασίας.

Το αποτέλεσμα της ανάλυσης έδειξε ότι η μεταφορά του video εμφανίζει έναν αυτό-όμοιο χαρακτήρα και ότι το μήκος των πλαισίων ακολουθούν μια Pareto κατανομή, τουλάχιστον όσον αφορά την ουρά της κατανομής. Οι συγγραφείς έδειξαν πώς η υψηλή μεταβλητότητα του μήκους των πλαισίων σχετίζεται με την δράση της ταινίας: Η ταινία περιέχει σκηνές με λίγη δράση, σκηνές με περισσότερη δράση και σκηνές με εξαιρετικά γρήγορη δράση. Όλα αυτά σχετίζονται με την heavy-tailed κατανομή του κωδικοποιημένου video.

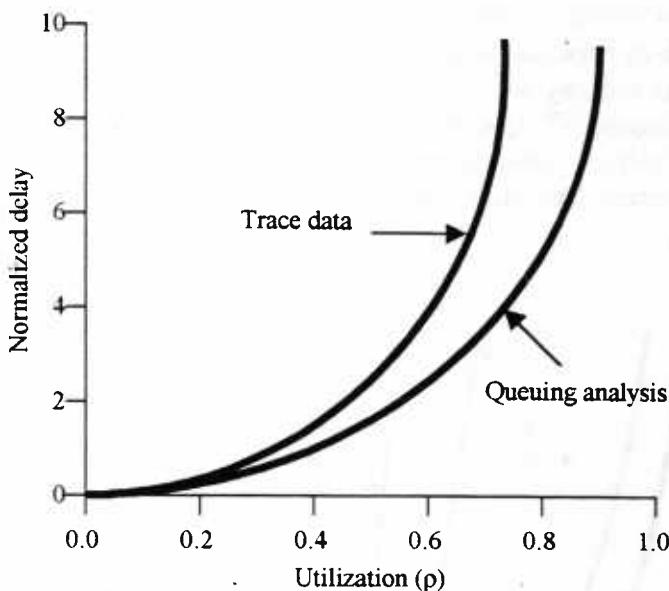
Τα αποτελέσματα μιας ευρύτερης ανάλυσης έδειξαν ότι το χαρακτηριστικό της εξάρτησης μακρού χρόνου (long-range dependence) κληρονομείται από την video κίνηση μεταβλητού ρυθμού bits, και είναι ανεξάρτητη από τους αλγορίθμους κωδικοποίησης και το επιπέδο δράσης σκηνών.

## 6. ΣΥΝΕΠΕΙΕΣ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ ΣΤΗΝ ΑΠΟΔΟΣΗ ΔΙΚΤΥΩΝ

Τα σημερινά δίκτυα και οι υπολογιστές υψηλών ταχυτήτων παρέχουν τη δυνατότητα της γρήγορης συλλογής και ανάλυσης μεγάλου όγκου δεδομένων κίνησης, έτσι ώστε να διαπιστωθεί η ύπαρξη ή όχι αυτό-ομοιότητας σε ένα συγκεκριμένο περιβάλλον και να εκτιμηθούν οι παράμετροί του. Το πραγματικά δύσκολο σημείο συνίσταται στην κατανόηση των συνεπειών της αυτό-ομοιότητας στην απόδοση των δικτύων.

### 6.1 ΑΝΑΛΥΣΗ ETHERNET & ISDN ΔΕΔΟΜΕΝΩΝ ΚΙΝΗΣΗΣ

Για την έρευνα που δημοσιεύτηκε στα [ERRA96] και [ERRA94], χρησιμοποιήθηκε ένας μεγάλος όγκος Ethernet και ISDN δεδομένων κίνησης. Τα αποτελέσματα της έρευνας απέδειξαν ότι υπάρχει σημαντική διαφορά ανάμεσα στον πραγματικό χρόνο αναμονής πακέτων και στην εκτίμηση του χρόνου αναμονής που προκύπτει χρησιμοποιώντας τη συμβατική θεωρία αναμονής, όπως φαίνεται στο Σχήμα 3.5. Έτσι, σύμφωνα με τα αποτελέσματα της συμβατικής ανάλυσης ουρών αναμονής όταν ο βαθμός χρήσης ενός εξυπηρετητή υπερβεί το 80%, τότε ο χρόνος αναμονής των πακέτων σημειώνει μια απότομη αύξηση. Αντίθετα, σύμφωνα με τα πραγματικά δεδομένα κίνησης, ο χρόνος αναμονής των πακέτων σημειώσει απότομη αύξηση όταν ο βαθμός χρήσης του εξυπηρετητή έφτασε στο ποσοστό των 60%.



Σχήμα 3.5 Αποτελέσματα μέσου χρόνου αναμονής.

## 6.2 ΑΠΑΙΤΗΣΕΙΣ ΧΩΡΗΤΙΚΟΤΗΤΑΣ

Οι δύο προηγούμενες αναλύσεις βασίζονται στη σύγκριση της συμπεριφοράς πραγματικών δεδομένων κίνησης με μοντέλα τα οποία στηρίζονται στις υποθέσεις Poisson. Στο άρθρο [NORR94] παρουσιάζεται μια προσπάθεια ανάπτυξης ενός αξιόπιστου αναλυτικού μοντέλου της αυτό-όμοιας συμπεριφοράς.

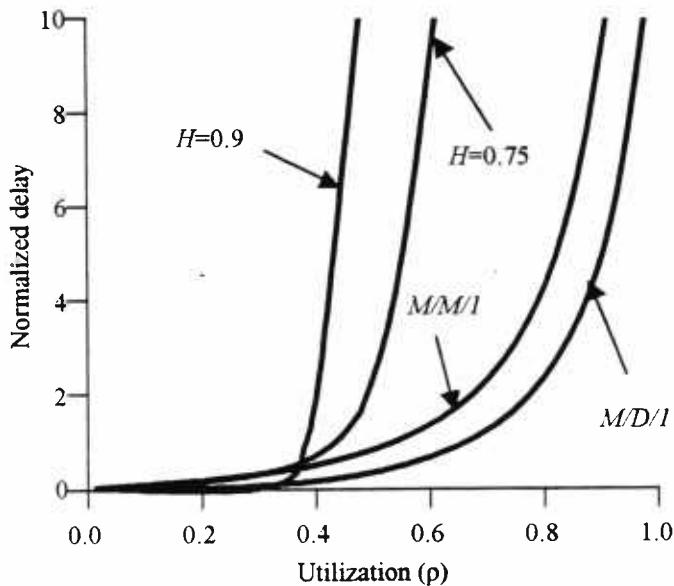
Ο Νορτος χρησιμοποίησε μια αυτό-όμοια διαδικασία προκειμένου να κατασκευάσει ένα μοντέλο τροφοδοσίας ενός ενταμιευτή μη πεπερασμένης χωρητικότητας για τον οποίο ίσχυαν σταθεροί χρόνοι εξυπηρέτησης. Ένα σημαντικό συμπέρασμα το οποίο αφορά άμεσα τις επιπτώσεις απόδοσης είναι το εξής: Κάτω από συγκεκριμένες υποθέσεις, οι απαιτήσεις χωρητικότητας του ενταμιευτή  $q$  ως συνάρτηση του συντελεστή χρησιμοποίησης  $\rho$  ικανοποιεί τη σχέση:

$$q = \frac{\rho^{1/2(1-H)}}{(1-\rho)^{H/(1-H)}}$$

όπου  $H$  είναι η παράμερος Hurst. Για  $H=0.5$ , αυτή η σχέση απλοποιείται στην  $q=\rho/(1-\rho)$ , η οποία από τη θεωρία ουρών αναμονής χαρακτηρίζει ένα σύστημα με εκθετικούς χρόνους μεταξύ των αφίξεων και εκθετικούς χρόνους εξυπηρέτησης (M/M/1). Για ένα σύστημα με σταθερούς χρόνους εξυπηρέτησης (M/D/1), ισχύει η σχέση:

$$q = \frac{\rho}{1-\rho} - \frac{\rho^2}{2(1-\rho)}$$

στο Σχήμα 3.6 εμφανίζονται τα αποτελέσματα για  $H=0.9$  και  $0.75$  τα οποία συγκρίνονται με τις περιπτώσεις  $M/M/1$  και  $M/D/1$ . Όπως φαίνεται οι απαιτήσεις χωρητικότητας του ενταμιευτή αρχίζουν να αυξάνονται απότομα σε χαμηλά επίπεδα του βαθμού χρήσης για μεγάλες τιμές του  $H$  (ύπαρξη αυτό-ομοιότητας). Το γεγονός αυτό έχει την εξής συνέπεια στον σχεδιασμό των ενταμιευτών: Σε περιπτώσεις υψηλού βαθμού χρήσης, απαιτούνται ενταμιευτές πολύ υψηλότερης χωρητικότητας από αυτούς που προτείνονται από την κλασσική θεωρία ουρών αναμονής.



Σχήμα 3.6 Αυτό-όμοιο μοντέλο αποθήκευσης.

## 7. Η ΕΦΑΡΜΟΣΙΜΟΤΗΤΑ ΤΩΝ ΑΥΤΟ-ΟΜΟΙΩΝ ΜΟΝΤΕΛΩΝ ΚΙΝΗΣΗΣ

Τα αποτελέσματα πολλών ερευνών αποδεικνύουν ότι η κίνηση ενός μεγάλου αριθμού δικτύων χαρακτηρίζεται από ένα αυτό-όμοιο πρότυπο (pattern). Το ερώτημα που προκύπτει είναι πόσο “έγκυρα” είναι αυτού του είδους τα υποδείγματα κίνησης και κάτω από ποιές συνθήκες η ύπαρξη του φαινομένου της αυτό-ομοιότητας πρέπει να ληφθεί υπ’ όψη κατά τη διεξαγωγή ανάλυσης απόδοσης (performance analysis). Τα ζητήματα αυτά αποτελούν τρέχοντα αντικείμενα έρευνας.

Τα αποτελέσματα του συνεδρίου SIGMETRICS 1995 σχετικά με την αξιοποίηση των αυτό-όμοιων μοντέλων αναφέρουν ότι σε ορισμένα περιβάλλοντα δικτύων η επίδραση του φαινομένου της αυτό-ομοιότητας επηρεάζει σημαντικά την απόδοση (performance) των δικτύων ενώ σε άλλα όχι. Για παράδειγμα στο άρθρο [RYU96] παραθέτονται στοιχεία από τα οποία προκύπτουν ότι η αυτό-ομοιότητα μπορεί να αγνοηθεί στην περίπτωση μεταφοράς VBR (Variable Bit Rate) κίνησης πάνω από ATM δίκτυα όσον αφορά τον σχεδιασμό της χωρητικότητας των ενταμιευτών (buffers). Στο άρθρο [LIVN93] υποστηρίζεται ότι σε πολλές περιπτώσεις, η ύπαρξη αυτό-ομοιότητας είτε στους χρόνους μεταξύ των αφίξεων, είτε στους χρόνους εξυπηρέτησης, μπορεί να έχει σημαντική επίδραση στις ουρές αναμονής.

Μια οπτική η οποία μπορεί να αποδειχθεί σημαντική δίδεται στο [RYU97]. Οι συγγραφείς κάνουν έναν διαχωρισμό μεταξύ των εννοιών της αυτό-ομοιότητας σε επίπεδο εφαρμογής και σε επίπεδο δικτύου. Η αυτό-ομοιότητα επιπέδου εφαρμογής προέρχεται από μια πηγή η οποία παρουσιάζει αυτό-ομοιότητα για ένα μεγάλο εύρος κλιμάκων χρόνου, χωρίς καμμία αλληλεπίδραση με το δίκτυο. Με αυτόν τον τρόπο η αυτό-ομοιότητα κληρονομείται από την πηγή των δεδομένων. Η αυτό-ομοιότητα επιπέδου δικτύου εμφανίζεται ως αποτέλεσμα ενός αριθμού αλληλεπιδράσεων με το δίκτυο (ή το Internet).

Όπως έχει ήδη αναφερθεί ο Ryu υποστηρίζει ότι η αυτό-ομοιότητα της VBR κίνησης μπορεί να αγνοηθεί κατά τον σχεδιασμό της χωρητικότητας των ενταμιευτών. Κάτι τέτοιο μπορεί να υποδεικνύει ότι, τουλάχιστον σε ορισμένες περιπτώσεις, η αυτό-ομοιότητα επιπέδου εφαρμογής (κλασσικό αντιπρόσωπο της οποίας αποτελεί η VBR κίνηση), πρέπει να «χρησιμοποιηθεί» με διαφορετικό τρόπο από την αυτό-ομοιότητα επιπέδου δικτύου. Επίσης, επειδή η συμπεριφορά της κίνησης η οποία είναι αυτό-όμοια σε επίπεδο εφαρμογής δεν οφείλει τη ύπαρξη της στις συνθήκες του δικτύου κάτω από το οποίο μεταφέρεται, μπορεί να εξεταστεί κάτω από την θεώρηση του ελέγχου αποδοχής (admission control) αλλά και της ποιότητας υπηρεσίας (QoS). Από την άλλη πλευρά, η αυτό-ομοιότητα της κίνησης η οποία προκαλείται λόγω δικτύου, επηρεάζεται από ένα πλήθος παραγόντων όπως ο βαθμός συμφόρησης, ο αριθμός των ταυτόχρονων χρηστών, το μέγεθος του αρχείου του αιτείται, οι πολιτικές επαναμετάδοσης κλπ. Το γεγονός αυτό δυσκολεύει την προσπάθεια εξαγωγής συμπερασμάτων τα οποία σχετίζονται με την απόδοση. Τέλος, το πόσο σημαντικός μπορεί να είναι ο διαχωρισμός μεταξύ της αυτό-ομοιότητας επιπέδου εφαρμογής και δικτύου, αποτελεί ανοικτό αντικείμενο έρευνας.

## ΚΕΦΑΛΑΙΟ 4

### ΕΚΤΙΜΗΣΗ ΒΑΘΜΟΥ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ

#### 1. ΜΕΘΟΔΟΙ ΕΚΤΙΜΗΣΗΣ ΑΥΤΟ-ΟΜΟΙΩΝ ΔΕΔΟΜΕΝΩΝ ΚΙΝΗΣΗΣ

Πολλές προσεγγίσεις έχουν αναπτυχθεί προκειμένου να προσδιοριστεί εάν μια χρονολογική σειρά πραγματικών δεδομένων είναι αυτό-όμοια, και αν όντως είναι, να εκτιμηθεί η παράμετρος αυτό-ομοιότητας  $H$ . Σε αυτήν την ενότητα παρουσιάζονται δύο από τις πιο γνωστές τέτοιες προσεγγίσεις.

##### 1.1 ΜΕΘΟΔΟΣ ΑΘΡΟΙΣΤΙΚΩΝ ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ

Όπως έχει ήδη αναφερθεί οι διακυμάνσεις των αθροιστικών χρονολογικών σειρών  $\mathbf{x}^{(m)}$  μιας αυτό-όμοιας διαδικασίας, ικανοποιούν την ακόλουθη σχέση για μεγάλα  $m$ :

$$Var(\mathbf{x}^{(m)}) = \frac{Var(\mathbf{x})}{m^\beta}$$

όπου  $H=1-(\beta/2)$ . Η σχέση αυτή μπορεί να ξαναγραφεί ως:

$$\log[Var(\mathbf{x}^{(m)})] \sim \log[Var(\mathbf{x})] - \beta \log(m)$$

Επειδή η  $\log[Var(\mathbf{x})]$  είναι μια σταθερά ανεξάρτητη από το  $m$ , αν απεικονίσουμε γραφικά τα  $Var(\mathbf{x}^{(m)})$  και  $m$  σε ένα log-log γράφημα, το αποτέλεσμα θα είναι μια ευθεία γραμμή με κλίση  $-\beta$ . Το γράφημα κατασκευάζεται από μια χρονολογική σειρά  $\mathbf{x}(t)$  δημιουργώντας τις αθροιστικές διαδικασίες σε διαφορετικά επίπεδα άθροισης  $m$  και υπολογίζοντας τις διακυμάνσεις. Για τιμές κλίσης μεταξύ  $-1$  και  $0$ , υπάρχει ένδειξη αυτό-ομοιότητας.

##### 1.2 R/S ΑΝΑΛΥΣΗ

Ο σκοπός της διεξαγωγής R/S ανάλυσης σε πραγματικά δεδομένα είναι η διαπίστωση του βαθμού της αυτό-ομοιότητας (παράμετρος Hurst) της διαδικασίας η οποία παρήγαγε τα πραγματικά δεδομένα. Πρακτικά, πρόκειται για μια γραφική, ευρετική

προσέγγιση η οποία προσπαθεί να εκμεταλλευτεί όσο το δυνατό περισσότερο την πληροφορία που κρύβεται στα πραγματικά δεδομένα.

Για ένα σύνολο παρατηρήσεων ( $X_k : k = 1, 2, 3, \dots, N$ ) με μέση τιμή δείγματος  $X(N)$  και διακύμανση δείγματος  $S^2(N)$ , το ανακλιμακούμενο προσαρμοσμένο εύρος ή  $R/S$  στατιστική, δίνεται από τον τύπο:

$$R(N)/S(N) = 1/S(N) [\max(0, W_1, W_2, \dots, W_N) - \min(0, W_1, W_2, \dots, W_N)], \quad (4.1)$$

με  $W_k = (X_1 + X_2 + \dots + X_k) - kX(N), k = 1, 2, \dots, N$ .

Ο Hurst έδειξε ότι πολλές χρονολογικές σειρές, ικανοποιούν τη σχέση:

$$E[R(N)/S(N)] \sim cN^H, \quad \text{όταν } N \rightarrow \infty,$$

με *Hurst παράμετρο*  $H > 0.5$ , και  $c$  μια πεπερασμένη θετική σταθερά ανεξάρτητη από το  $N$ . Αντίθετα, αν οι παρατηρήσεις  $X_k$  προέρχονται από ένα short-range dependent μοντέλο, τότε οι Mandelbrot και Van Ness έδειξαν ότι:

$$E[R(N)/S(N)] \sim dN^{0.5}, \quad \text{όταν } N \rightarrow \infty,$$

Όπου  $d$  μια πεπερασμένη θετική σταθερά, ανεξάρτητη από το  $N$ .

Συγκεκριμένα η διαδικασία διεξαγωγής  $R/S$  ανάλυσης σε ένα δείγμα  $N$  παρατηρήσεων ( $X_k : k = 1, 2, 3, \dots, N$ ) έχει ως εξής: Υποδιαιρούμε το δείγμα σε  $K$  μη επικαλυπτόμενες ομάδες των  $m=N/K$  στοιχείων και υπολογίζουμε το ανακλιμακούμενο προσαρμοσμένο εύρος  $R(t_i, m) / S(t_i, m)$  για κάθε ένα από τα  $t_1 = 1$ ,  $t_2 = N/K + 1$ ,  $t_3 = 2N/K + 1, \dots$  τα οποία ικανοποιούν τη συνθήκη  $(t_i - 1) + m \leq N$ .

Εδώ, η  $R/S$  στατιστική  $R(t_i, m) / S(t_i, m)$  ορίζεται όπως στη σχέση (4.1), αντικαθιστώντας το  $W_k$  με  $W_{ti} + k - W_{ti}$ , και ορίζοντας το  $S^2(t_i, m)$  ως τη δειγματική διακύμανση των  $X_{ti+1}, X_{ti+2}, \dots, X_{ti+1}$ . Έτσι, για ένα ορισμένο επίπεδο άθροισης  $m$ , προκύπτουν  $K$  τιμές της  $R/S$ . Στη συνέχεια λογαριθμούμε ως προς  $m$ . Παριστάνοντας γραφικά τις τιμές  $\log(R(t_i, m) / S(t_i, m))$  σε σχέση με τις τιμές  $\log(m)$ , προκύπτει το ανακλιμακούμενο προσαρμοσμένου εύρους διάγραμμα (rescaled adjusted range plot) -το οποίο ονομάζεται και *Rox* διάγραμμα των  $R/S$ . Χρησιμοποιώντας την μέθοδο των ελαχίστων τετραγώνων, προσαρμόζουμε ένα ευθύγραμμο τμήμα μεταξύ των σημείων του διαγράμματος. Η κλίση του ευθύγραμμου τμήματος δίνει μια εκτίμηση της Hurst παραμέτρου.

Τα παραπάνω χαρακτηριστικά και ιδιότητες ανακαλύφθηκαν από τον Hurst στην προσπάθειά του να ομαλοποιήσει τη στάθμη του ποταμού Νείλου. Συγκεκριμένα η ανακάλυψή του μπορεί να περιγραφεί ως εξής: Ας υποθέσουμε ότι έχουμε μια δεξαμενή και θέλουμε να εξασφαλίσουμε ομοιόμορφη στάθμη νερού για την δεξαμενή για το χρονικό διάστημα από  $t$  έως  $t+n$ . Αυτό σημαίνει ότι η δεξαμενή θα πρέπει να περιέχει την ίδια ποσότητα νερού το χρόνο  $t$  και  $t+n$ . Συμβολίζουμε με  $Y_i$  την εισροή νερού στο χρόνο  $i$  και με  $X_j = \sum_{i=1}^j Y_i$  τη συνολική εισροή νερού μέχρι το

χρόνο  $j$ . Μπορεί να αποδειχθεί ότι η χωρητικότητα που ικανοποιεί τους παραπάνω περιορισμούς είναι ίση με:

$$R(t,m) = \max_{0 \leq i \leq m} \left[ X_{t+i} - X_t - \frac{i}{m} (X_{t+m} - X_t) \right] - \min_{0 \leq i \leq m} \left[ X_{t+i} - X_t - \frac{i}{m} (X_{t+m} - X_t) \right]$$

Το  $R(t,m)$  είναι το αναπροσαρμοσμένο εύρος και προκειμένου να γίνει ανεξάρτητο κλίμακας διαιρείται με:

$$S(t,m) = \sqrt{\frac{1}{m} \sum_{i=t+1}^{t+m} \left( Y_i - \bar{Y}_{t,m} \right)^2}$$

όπου  $\bar{Y}_{t,m} = \frac{1}{m} \sum_{i=t+1}^{t+m} Y_i^2$ .

Ο λόγος  $R/S = \frac{R(t,m)}{S(t,m)}$  αποτελεί το ανακλιμακούμενο προσαρμοσμένο εύρος

(rescaled adjusted range ή R/S-statistic). Ο Hurst απεικόνισε γραφικά τις λογαριθμημένες τιμές του  $R/S$  για διάφορες τιμές του  $m$ . Έτσι, παρατήρησε ότι για μεγάλες τιμές του  $m$ , οι τιμές των  $\log R/S$ , συσπειρώνονται γύρω από μια ευθεία γραμμή με κλίση μεγαλύτερη από 0.5. Σε όρους πιθανοτήτων αυτό σημαίνει ότι για μεγάλα  $m$  ισχύει:

$$\log E[R/S] \approx a + H \log m, \text{ με } H > \frac{1}{2}$$

## 2. ΥΛΟΠΟΙΗΣΗ ΕΡΓΑΛΕΙΩΝ ΕΚΤΙΜΗΣΗΣ ΤΟΥ ΒΑΘΜΟΥ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ

Η πραγματοποίηση μετρήσεων στο δίκτυο δεδομένων του Πανεπιστημίου είχε ως αποτέλεσμα τη δημιουργία ενός πλήθους χρονολογικών σειρών δεδομένων κίνησης. Στόχος της παρούσας εργασίας είναι η μελέτη των δεδομένων κίνησης σε σχέση με το φαινόμενο της αυτό-ομοιότητας, με την προοπτική της συμβολής στην εξαγωγή συμπερασμάτων σχετικά με την συμπεριφορά και τα χαρακτηριστικά της κίνησης των δικτύων υπολογιστών.

Για το σκοπό αυτό, και προκειμένου να εκτιμηθεί ο βαθμός αυτό-ομοιότητας των χρονολογικών σειρών υλοποιήθηκαν δύο διαδικασίες εκτίμησης της παραμέτρου  $H$  (βαθμός αυτό-ομοιότητας) χρησιμοποιώντας τη δυνατότητα κατασκευής μακροεντολών στο πακέτο επεξεργασίας λογιστικών φύλλων Excel.

Η πρώτη διαδικασία ονομάζεται *AggregatedSeries* και αποτελεί μια προσπάθεια υλοποίησης της μεθόδου εκτίμησης της παραμέτρου  $H$  σύμφωνα με την μέθοδο των Αθροιστικών Χρονολογικών Σειρών. Η υλοποίηση της μεθόδου περιλαμβάνει και την δημιουργία του διαγράμματος Διακύμανσης-Χρόνου.

Η δεύτερη διαδικασία ονομάζεται *RS\_Statistic* και υλοποιεί την μέθοδο εκτίμησης της παραμέτρου  $H$  σύμφωνα με την μέθοδο της *R/S* στατιστικής, ενώ υπολογίζει και τα *POX* διαγράμματα της μεθόδου.

Στις επόμενες παραγράφους ακολουθεί η περιγραφή των υλοποιήσεων των διαδικασιών, η διαδικασία εκτέλεσης τους στο περιβάλλον του Excel, καθώς και η ερμηνεία των αποτελεσμάτων της εκτέλεσης. Σημειώνεται ότι ο κώδικας των διαδικασιών παρατίθεται στο Παράρτημα 2.

## 2.1 ΠΕΡΙΓΡΑΦΗ ΤΗΣ ΔΙΑΔΙΚΑΣΙΑΣ *AGGREGATED SERIES*

Όπως έχει ήδη αναφερθεί μια από τις μεθόδους εκτίμησης του βαθμού αυτό-ομοιότητας χρονολογικών σειρών είναι η μέθοδος των αθροιστικών χρονολογικών σειρών (*Aggregated Series*). Από την μέθοδο αυτή προκύπτει το γράφημα διακύμανσης-χρόνου. Αυτό παρέχει μια οπτική ένδειξη για την ύπαρξη αυτό-ομοιότητας.

Για την υλοποίηση της μεθόδου των *Aθροιστικών Χρονολογικών Σειρών*, και την κατασκευή των διαγραμμάτων διακύμανσης-χρόνου, ακολουθήθηκε η εξής διαδικασία:

Η χρονολογική σειρά  $x(t)$  εισάγεται στο πακέτο επεξεργασίας λογιστικών φύλλων Excel. Βάσει της  $x(t)$  δημιουργούνται οι αθροιστικές χρονολογικές σειρές  $x^{(m)} = \{x_k^{(m)}, k = 0, 1, 2, \dots\}$  για διάφορα επίπεδα άθροισης  $m = 100, 200, 400, 800, 1600$  (ανάλογα με το μήκος της χρονολογικής σειράς). Αυτές τοποθετούνται στις διπλανές στήλες του λογιστικού φύλλου που περιέχει την  $x(t)$ .

Γενικά, το  $k$ -οστό στοιχείο της  $m$ -αθροιστικής χρονολογικής σειράς, ορίζεται ως:

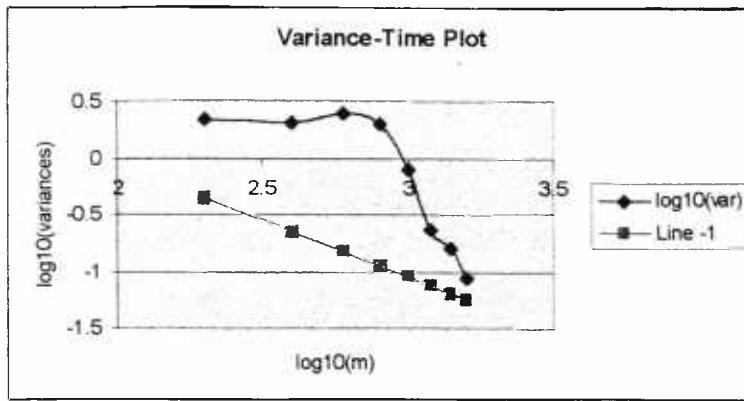
$$x_k^{(m)} = \frac{1}{m} \sum_{i=km-(m-1)}^{km} x_i$$

Έτσι για παράδειγμα το 10 στοιχείο της  $x^{(10)}$ , είναι το άθροισμα των 10 πρώτων στοιχείων της  $x(t)$  διαιρεμένο κατά 10, το 20 στοιχείο της  $x^{(10)}$ , είναι το άθροισμα των 10 επόμενων στοιχείων της  $x(t)$  διαιρεμένο κατά 10 κλπ.

Για κάθε αθροιστική χρονολογική σειρά  $x^{(m)}$ , υπολογίζεται η διακύμανσή της  $\text{Var}(x^{(m)})$ , καθώς και η λογαριθμημένη διακύμανσή της  $\log[\text{Var}(x^{(m)})]$ .

### 2.1.1 Διάγραμμα Διακύμανσης-Χρόνου

Απεικονίζοντας γραφικά τις τιμές  $\log[\text{Var}(x^{(m)})]$  σε σχέση με τις τιμές  $\log(m)$ , προκύπτει το διάγραμμα διακύμανσης-χρόνου της  $x(t)$ . Παρακάτω παρατίθεται ένα παράδειγμα γραφήματος διακύμανσης-χρόνου όπως προκύπτει από την εκτέλεση της διαδικασίας *AggregatedSeries*.



Σχήμα 4.1 Διάγραμμα διακύμανσης-χρόνου (Variance-Time Plot).

Όπως έχει ήδη αναφερθεί η σχέση που συνδέει τα  $\text{Var}(\mathbf{x}^{(m)})$ ,  $\text{Var}(\mathbf{x})$ , και  $m$ , είναι η εξής:

$$\log[\text{Var}(\mathbf{x}^{(m)})] \sim \log[\text{Var}(\mathbf{x})] - \beta \log(m) \quad (4.2)$$

Η γραμμή με κλίση  $-1$  (Line  $-1$ ) που εμφανίζεται στο παραπάνω διάγραμμα διακύμανσης-χρόνου, έχει την εξής σημασία: Είναι γνωστό ότι ισχύει η σχέση  $H=1-\beta/2$ . Ετσι, όταν  $\beta=-1$ , τότε  $H=0.5$ . Θεωρώντας ότι αυτό-ομοιότητα εμφανίζεται όταν  $H>0.5$ , τότε, αντικαθιστώντας στην (4.2) όπου  $\beta=-1$ , ισχύει η σχέση:

$$\log[\text{Var}(\mathbf{x}^{(m)})] \sim \log[\text{Var}(\mathbf{x})] + \log(m) \quad (4.3)$$

Έτσι, απεικονίζοντας γραφικά όλα τα ζεύγη  $(\log(m), \log[\text{Var}(\mathbf{x})] + \log(m))$ , προκύπτει η γραμμή με κλίση  $-1$ , πάνω από την οποία πρέπει να βρίσκονται οι τιμές  $\log[\text{Var}(\mathbf{x}^{(m)})]$ , προκειμένου να ισχύει ότι  $H>0.5$ .

### 2.1.2 Εκτίμηση του βαθμού αυτό-ομοιότητας $H$

Εφαρμόζοντας τη μέθοδο των ελαχίστων τετραγώνων στο διάγραμμα διακύμανσης-χρόνου, προκύπτει μια ευθεία γραμμή με κλίση  $-\beta$  για την οποία ισχύει η σχέση  $H=1-\beta/2$ . Για τιμές κλίσης μεταξύ  $-1$  και  $0$  (δηλ. για  $H$  μεταξύ  $0.5$  και  $1$ ), υπάρχει ένδειξη αυτό-ομοιότητας.

Η εκτίμηση για τον βαθμό της αυτό-ομοιότητας  $H$ , προκύπτει από τον τύπο  $H=1-\beta/2$ , αν όπου  $\beta$ , βάλουμε το μέσο όρο των  $\beta$  που προκύπτουν για κάθε αθροιστική χρονολογική σειρά.

Σημειώνεται ότι το επίπεδο άθροισης  $m$ , πρέπει να ικανοποιεί τη σχέση  $m \leq N/2^1$ , όπου  $N$  το πλήθος των στοιχείων της χρονολογικής σειράς. Η εξήγηση δίνεται μέσα από ένα παράδειγμα: Έστω μια χρονολογική σειρά  $\mathbf{x}$  με  $N=1000$  το πλήθος στοιχείων. Αν υποθέσουμε ότι  $m=600$ , τότε η αθροιστική χρονολογική σειρά  $\mathbf{x}^{(600)}$ , περιλαμβάνει ένα μόνο στοιχείο  $\mathbf{x}_1^{(600)} = \frac{1}{600} \sum_{i=1}^{600} \mathbf{x}_i$ . Όμως, επειδή δεν ορίζεται η διακύμανση μιας σταθεράς, θα πρέπει κάθε αθροιστική χρονολογική σειρά να έχει τουλάχιστον δύο στοιχεία, ώστε να μπορεί να υπολογιστεί η διακύμανσή της.

## 2.2 ΑΠΟΤΕΛΕΣΜΑΤΑ ΕΚΤΕΛΕΣΗΣ ΤΗΣ AGGREGATED SERIES

Στο τρέχον φύλλο εργασίας εισάγεται (import) η χρονολογική σειρά. Αυτή πρέπει να αρχίζει από τη δεύτερη γραμμή και καταλαμβάνει τις δύο πρώτες στήλες του φύλλου εργασίας.

Η εκτέλεση της διαδικασίας AggregatedSeries έχει ως συνέπεια τη δημιουργία ενός αριθμού στηλών στο φύλλο εργασίας:

Η πρώτη στήλη έχει τον τίτλο *H Aggr.Series* βρίσκεται στην 4<sup>η</sup> στήλη και σε κάθε κελί της περιλαμβάνει την εκτίμηση της παραμέτρου  $H$  για κάθε ένα από τα επίπεδα άθροισης. Το τελευταίο στοιχείο της στήλης αποτελεί την τελική εκτίμηση της  $H$  (προκύπτει από την μέση τιμή των εκτιμήσεων  $H$  για όλα τα επίπεδα άθροισης), και εμφανίζεται με μπλε χρώμα.

Οι επόμενες τρεις στήλες που προκύπτουν από την εκτέλεση της διαδικασίας αφορούν την κατασκευή του διαγράμματος Διακύμανσης-Χρόνου. Στην 7<sup>η</sup> στήλη του φύλλου εργασίας εμφανίζονται οι τιμές του οριζόντιου άξονα του διαγράμματος Διακύμανσης-Χρόνου. Ουσιαστικά πρόκειται για τις λογαριθμημένες τιμές των επιπέδων άθροισης  $m$ . Στην 8<sup>η</sup> στήλη εμφανίζονται οι τιμές του κατακόρυφου άξονα του διαγράμματος οι οποίες περιλαμβάνουν τις λογαριθμημένες διακυμάνσεις των αθροιστικών χρονολογικών σειρών. Η δημιουργία ενός γραφήματος με τα σημεία που ορίζουν οι στήλες 7 και 8 έχει ως συνέπεια την κατασκευή του διαγράμματος Διακύμανσης-Χρόνου. Επιπλέον, η απεικόνιση των σημείων που ορίζονται από τις στήλες 7 και 9, στο ίδιο διάγραμμα, εμφανίζει τη γραμμή με κλίση -1, η οποία περιλαμβάνει τα σημεία για τα οποία ισχύει  $H=0.5$ .

<sup>1</sup> Παρόμοιος περιορισμός δεν ισχύει για την μέθοδο της R/S ανάλυσης, όπου μια χρονολογική σειρά των  $N$  στοιχείων μπορεί να «διαιρεθεί» σε  $k=1$  τμήματα των  $m=N$  στοιχείων (όλα τα στοιχεία της χρονολογικής σειράς).

A	B	C	D	E	F	G	H	I
1	Time Series		H Aggr.Series	Var-Time Diagram			Line -1	
2	9:32:59	0	0,644222	2,60206	0,233023	-0,51753		
3	9:33:09	0	0,669767	2,90309	0,167141	-0,81856		
4	9:33:19	0	0,444769	3,079181	-1,33478	-0,99465		
5	9:33:29	0	0,625262	3,20412	-0,31688	-1,11959		
6	9:33:39	0						
7	9:33:49	6,40641	0,596005					
8	9:33:59	0						

Από την 18<sup>η</sup> στήλη εμφανίζονται οι αθροιστικές χρονολογικές σειρές οι οποίες φέρουν τον τίτλο  $x(m)$ , όπου m το επίπεδο άθροισης που αντιστοιχεί στην αθροιστική χρονολογική σειρά. Δύο στήλες μετά την τελευταία αθροιστική χρονολογική σειρά εμφανίζεται η στήλη με τις εκτιμήσεις του β οι οποίες αντιστοιχούν σε κάθε επίπεδο άθροισης m.

S	T	U	V	W	X	Y
1	$x(400)$	$x(800)$	$x(1200)$	$x(1600)$	Est. $\beta$	
2	3,19292	3,29077	3,67664	4,39326	0,71156	
3	3,38862	5,49576	3,98081	3,41135	0,66047	
4	4,44837	2,69964			1,11046	
5	6,54315	4,12305			0,74948	
6	3,10172					
7	2,29756					
8	4,67884					
9	3,56727					

## 2.3 ΠΕΡΙΓΡΑΦΗ ΤΗΣ ΔΙΑΔΙΚΑΣΙΑΣ RS\_STATISTIC

Όπως έχει ήδη αναφερθεί τα αποτελέσματα της εφαρμογής R/S ανάλυσης σε μια χρονολογική σειρά, δίνει μια ένδειξη για το βαθμό ύπαρξης του φαινομένου της αυτό-ομοιότητας στην χρονολογική σειρά.

Για την εφαρμογή R/S ανάλυσης και τη δημιουργία του διαγράμματος ROX των χρονολογικών σειρών οι οποίες περιλαμβάνουν τα στοιχεία των μετρήσεων του δικτύου δεδομένων του Πανεπιστημίου, ακολουθήθηκε η εξής διαδικασία:

Η χρονολογική σειρά  $x(t)$  εισάγεται στο πακέτο επεξεργασίας λογιστικών φύλλων Excel. Η  $x(t)$  υποδιαιρείται σε  $k$  τμήματα των  $N$  στοιχείων. Έτσι αν η  $x(t)$  περιλαμβάνει 500 στοιχεία και  $k=10$ , τότε  $N=50$ . Για κάθε τμήμα  $i$  ( $1 \leq i \leq k$ ) της χρονολογικής σειράς υπολογίζονται και τοποθετούνται σε κελιά η μέση τιμή  $\overline{X(i)}$  καθώς και η τυπική απόκλιση  $sd(i)$  των  $N$  στοιχείων του  $i$  τμήματος.

Για κάθε τμήμα  $i$  της χρονολογικής σειράς δημιουργούνται νέες χρονολογικές σειρές οι οποίες τοποθετούνται σε διτλανές στήλες του φύλλου εργασίας. Η  $j$  τιμή της νέας χρονολογικής σειράς που αντιστοιχεί στο τμήμα  $i$  της  $x(t)$  υπολογίζεται από τον τύπο

$$\sum_{m=(i-1)N+1}^{(i-1)N+(j-1)} x_m - j\overline{X(i)} \quad \text{όπου } i = 1 \dots k \text{ και } j = 1 \dots N.$$

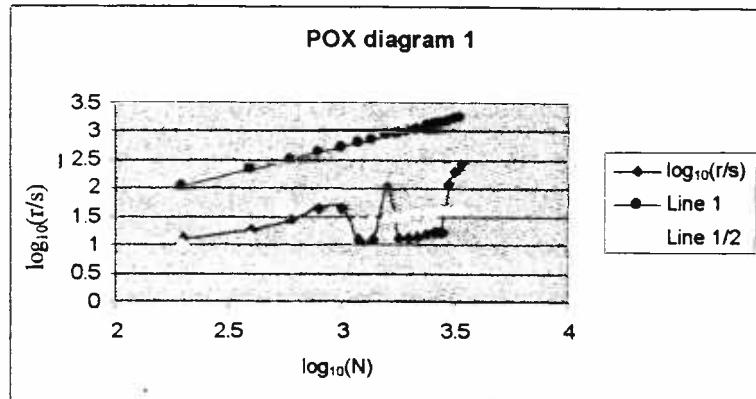
Με αυτό τον τρόπο δημιουργούνται  $k$  νέες χρονολογικές σειρές των  $N$  στοιχείων. Για κάθε μια από της  $k$  νέες χρονολογικές σειρές υπολογίζονται και τοποθετούνται σε κελιά η ελάχιστη ( $min(i)$ ) και η μέγιστη ( $max(i)$ ) τιμή τους. Η R/S στατιστική για κάθε μια από τις  $k$  χρονολογικές σειρές υπολογίζεται από τον τύπο:  $R/S = \frac{max(i) - min(i)}{sd(i)}$  όπου  $i=1 \dots k$ . Λαμβάνοντας τη μέση τιμή των R/S  $ER/S$ , και λογαριθμώντας τη, προκύπτει μια εκτίμηση του βαθμού της αυτό-ομοιότητας  $H$ :

$$\hat{H} = \frac{\log ER/S}{\log N - \log 2} \quad (4.4)$$

Επαναλαμβάνοντας την προηγούμενη διαδικασία για διάφορες τιμές του  $k$  (και συνεπώς και του  $N$ ), προκύπτει μια τιμή του  $\log ER/S$  για κάθε διαφορετική τιμή του  $k$ .

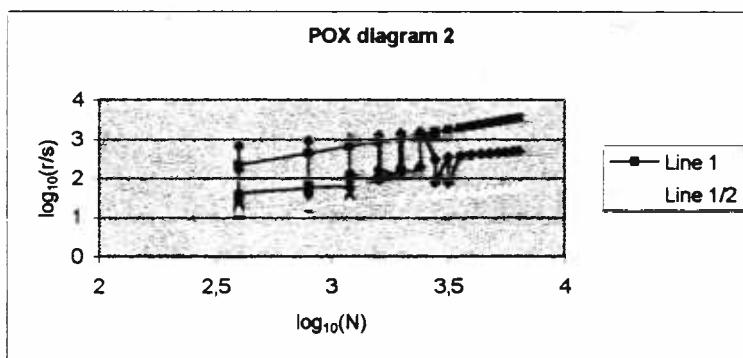
### 2.3.1 Διαγράμματα POX

Απεικονίζοντας σε ένα διάγραμμα τις τιμές των  $\log R/S$  σε σχέση με το  $\log N$ , προκύπτει το διάγραμμα POX τύπου 1. Παρακάτω παρατίθεται η μορφή ενός τέτοιου διαγράμματος, όπως προέκυψε ακολουθώντας την παραπάνω διαδικασία για  $N=200, 400, 600, 800, \dots, 3400$ .



Σχήμα 4.2 Διάγραμμα POX τύπου 1.

Εναλλακτικά, απεικονίζοντας γραφικά τις τιμές των  $\log R/S$  σε σχέση με το  $\log N$ , προκύπτει το διάγραμμα POX τύπου 2. Η διαφορά των δύο διαγραμμάτων έγκειται στο ότι το POX τύπου 2 περιλαμβάνει για κάθε διαφορετικό επίπεδο άθροισης  $N$  το σύνολο των  $k$  τιμών  $R/S$ , και όχι την μέση τιμή αυτών όπως συμβαίνει στο POX τύπου 1. Παρακάτω παρατίθεται η μορφή ενός διαγράμματος POX τύπου 2, για  $N=200, 400, 600, 800, \dots, 3400$ .



Σχήμα 4.3 Διάγραμμα POX τύπου 2.

Όπως έχει ήδη αναφερθεί προκειμένου μια χρονολογική σειρά να χαρακτηρίζεται ως αυτό-όμοια, θα πρέπει να ισχύει για τον βαθμό αυτό-ομοιότητας  $H$ ,  $0.5 \leq H \leq 1$ .

Σε περίπτωση που για κάθε επίπεδο άθροισης  $N$ , η εκτίμηση του βαθμού αυτό-ομοιότητας  $H$  ισούται με 1, τότε από την (4.4) ισχύει  $\log R/S = \log N - \log 2$ . Αν στο POX διάγραμμα τύπου 1 απεικονίσουμε όλα τα σημεία  $(\log N, \log N - \log 2)$ , τότε προκύπτει η γραμμή με κλίση 1. Σε περίπτωση που όλες οι εκτιμήσεις  $H$  μιας χρονολογικής σειράς για κάθε επίπεδο άθροισης  $N$ , είναι ίσες με 1, τότε τα σημεία

( $\log N, \log ER/S$ ) και ( $\log N, \log N - \log 2$ ) συμπίπτουν. Ανάλογα προκύπτει και η γραμμή με κλίση 1 για το POX διάγραμμα τύπου 2.

Αντίστοιχα, για την γραμμή με κλίση  $\frac{1}{2}$  του POX διαγράμματος τύπου 1, ισχύουν τα εξής: Σε περίπτωση που για κάθε επίπεδο άθροισης  $N$ , η εκτίμηση του βαθμού αυτό-ομοιότητας  $\hat{H}$  ισούται με  $1/2$ , τότε από την (4.4) ισχύει  $\log ER/S = 0.5 * (\log N - \log 2)$ . Αν στο POX διάγραμμα τύπου 1 απεικονίσουμε όλα τα σημεία ( $\log N, 0.5 * (\log N - \log 2)$ ), τότε προκύπτει η γραμμή με κλίση  $1/2$ . Σε περίπτωση που όλες οι εκτιμήσεις  $\hat{H}$  μιας χρονολογικής σειράς για κάθε επίπεδο άθροισης  $N$ , είναι ίσες με  $1/2$ , τότε τα σημεία ( $\log N, \log ER/S$ ) και ( $\log N, 0.5 * (\log N - \log 2)$ ) συμπίπτουν. Ανάλογα προκύπτει και η γραμμή με κλίση  $1/2$  για το POX διάγραμμα τύπου 2.

Από τα παραπάνω προκύπτει ότι αν οι τιμές  $\log ER/S$  (ή  $\log R/S$ ) μιας χρονολογικής σειράς βρίσκονται μεταξύ των γραμμών με κλίση 1 και  $\frac{1}{2}$ , (όπως αυτά προκύπτουν από την  $R/S$  ανάλυση), τότε υπάρχει ένδειξη για την ύπαρξη αυτό-ομοιότητας. Ο βαθμός της αυτό-ομοιότητας είναι μεγαλύτερος όσο οι τιμές  $\log ER/S$  (ή  $\log R/S$ ) βρίσκονται πιο κοντά στη γραμμή με κλίση 1, ενώ όσο αυτές προσεγγίζουν την γραμμή με κλίση  $\frac{1}{2}$ , μειώνεται και ο βαθμός αυτό-ομοιότητας.

Η τελική εκτίμηση της παραμέτρου  $H$ , προκύπτει λαμβάνοντας τη μέση τιμή των εκτιμήσεων της  $H$  οι οποίες αντιστοιχούν σε κάθε ένα από τα  $N$ . Αυτές είναι της μορφής:

$$\hat{H} = \frac{\log ER/S}{\log N - \log 2}$$

## 2.4 ΑΠΟΤΕΛΕΣΜΑΤΑ ΕΚΤΕΛΕΣΗΣ ΤΗΣ RS\_STATISTIC

Η εκτέλεση της διαδικασίας *RS\_Statistic* δημιουργεί τις εξής στήλες:

Στην 5<sup>η</sup> στήλη του φύλλου εργασίας η οποία φέρει τον τίτλο *HR/SStatistic* εμφανίζονται οι εκτιμήσεις της παραμέτρου  $H$  για κάθε ένα από τα επίπεδα άθροισης που χρησιμοποιούνται. Το τελευταίο στοιχείο της στήλης αποτελεί την τελική εκτίμηση της  $H$  η οποία προκύπτει από την μέση τιμή των εκτιμήσεων  $H$  για όλα τα επίπεδα άθροισης και εμφανίζεται με μπλε χρώμα.

	A	B	C	D	E
1	Time Series				H RS Stat
2	9032069	0			0.68575
3	9033009	0			0.70737
4	9033019	0			0.68489
5	9033029	0			0.69452
6	9033039	0			0.71906
7	9033049	6,406			0.72106
8	9033069	0			0.69326
9	9034009	0			0.6879
10	9034019	0			
11	9034029	0			0.69923
12	9034039	0			

Οι στήλες 11-16 αφορούν τα POX διαγράμματα τύπου 1 και 2 της μεθόδου. Η 11<sup>η</sup> στήλη περιλαμβάνει τις τιμές του οριζόντιου άξονα του διαγράμματος POX τύπου 2, ενώ η 12<sup>η</sup> τις τιμές του κατακόρυφου άξονα. Ουσιαστικά πρόκειται για τα ζεύγη τιμών ( $\log(m), \log(r/s)$ ). Η 15<sup>η</sup> στήλη περιλαμβάνει τις τιμές του οριζόντιου άξονα του διαγράμματος POX τύπου 1, ενώ η 16<sup>η</sup> τις τιμές του κατακόρυφου άξονα του ίδιου διαγράμματος. Τα σημεία που σχηματίζουν οι δύο στήλες είναι τα ( $\log(m), \log(E/r/s)$ ).

Η απεικόνιση των σημείων που ορίζονται από τις στήλες 15 - 13 και 15 - 14 στα POX διαγράμματα έχει ως συνέπεια την εμφάνιση των γραμμών με κλίση 1 και  $\frac{1}{2}$ . Υπενθυμίζεται ότι η γραμμή με κλίση 1 αντιπροσωπεύει τα σημεία για τα οποία ισχύει  $H=1$ , ενώ η γραμμή με κλίση  $\frac{1}{2}$  τα σημεία για τα οποία  $H=0.5$ .

K	L	M	N	O	P	
POX	Diagram1	Line 1	Line 1/2	POX	Diagram2	
2,6021	1,48436	2,30103	1,15051	2,6021	1,57793	
2,6021	1,3394	2,60206	1,30103	2,9031	1,84061	
2,6021	1,80905	2,77815	1,38908	3,0792	1,90273	
2,6021	1,8074	2,90309	1,45154	3,2041	2,01626	
2,6021	1,21312	3	1,5	3,301	2,15719	
2,6021	1,02361	3,07918	1,53959	3,3802	2,22028	
2,6021	1,82936	3,14813	1,57306	3,4472	2,18108	
2,6021	1,43692	3,20412	1,60206	3,5051	2,20412	
2,9031	1,69296					

Από την 18<sup>η</sup> στήλη εμφανίζονται χρονολογικές σειρές οι οποίες φέρουν τον τίτλο  $\alpha \Rightarrow \beta$  όπου  $[\alpha, \beta]$  το διάστημα τιμών της υπό ανάλυσης χρονολογικής σειράς βάσει του οποίου κατασκευάζεται η χρονολογική σειρά. Έτσι, αν η υπό ανάλυση χρονολογική σειρά έχει N το πλήθος στοιχεία, τότε για κάθε επίπεδο άθροισης m δημιουργούνται  $N/m$  χρονολογικές σειρές των m στοιχείων η κάθε μία.

	AB	AC	AD	AE
1	1 έως 800	801 έως 1600	1601 έως 2400	2401 έως 3200
2	□3,2908	□5,49576001	□2,6998408	2,278946562
3	□6,5815	□10,99152	1,0071284	□1,846106875
4	□9,8723	□16,48728	24,7338876	□6,969180313
5	□13,163	□15,5830401	22,0342468	□10,09221375
6	□16,454	□21,0788001	39,434606	□14,21526719
7	□13,338	□26,5746601	36,7349652	□18,33832063
8	□16,629	□32,0703201	40,4353244	□18,06137406
9	□19,92	□37,5660801	44,1420936	□20,1844275
10	□23,211	□43,0618401	41,4424628	□24,30748094
11	□26,501	□42,1576001	38,742812	□28,43053438
12	□29,792	□47,8533601	36,0431712	□32,55358781
13	□26,876	□53,1491202	33,3435304	□36,67664125

Τέλος, όσον αφορά και τις δύο διαδικασίες *AggregatedSeries* και *RS\_Statistic*, σημειώνονται τα εξής:

- Τα στοιχεία των εκτιμήσεων και των διαγραμμάτων της μίας διαδικασίας συνεχίζουν να εμφανίζονται και μετά την εκτέλεση της άλλης. Αντίθετα, οι στήλες των χρονολογικών σειρών οι οποίες δημιουργούνται από την εκτέλεση μιας διαδικασίας αντικαθιστούνται από τις στήλες των χρονολογικών σειρών της διαδικασίας που εκτελέστηκε πιο πρόσφατα.
- Για την υλοποίηση των δύο παραπάνω μεθόδων χρησιμοποιήθηκε ο εξής τύπος για την συνάρτηση διακύμανσης:  
Η διακύμανση της χρονολογικής σειράς  $\mathbf{x}_i$ , με  $i=1,\dots,N$ , ορίζεται ως:

$$Var(\mathbf{x}) = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{x}_i - \mu)^2$$

όπου  $\mu$ , ο μέσος όρος της  $\mathbf{x}$ .

## 2.5 ΕΥΡΟΣ ΤΙΜΩΝ ΤΗΣ ΕΚΤΙΜΗΣΗ ΤΗΣ $H$ ΠΑΡΑΜΕΤΡΟΥ

Σε ότι αφορά τις τιμές που μπορεί να πάρει η παράμετρος  $H$  (βαθμός αυτό-ομοιότητας), κατά την εκτίμησή της από τις μεθόδους των Αθροιστικών Χρονολογικών Σειρών και της *R/S* Στατιστικής προκύπτουν τα εξής:

**Μέθοδος Αθροιστικών Χρονολογικών Σειρών:**

Όπως έχει ήδη αναφερθεί, ισχύουν οι παρακάτω σχέσεις για την εκτίμηση της παραμέτρου  $H$ :

$$\hat{H} = 1 - \frac{\hat{\beta}}{2} \quad (4.5)$$

$$\text{όπου } \hat{\beta} = \frac{\log Var(\mathbf{x}) - \log Var(\mathbf{x}^{(m)})}{\log m} \quad (4.6)$$

όπου  $\mathbf{x}$  η χρονολογική σειρά,  $m$ , το επίπεδο άθροισης, και  $\mathbf{x}^{(m)}$  η  $m$ -αθροιστική χρονολογική σειρά.

Για να ισχύει  $\hat{H} < 0$ , θα πρέπει όπως προκύπτει από την (4.5) να ισχύει  $\hat{\beta} > 2$ , δηλαδή βάσει της (4.6), να ισχύει:

$$\log Var(\mathbf{x}) - \log Var(\mathbf{x}^{(m)}) > 2 \log(m) \quad (4.7)$$

Είναι φανερό ότι εφ' όσον υπάρχουν χρονολογικές σειρές  $\mathbf{x}$  που να ικανοποιούν την (4.7), τότε υπάρχουν και περιπτώσεις όπου  $\hat{H} < 0$ .

Επίσης, ισχύει ότι  $\hat{H} > 1$  αν  $\hat{\beta} < 0$  (4.5), δηλαδή από (4.6) όταν

$$\log Var(x^{(m)}) > \log Var(x) \quad (4.8)$$

Εφ' όσον, όπως αποδεικνύεται παρακάτω, υπάρχουν χρονολογικές σειρές  $x$  που να ικανοποιούν την (4.8), τότε δικαιολογούνται τιμές της  $\hat{H}$  μεγαλύτερες της μονάδας.

### Μέθοδος R/S Στατιστικής:

Για την εκτίμηση της παραμέτρου  $H$  με την μέθοδο της R/S Στατιστικής ισχύει:

$$\hat{H} = \frac{\log R/S}{\log m - \log 2} \quad (4.9)$$

όπου  $R/S$  η r/s τιμή της  $m$ -αθροιστικής χρονολογικής σειράς όπως κατασκευάζεται από την μέθοδο, και  $m$  το επίτεδο άθροισης.

Έτσι, για να ισχύει  $\hat{H} > 1$ , θα πρέπει βάσει της (4.9) να ισχύει:

$$\log R/S > \log m - \log 2 \quad (4.10)$$

Όπως φαίνεται παρακάτω, από τις εκτιμήσεις  $\hat{H}$  πραγματικών δεδομένων κίνησης, υπάρχουν χρονολογικές σειρές που να ικανοποιούν την (4.10), και άρα να εμφανίζουν σύμφωνα με την μέθοδο της R/S Στατιστικής εκτιμήσεις του βαθμού αυτό-ομοιότητας μεγαλύτερες της μονάδας.

Αντίστοιχα, για να ισχύει  $\hat{H} < 0$ , θα πρέπει βάσει της (4.9) να ισχύει

$$\text{ή } \log R/S < 0$$

$$\text{ή } \log(m) - \log 2 < 0,$$

συνθήκες οι οποίες είναι αδύνατο να ικανοποιηθούν. Έτσι, συμπεραίνεται ότι οι εκτιμήσεις του βαθμού αυτό-ομοιότητας σύμφωνα με την R/S Στατιστική, δεν μπορούν να είναι ποτέ αρνητικές.

### 3. ΕΛΕΓΧΟΣ ΤΩΝ ΕΡΓΑΛΕΙΩΝ ΕΚΤΙΜΗΣΗΣ ΤΟΥ ΒΑΘΜΟΥ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ

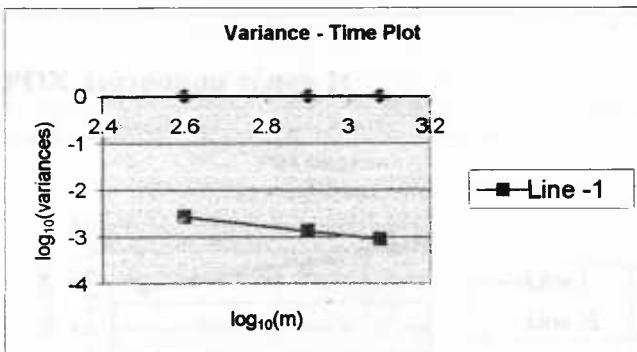
Προκειμένου να ελεγχθεί η ορθότητα των αποτελεσμάτων των υλοποιήσεων των δύο μεθόδων εκτίμησης του βαθμού αυτό-ομοιότητας, οι δύο μέθοδοι ελέγχθηκαν με χρονολογικές σειρές των οποίων η συμπεριφορά τους όσον αφορά την ύπαρξη αυτό-ομοιότητας είναι εκ φύσεως γνωστή. Παρακάτω παρατίθενται τα αποτελέσματα:

**Χρονολογική Σειρά :**  $x_t = 18,9 \{t = 2 \dots 2501\}$

**Μέθοδος Αθροιστικών Χ.Σ.:**  $\hat{H} = 1$

**Μέθοδος R/S Στατιστικής:**  $\hat{H} = 1$

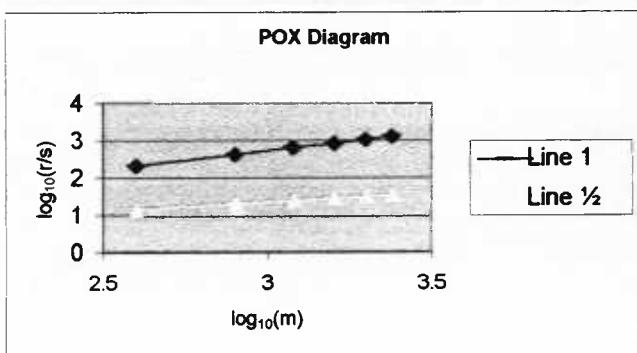
**Variance – Time Διάγραμμα:**



Στο διπλανό διάγραμμα οι διακυμάνσεις είναι μηδενικές και βρίσκονται πάνω από τη γραμμή με κλίση -1 (Line -1).

Ετσι προκύπτει ότι η χρονολογική σειρά είναι ακριβώς αυτό-όμοια (exactly self-similar) όπως αναμενόταν.

**POX Διάγραμμα τύπου 1:**



Στο διπλανό διάγραμμα οι τιμές της R/S στατιστικής συμπίπτουν με τη γραμμή με κλίση 1 (Line 1).

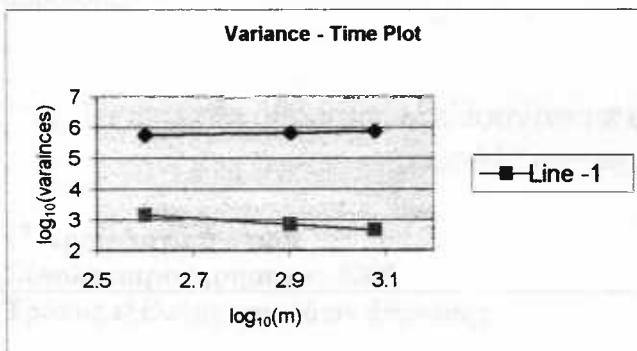
Ετσι προκύπτει ότι η Χ.Σ. είναι ακριβώς αυτό-όμοια (exactly self-similar) όπως αναμενόταν.

Χρονολογική Σειρά :  $x_t = t \{t = 2, 2.2501\}$

Μέθοδος Αθροιστικών Χ.Σ.:  $H = 1.0147$

Μέθοδος R/S Στατιστικής:  $H = 0.9772$

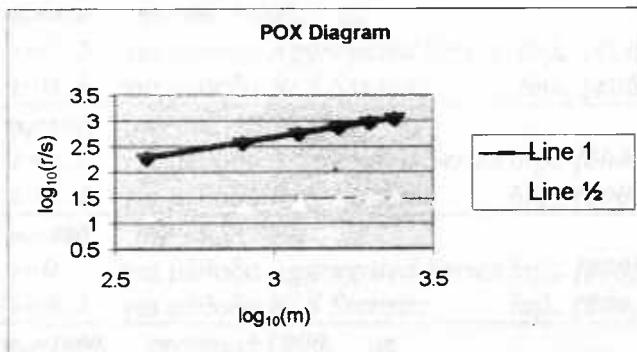
Variance – Time Διάγραμμα:



Στο διπλανό διάγραμμα οι διακυμάνσεις έχουν μεγάλη απόσταση από τη γραμμή με κλίση -1 (Line -1).

Ετσι προκύπτει ότι η χρονολογική σειρά χαρακτηρίζεται από μεγάλο βαθμό αυτό-ομοιότητας όπως αναμενόταν.

POX Διάγραμμα τύπου 1:



Στο διπλανό διάγραμμα οι τιμές της R/S στατιστικής βρίσκονται πολύ κοντά με τη γραμμή με κλίση 1 (Line 1).

Ετσι προκύπτει ότι η χρονολογική σειρά χαρακτηρίζεται από μεγάλο βαθμό αυτό-ομοιότητας, όπως αναμενόταν.

#### 4. Η ΕΠΙΔΡΑΣΗ ΤΟΥ ΕΠΙΠΕΔΟΥ ΑΘΡΟΙΣΗΣ ( $m$ ) ΣΤΙΣ ΕΚΤΙΜΗΣΕΙΣ ΤΟΥ ΒΑΘΜΟΥ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ

Στην παρούσα παράγραφο επιχειρείται να διαπιστωθεί ο βαθμός και το είδος της επίδρασης της επιλογής του επιπέδου άθροισης ( $m$ ), στις εκτιμήσεις της παραμέτρου  $H$  (βαθμός αυτό-ομοιότητας).

Προκειμένου να ανιχνευθεί η σχέση μεταξύ  $m$  και  $H$ , εκτιμήθηκαν οι βαθμοί αυτό-ομοιότητας τριών διαφορετικών χρονολογικών σειρών, για διαφορετικά επίπεδα άθροισης.

**A. Τα επίπεδα άθροισης εξελίσσονται με αθροιστικό τρόπο της μορφής**  
 $m_0=M, m_i=m_{i-1}+M$

1<sup>η</sup> Χρονολογική σειρά

Σύνολο παρατηρήσεων: 3080

Τρόπος εξέλιξης επιπέδων άθροισης	Μέθοδος Aggr.Ser.	Μέθοδος R/S Stat.
$m_0=200, m_i=m_{i-1}+200$ , με $i=0..6$ για μέθοδο Aggregated Series δηλ. $\{200, 400, \dots, 1400\}$ $i=0..14$ για μέθοδο R/S Statistic δηλ. $(200, 400, \dots, 3000)$	0.949	0.889
$m_0=400, m_i=m_{i-1}+400$ , με $i=0..2$ για μέθοδο Aggregated Series δηλ. $\{400, 800, 1200\}$ $i=0..6$ για μέθοδο R/S Statistic δηλ. $(400, 800, \dots, 2800)$	0.968	0.881
$m_0=600, m_i=m_{i-1}+600$ , με $i=0..1$ για μέθοδο Aggregated Series δηλ. $\{600, 1200\}$ $i=0..4$ για μέθοδο R/S Statistic δηλ. $(600, 1200, \dots, 3000)$	0.961	0.895
$m_0=800, m_i=m_{i-1}+800$ , με $i=0$ για μέθοδο Aggregated Series δηλ. $\{800\}$ $i=0..2$ για μέθοδο R/S Statistic δηλ. $(800, 1600, 2400)$	0.974	0.894
$m_0=1000, m_i=m_{i-1}+1000$ , με $i=0$ για μέθοδο Aggregated Series δηλ. $\{1000\}$ $i=0..2$ για μέθοδο R/S Statistic δηλ. $(1000, 2000, 3000)$	0.924	0.913
$m_0=1500, m_i=m_{i-1}+1500$ , με $i=0$ για μέθοδο Aggregated Series δηλ. $\{1500\}$ $i=0..1$ για μέθοδο R/S Statistic δηλ. $(1500, 3000)$	0.872	0.924

## 2<sup>η</sup> Χρονολογική Σειρά

Σύνολο παρατηρήσεων: 10876

Τρόπος εξέλιξης επιπέδων άθροισης	Μέθοδος Aggr.Ser.	Μέθοδος R/S Stat.
$m_0=200$ , $m_i=m_{i-1}+200$ , με $i=0..27$ για μέθοδο Aggregated Series δηλ. {200,400,...,10800}	0.761	-
$m_0=600$ , $m_i=m_{i-1}+600$ , με $i=0..8$ για μέθοδο Aggregated Series δηλ. {600,1200,...,5400} $i=0..17$ για μέθοδο R/S Statistic δηλ. (600,1200,...,10800)	0.779	0.671
$m_0=800$ , $m_i=m_{i-1}+800$ , με $i=0..6$ για μέθοδο Aggregated Series δηλ. {800,1600,...,4800} $i=1..13$ για μέθοδο R/S Statistic δηλ. (800,1600,10400)	0.75	0.66
$m_0=1000$ , $m_i=m_{i-1}+1000$ , με $i=0..4$ για μέθοδο Aggregated Series δηλ. {1000,2000,...,5000} $i=0..9$ για μέθοδο R/S Statistic δηλ. (1000,2000,...,10000)	0.746	0.659
$m_0=2000$ , $m_i=m_{i-1}+2000$ , με $i=0..1$ για μέθοδο Aggregated Series δηλ. {2000,4000} $i=0..4$ για μέθοδο R/S Statistic δηλ. (2000,4000,...,10000)	0.715	0.652
$m_0=4000$ , $m_i=m_{i-1}+4000$ , με $i=0$ για μέθοδο Aggregated Series δηλ. {4000} $i=0..1$ για μέθοδο R/S Statistic δηλ. (4000,8000)	0.611	0.593

## 3<sup>η</sup> Χρονολογική σειρά

Σύνολο παρατηρήσεων: 10876

Τρόπος εξέλιξης επιπέδων άθροισης	Μέθοδος Aggr.Ser.	Μέθοδος R/S Stat.
$m_0=200$ , $m_i=m_{i-1}+200$ , με $i=0..27$ για μέθοδο Aggregated Series δηλ. {200,400,...,10800}	0.449	-
$m_0=600$ , $m_i=m_{i-1}+600$ , με $i=0..8$ για μέθοδο Aggregated Series δηλ. {600,1200,...,5400} $i=0..17$ για μέθοδο R/S Statistic δηλ. (600,1200,...,10800)	0.452	0.496
$m_0=800$ , $m_i=m_{i-1}+800$ , με $i=0..6$ για μέθοδο Aggregated Series δηλ. {800,1600,...,4800} $i=1..13$ για μέθοδο R/S Statistic δηλ. (800,1600,10400)	0.479	0.548
$m_0=1000$ , $m_i=m_{i-1}+1000$ , με $i=0..4$ για μέθοδο Aggregated Series δηλ. {1000,2000,...,5000} $i=0..9$ για μέθοδο R/S Statistic δηλ. (1000,2000,...,10000)	0.451	0.493
$m_0=2000$ , $m_i=m_{i-1}+2000$ , με $i=0..1$ για μέθοδο Aggregated Series δηλ. {2000,4000} $i=0..4$ για μέθοδο R/S Statistic δηλ. (2000,4000,...,10000)	0.417	0.48
$m_0=4000$ , $m_i=m_{i-1}+4000$ , με $i=0$ για μέθοδο Aggregated Series δηλ. {4000} $i=0..1$ για μέθοδο R/S Statistic δηλ. (4000,8000)	0.362	0.458

Παρατηρώντας τους παραπάνω πίνακες φαίνεται ότι για μικρά επίπεδα άθροισης  $m$  οι εκτιμήσεις της παραμέτρου  $H$  που προκύπτουν και από τις δύο μεθόδους εκτιμήσεις δεν παρουσιάζουν σημαντικές διαφορές.

Συγκεκριμένα για την 1<sup>η</sup> χρονολογική σειρά και για επίπεδα άθροισης τα οποία εκκινούν και αυξάνονται κατά 200, 400, 600 και 800, οι εκτιμήσεις της  $H$  που προκύπτουν από την μέθοδο *Aggregated Series* έχουν σχεδόν την ίδια τιμή. Το ίδιο ισχύει και για την μέθοδο *R/S Statistic*. Όμως οι εκτιμήσεις της μεθόδου *Aggregated Series* για επίπεδα άθροισης  $m$  που εκκινούν και αυξάνονται από 1000 και 1500 προκύπτουν μειωμένες σε σχέση με αυτές των μικρότερων επιπέδων άθροισης. Επίσης, παρατηρείται ότι οι εκτιμήσεις της μεθόδου *R/S Statistic* για μεγάλα επίπεδα άθροισης (1000 και 1500), εμφανίζονται αυξημένες σε σχέση με τα μικρότερα επίπεδα άθροισης.

Παρόμοιες παρατηρήσεις ισχύουν και για τη 2<sup>η</sup> χρονολογική σειρά η οποία περιλαμβάνει 10876 παρατηρήσεις. Και οι δύο μέθοδοι εκτίμησης για μικρά επίπεδα αθροίσεων δίνουν εκτιμήσεις οι οποίες κυμαίνονται στα ίδια επίπεδα τιμών. Αντίθετα παρατηρείται ότι για επίπεδα άθροισης τα οποία εκκινούν και αυξάνονται κατά 4000 οι εκτιμήσεις και των δύο μεθόδων εμφανίζονται μειωμένες σε σχέση με τις εκτιμήσεις των μικρότερων επιπέδων.

Ίδιες παρατηρήσεις ισχύουν και για την 3<sup>η</sup> χρονολογική σειρά. Οι εκτιμήσεις και των δύο μεθόδων για μεγάλα επίπεδα άθροισης (2000 και 4000) εμφανίζονται μειωμένες σε σχέση με τις εκτιμήσεις των μικρότερων επιπέδων άθροισης οι οποίες δεν παρουσιάζουν σημαντικές διαφορές.

## B. Τα επίπεδα άθροισης εξελίσσονται με πολλαπλό τρόπο της μορφής $m_0=M, m_i=2*m_{i-1}$

### 1<sup>η</sup> Χρονολογική σειρά

**Σύνολο παρατηρήσεων: 3080**

Τρόπος εξέλιξης επιπέδων άθροισης	Μέθοδος Aggr.Ser.	Μέθοδος R/S Stat.
$m_0=200, m_i=2*m_{i-1}, \text{ με}$ $i=0..2$ για μέθοδο <i>Aggregated Series</i> δηλ. $\{200, 400, 800\}$ $i=0..3$ για μέθοδο <i>R/S Statistic</i> δηλ. $\{200, 400, \dots, 1600\}$	0.967	0.833
$m_0=400, m_i=2*m_{i-1}, \text{ με}$ $i=0..1$ για μέθοδο <i>Aggregated Series</i> δηλ. $\{400, 800\}$ $i=0..2$ για μέθοδο <i>R/S Statistic</i> δηλ. $\{400, 800, 1600\}$	0.968	0.834
$m_0=600, m_i=2*m_{i-1}, \text{ με}$ $i=0..1$ για μέθοδο <i>Aggregated Series</i> δηλ. $\{600, 1200\}$ $i=0..2$ για μέθοδο <i>R/S Statistic</i> δηλ. $\{600, 1200, 2400\}$	0.961	0.867
$m_0=800, m_i=2*m_{i-1}, \text{ με}$ $i=0$ για μέθοδο <i>Aggregated Series</i> δηλ. $\{800\}$ $i=0..2$ για μέθοδο <i>R/S Statistic</i> δηλ. $\{800, 1600, 2400\}$	0.974	0.873
$m_0=1000, m_i=2*m_{i-1}, \text{ με}$ $i=0$ για μέθοδο <i>Aggregated Series</i> δηλ. $\{1000\}$ $i=0..2$ για μέθοδο <i>R/S Statistic</i> δηλ. $\{1000, 2000, 3000\}$	0.924	0.902

$m_0=1500$ , $m_i=2*m_{i-1}$ , με $i=0$ για μέθοδο Aggregated Series δηλ. {1500}; $i=0..1$ για μέθοδο R/S Statistic δηλ. (1500,3000);	0.872	0.924
---	-------	-------

## 2<sup>η</sup> Χρονολογική σειρά

Σύνολο παρατηρήσεων: 10876

Τρόπος εξέλιξης επιπέδων άθροισης	Μέθοδος Aggr.Ser.	Μέθοδος R/S Stat.
$m_0=200$ , $m_i=2*m_{i-1}$ , με $i=0..4$ για μέθοδο Aggregated Series δηλ. {200,400, ...,3200}; $i=0..5$ για μέθοδο R/S Statistic δηλ. (200,400,...,6400);	0.791	0.594
$m_0=600$ , $m_i=2*m_{i-1}$ , με $i=0..3$ για μέθοδο Aggregated Series δηλ. {600,1200,...,4800}; $i=0..4$ για μέθοδο R/S Statistic δηλ. (600,1200,...,9600);	0.799	0.644
$m_0=800$ , $m_i=2*m_{i-1}$ , με $i=0..2$ για μέθοδο Aggregated Series δηλ. {800,1600,...,3200}; $i=0..3$ για μέθοδο R/S Statistic δηλ. (800,1600,...,6400);	0.792	0.615
$m_0=1000$ , $m_i=2*m_{i-1}$ , με $i=0..2$ για μέθοδο Aggregated Series δηλ. {1000,2000,4000}; $i=0..3$ για μέθοδο R/S Statistic δηλ. (1000,2000,...,8000);	0.725	0.582
$m_0=2000$ , $m_i=2*m_{i-1}$ , με $i=0..1$ για μέθοδο Aggregated Series δηλ. {2000,4000}; $i=0..2$ για μέθοδο R/S Statistic δηλ. (2000,4000,8000);	0.685	0.593
$m_0=4000$ , $m_i=2*m_{i-1}$ , με $i=0$ για μέθοδο Aggregated Series δηλ. {4000}; $i=0..1$ για μέθοδο R/S Statistic δηλ. (4000,8000);	0.549	0.568

## 3<sup>η</sup> Χρονολογική σειρά

Σύνολο παρατηρήσεων: 10876

Τρόπος εξέλιξης επιπέδων άθροισης	Μέθοδος Aggr.Ser.	Μέθοδος R/S Stat.
$m_0=200$ , $m_i=2*m_{i-1}$ , με $i=0..4$ για μέθοδο Aggregated Series δηλ. {200,400, ...,3200}; $i=0..5$ για μέθοδο R/S Statistic δηλ. (200,400,...,6400);	0.451	0.506
$m_0=600$ , $m_i=2*m_{i-1}$ , με $i=0..3$ για μέθοδο Aggregated Series δηλ. {600,1200,...,4800}; $i=0..4$ για μέθοδο R/S Statistic δηλ. (600,1200,...,9600);	0.471	0.512
$m_0=800$ , $m_i=2*m_{i-1}$ , με $i=0..2$ για μέθοδο Aggregated Series δηλ. {800,1600,...,3200}; $i=0..3$ για μέθοδο R/S Statistic δηλ. (800,1600,...,6400);	0.466	0.503
$m_0=1000$ , $m_i=2*m_{i-1}$ , με $i=0..2$ για μέθοδο Aggregated Series δηλ. {1000,2000,4000}; $i=0..3$ για μέθοδο R/S Statistic δηλ. (1000,2000,...,8000);	0.427	0.481
$m_0=2000$ , $m_i=2*m_{i-1}$ , με $i=0..1$ για μέθοδο Aggregated Series δηλ. {2000,4000}; $i=0..2$ για μέθοδο R/S Statistic δηλ. (2000,4000,8000);	0.417	0.473
$m_0=4000$ , $m_i=2*m_{i-1}$ , με $i=0$ για μέθοδο Aggregated Series δηλ. {4000}; $i=0..1$ για μέθοδο R/S Statistic δηλ. (4000,8000);	0.362	0.458

Στην περίπτωση που τα επίπεδα άθροιστης εκκινούν από μια συγκεκριμένη τιμή  $m=200,600,800,1000$ , κλπ και κάθε επόμενη τιμή προκύπτει από την προηγούμενη πολλαπλασιαζόμενη επί 2, ισχύουν τα ίδια συμπεράσματα της περίπτωσης της αθροιστικής εξέλιξης των επιπέδων άθροιστης.

Το συμπέρασμα που πιθανώς προκύπτει είναι ότι οι εκτιμήσεις των μεγάλων επιπέδων άθροιστης διαφοροποιούνται σε σχέση με αυτές των μικρών επιπέδων, και για τις δύο μεθόδους εκτίμησης του βαθμού αυτό-ομοιότητας.

#### 4.1 ΣΥΓΚΡΙΣΗ ΑΘΡΟΙΣΤΙΚΟΥ & ΠΟΛΛΑΠΛΑΣΙΑΣΤΙΚΟΥ ΤΡΟΠΟΥ ΕΞΕΛΙΞΗΣ ΤΩΝ ΕΠΙΠΕΔΩΝ ΑΘΡΟΙΣΗΣ ( $m$ )

Συγκρίνοντας από τους παραπάνω πίνακες τον πολλαπλασιαστικό και τον αθροιστικό τρόπο εξέλιξης των επιπέδων άθροιστης, προκύπτει ότι: Τα αποτελέσματα των εκτιμήσεων της  $H$  είναι σχεδόν ίδια και για τους δύο τρόπους δημιουργίας των επιπέδων άθροιστης. Πρέπει ίσως να σημειωθεί ότι ο πολλαπλασιαστικός τρόπος εξέλιξης των  $m$  δίνει γρηγορότερα αποτελέσματα από τον αθροιστικό ο οποίος απαιτεί εν γένει μεγαλύτερο υπολογιστικό χρόνο προκειμένου να δώσει αποτελέσματα, λόγω λιγότερων βημάτων αθροίσεων.

Ένα άλλο σημείο που πρέπει να αναφερθεί είναι το εξής: Η μέθοδος εκτίμησης του βαθμού αυτό-ομοιότητας *R/S Statistic*, δίνει εκτιμήσεις οι οποίες προκύπτουν γενικά από περισσότερα επίπεδα άθροιστης σε σχέση με την μέθοδο εκτίμησης *Aggregated Series*. Αυτό οφείλεται στο γεγονός ότι δεν ορίζεται η διακύμανση μιας σταθεράς. Ο πολλαπλασιαστικός τρόπος εξέλιξης των επιπέδων άθροιστης μειώνει τον αριθμό των επιπέδων άθροιστης οι οποίες λαμβάνονται υπ' όψη στην μέθοδο *R/S Statistic* και δεν λαμβάνονται υπ' όψη από την μέθοδο *Aggregated Series*.

## ΚΕΦΑΛΑΙΟ 5

### ΣΥΛΛΟΓΗ ETHERNET ΚΙΝΗΣΗΣ ΑΠΟ ΤΟ ΔΙΚΤΥΟ ΔΕΔΟΜΕΝΩΝ ΤΟΥ Ο.Π.Α.

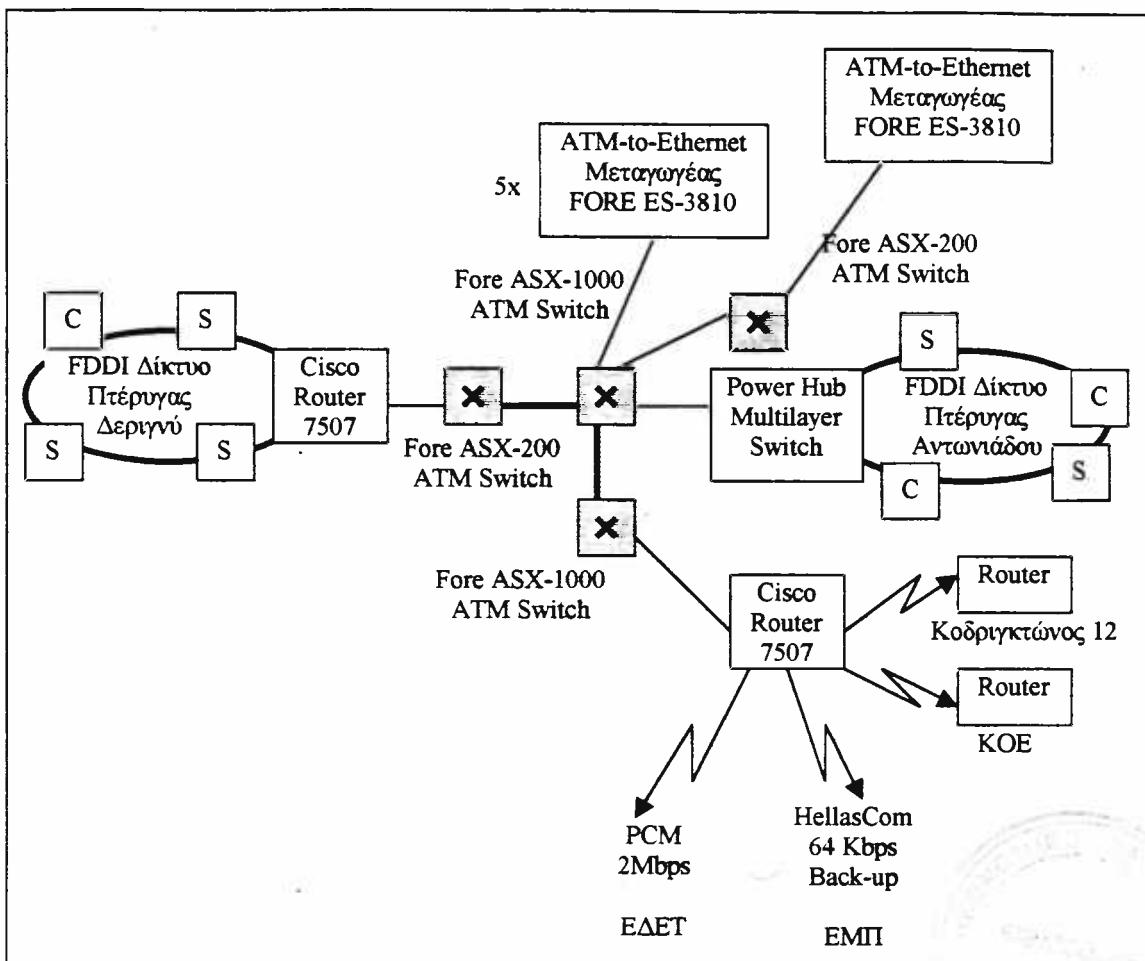
#### 1. ΔΙΚΤΥΟ ΔΕΔΟΜΕΝΩΝ ΤΟΥ Ο.Π.Α.

Τα βασικά χαρακτηριστικά στοιχεία του δικτύου δεδομένων του Πανεπιστημίου, έχουν ως εξής:

- **Δίκτυο κορμού τεχνολογίας ATM**, το οποίο επιτυγχάνει ταχύτητες 155 Mbps (δυνατότητα αναβάθμισης στα 622 Mbps). Το δίκτυο αποτελείται από τρεις μεταγωγείς ATM:
  1. Ένα μεταγωγέα ATM, ο οποίος χρησιμοποιείται και για τη σύνδεση με το Διαπανεπιστημιακό δίκτυο ATM στα 34 Mbps, στο ισόγειο της Πτέρυγας Αντωνιάδου.
  2. Ένα μεταγωγέα ATM, στην Πτέρυγα Αντωνιάδου.
  3. Ένα μεταγωγέα ATM, στην Πτέρυγα Δεριγνύ.
- Πέντε (5) μεταγωγείς ATM-to-Ethernet για την πρόσβαση των Εργαστηρίων του 3<sup>ου</sup> και 4<sup>ου</sup> ορόφου της Πτέρυγας Αντωνιάδου απευθείας στο δίκτυο κορμού ATM με ταχύτητες 10 και 100 Mbps.
- Τρία (3) δίκτυα FDDI, ένα (1) στην Πτέρυγα Δεριγνύ, ένα (1) στην Πτέρυγα Αντωνιάδου και ένα (1) στο κτίριο Δεριγνύ, τα οποία συνδέουν το υπάρχον δίκτυο διανομής με το δίκτυο κορμού ATM.
- Ένας μεταγωγέας πολλαπλών επιπέδων, ο οποίος έχει δυνατότητες γεφύρωσης, δρομολόγησης πολλαπλών πρωτοκόλλων και εικονικών δικτύων (VLANs). Ο μεταγωγέας αυτός παρέχει τη δυνατότητα σύνδεσης του δικτύου κορμού ATM και του δικτύου FDDI της Πτέρυγας Αντωνιάδου. Επίσης παρέχει την πρόσβαση των Εργαστηρίων της Πτέρυγας Αντωνιάδου στο δίκτυο κορμού ATM μέσω της δρομολόγησης των εικονικών δικτύων τους (VLANs & ELANs).
- Ένας δρομολογητής Cisco 7507, στην Πτέρυγα Αντωνιάδου που υποστηρίζει τις συνδέσεις WAN του Πανεπιστημίου με το κτίριο της Κοδριγκτώνος 12, το Κέντρο Ερευνας και το ΕΔΕΤ. Ο δρομολογητής συνδέεται με το δίκτυο κορμού ATM σε ένα μεταγωγέα ATM μέσω κατάλληλης κάρτας.



- Ένας δρομολογητής Cisco 7507, στην Πτέρυγα Δεριγνύ ο οποίος θα συνδέει το ATM δίκτυο κορμού με το δίκτυο FDDI της Πτέρυγας Δεριγνύ.
- Μεταγωγείς FDDI-to-Ethernet, για την πρόσβαση των τοπικών δικτύων στα δίκτυα FDDI.



Σχήμα 5.1 Η τοπολογία του δικτύου δεδομένων του Οικονομικού Πανεπιστημίου Αθηνών.

## 2. ΤΡΟΠΟΙ ΣΥΛΛΟΓΗΣ ΣΤΟΙΧΕΙΩΝ ΑΠΟ ΤΗΝ RMON MIB v1 (RFC1757)

Για τη συλλογή στοιχείων κίνησης από τα Ethernet (υπο-)δίκτυα του δικτύου δεδομένων του ΟΠΑ, υπάρχουν δύο δυνατότητες:

Η πρώτη αφορά την συσκευή LAN probe J3458A της HP (για περιγραφή τρόπου λειτουργίας βλ. Παράρτημα 1). Αυτή περιλαμβάνει υλοποιήσεις αντιπροσώπων (agents) για την RMON MIB v1 (RFC1757) & v2 (RFC2021).

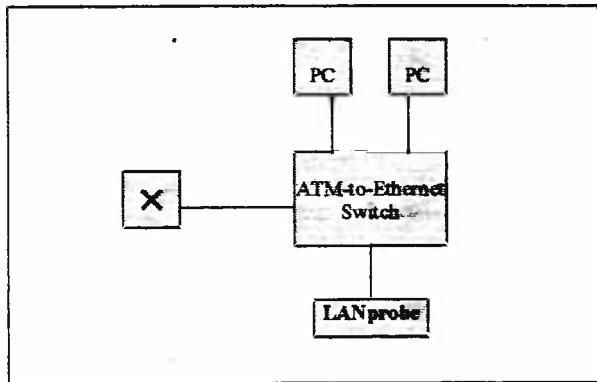
Η δεύτερη αφορά την εγκατάσταση των RMON αντιπροσώπων v1 (RFC1757) σε FDDI-to-Ethernet DEC μεταγωγείς (DEC switches).

Πρέπει να σημειωθεί ότι η χρησιμοποιούμενη τεχνολογία του δικτύου δεδομένων του Πανεπιστημίου (ATM backbone, FDDI δίκτυα), θέτει περιορισμούς στη χρήση του Ethernet LAN probe και στα πιθανά σενάρια διασύνδεσης της συσκευής στα επιμέρους τμήματα του δικτύου. Ιδιοί περιορισμοί τίθενται και για την εγκατάσταση των RMON αντιπροσώπων v1 (RFC 1757) σε FDDI-to-Ethernet DEC μεταγωγείς.

Παρακάτω παρουσιάζονται όλοι οι πιθανοί τρόποι συλλογής στοιχείων μετρήσεων κάνοντας χρήση του Ethernet LAN probe, αλλά και των αντιπροσώπων για την RMON MIB v1 (RFC1757).

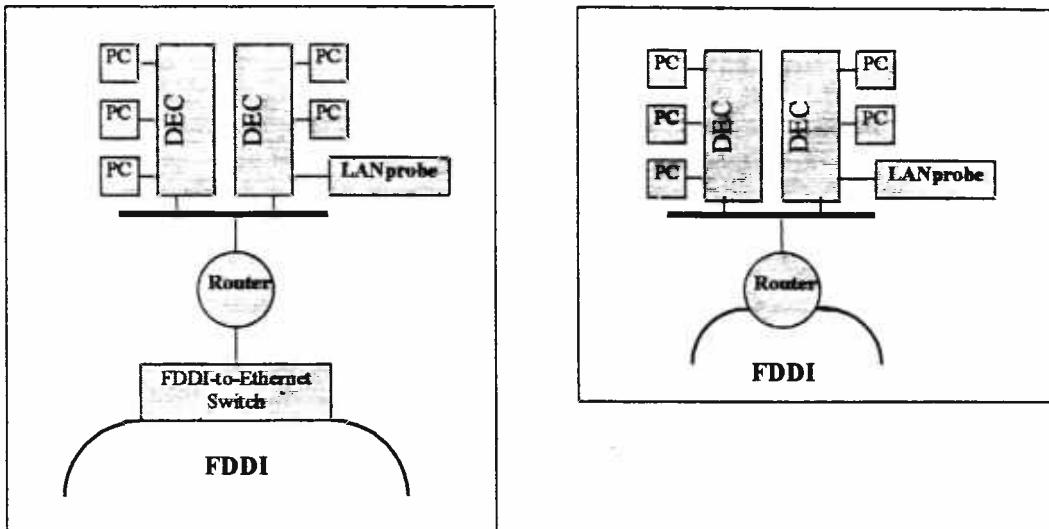
## 2.1 ΣΥΛΛΟΓΗ ΣΤΟΙΧΕΙΩΝ ΜΕ ΧΡΗΣΗ ΤΟΥ ETHERNET LAN PROBE

1. **Σύνδεση του LAN probe σε ATM-to-Ethernet μεταγωγέα.** Σε αυτή την περίπτωση εντοπίζεται η κίνηση που αφορά την επικοινωνία μεταξύ των υπολογιστικών συσκευών του συγκεκριμένου εικονικού δικτύου (VLAN), καθώς και η κίνηση από την επικοινωνία κάθε μιας από αυτές με το ATM link.



Σχήμα 5.2 Σύνδεση του LAN probe σε ATM-to-Ethernet μεταγωγέα.

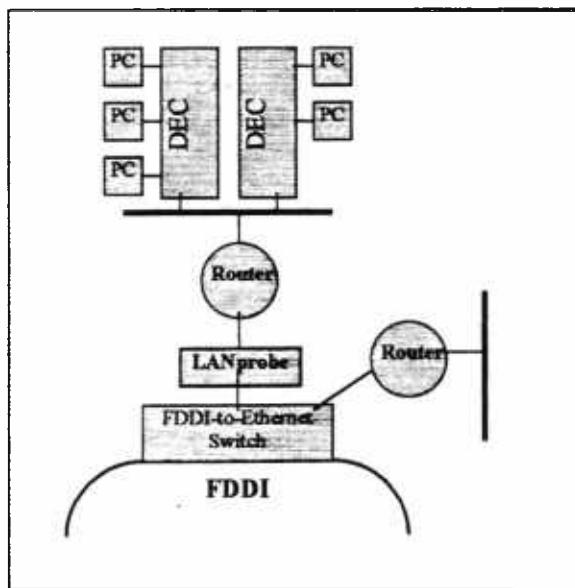
2. Σύνδεση του LAN probe σε Ethernet Hub Εργαστηρίου, του οποίου ο εξορισμού δρομολογητής (default router) συνδέεται με FDDI-to-Ethernet μεταγωγέα ή με FDDI δίκτυο. Σε αυτή την περίπτωση εντοπίζεται η κίνηση που αφορά την επικοινωνία μεταξύ των υπολογιστικών συσκευών του συγκεκριμένου Εργαστηρίου, καθώς και η κίνηση από την επικοινωνία κάθε μιας από αυτές με τον εξορισμού δρομολογητή.



Σχήμα 5.3 (α) Σύνδεση του LAN probe σε Ethernet Hub Εργαστηρίου, ο εξορισμού δρομολογητής του οποίου είναι συνδεδεμένος σε FDDI-to-Ethernet μεταγωγέα.

Σχήμα 5.3 (β) Σύνδεση του LAN probe σε Ethernet Hub Εργαστηρίου, ο εξορισμού δρομολογητής του οποίου είναι συνδεδεμένος σε FDDI δίκτυο.

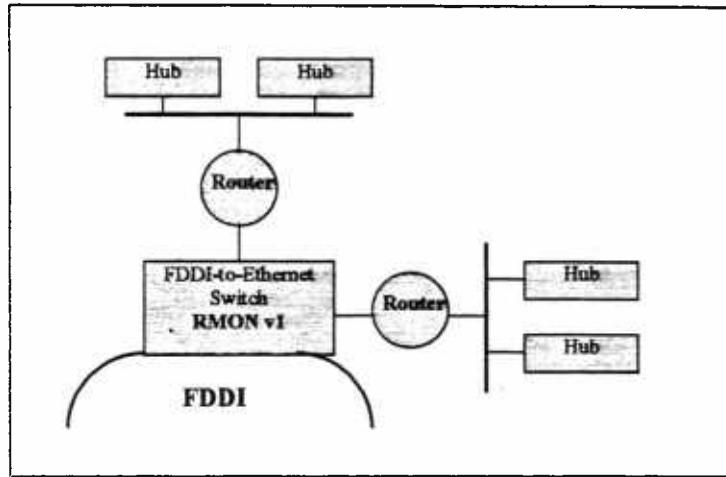
3. Σύνδεση του LAN probe σε δρομολογητή και σε FDDI-to-Ethernet μεταγωγέα ο οποίος συνδέεται με FDDI δίκτυο. Σε αυτή την περίπτωση εντοπίζεται η κίνηση που αφορά την επικοινωνία μεταξύ του δρομολογητή και του μεταγωγέα.



Σχήμα 5.4 Σύνδεση του LAN probe σε FDDI-to-Ethernet μεταγωγέα ο οποίος συνδέεται με FDDI δίκτυο και δρομολογητή ή δρομολογητές.

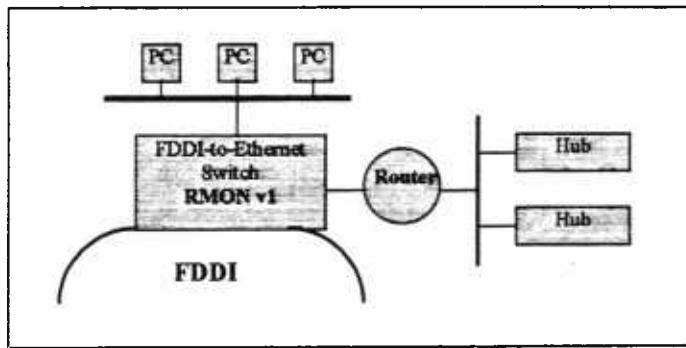
## 2.2 ΣΥΛΛΟΓΗ ΣΤΟΙΧΕΙΩΝ ΜΕΣΩ ΕΓΚΑΤΑΣΤΑΣΗΣ ΑΝΤΙΠΡΟΣΩΠΩΝ ΤΗΣ RMON MIB V1 (RFC1757)

1. *Εγκατάσταση αντιπροσώπου της RMON MIB v1 (RFC1757), σε FDDI-to-Ethernet μεταγωγέα στον οποίο συνδέονται ένας ή περισσότεροι δρομολογητές.* Σε αυτή την περίπτωση εντοπίζεται η κίνηση που αφορά την επικοινωνία των δρομολογητών με το FDDI δίκτυο, αλλά και των δρομολογητών μεταξύ τους.



Σχήμα 5.5 Εγκατάσταση αντιπροσώπου της RMON MIB v1 (RFC1757) σε FDDI-to-Ethernet μεταγωγέα στον οποίο συνδέονται ένας ή περισσότεροι δρομολογητές.

2. *Εγκατάσταση αντιπροσώπου της RMON MIB v1 (RFC1757), σε FDDI-to-Ethernet μεταγωγέα στον οποίο συνδέονται ένας ή περισσότεροι δρομολογητές, αλλά και Ethernet δίκτυο.* Σε αυτή την περίπτωση εντοπίζεται η κίνηση που αφορά την επικοινωνία των δρομολογητών και του Ethernet δικτύου με το FDDI δίκτυο, αλλά και η κίνηση που αφορά την επικοινωνία μεταξύ των υπολογιστικών συσκευών του Ethernet δικτύου και των δρομολογητών.



Σχήμα 5.5 Εγκατάσταση αντιπροσώπου της RMON MIB v1 (RFC1757) σε FDDI-to-Ethernet μεταγωγέα στον οποίο συνδέονται ένας ή περισσότεροι δρομολογητές, αλλά και Ethernet δίκτυο.

### 3. ΠΕΡΙΓΡΑΦΗ ΧΡΗΣΙΜΟΠΟΙΟΥΜΕΝΩΝ RMON MIB v1 (RFC1757) ΑΝΤΙΚΕΙΜΕΝΩΝ

Τα αντικείμενα της RMON MIB v1 για τα οποία συλλέχθηκαν στοιχεία στις μετρήσεις της κίνησης των δικτύων οι οποίες πραγματοποιήθηκαν είναι τα εξής:

#### **RMON MIB v1 (RFC 1757) ΠΕΡΙΓΡΑΦΗ ΑΝΤΙΚΕΙΜΕΝΟ**

MatrixSDOctets	Ο αριθμός των ψηφιοσυλλαβών (bytes) όλων των πακέτων που μεταφέρονται από την διεύθυνση πηγής στην διεύθυνση προοριζμού.
HostInOctets	Ο αριθμός των ψηφιοσυλλαβών (bytes) όλων των πακέτων που μεταφέρονται στην διεύθυνση του host, συμπεριλαμβανομένων και των ψηφιοσυλλαβών των λανθασμένων πακέτων.
HostOutOctets	Ο αριθμός των ψηφιοσυλλαβών (bytes) όλων των πακέτων που μεταφέρονται από την διεύθυνση του host προς όλους τους προορισμούς, εκτός των ψηφιοσυλλαβών των λανθασμένων πακέτων.
HostInPkts	Ο αριθμός των μη λανθασμένων πακέτων που μεταφέρονται στην διεύθυνση του host.
HostOutPkts	Ο αριθμός των πακέτων (συμπεριλαμβανομένων και των λανθασμένων) που μεταφέρονται από την διεύθυνση του host προς όλους τους προορισμούς.
EtherStatsOctets	Ο αριθμός των ψηφιοσυλλαβών (bytes) όλων των πακέτων (συμπεριλαμβανομένων και των λανθασμένων) που εμφανίστηκαν στο δίκτυο.
EtherStatsPkts	Ο αριθμός των πακέτων (συμπεριλαμβανομένων των λανθασμένων, των broadcast και multicast πακέτων) που εμφανίστηκαν στο δίκτυο.

#### 4. ΔΙΑΔΙΚΑΣΙΑ ΣΥΛΛΟΓΗΣ ΣΤΟΙΧΕΙΩΝ ΑΠΟ ΤΗΝ RMON MIB v1

Η διαδικασία συλλογής στοιχείων για τα αντικείμενα της RMON MIB v1 που περιγράφονται παραπάνω, μέσω της χρήσης του Ethernet Lan probe, περιλαμβάνει τα εξής βήματα:

1. Η Ethernet Lan Probe συσκευή ρυθμίζεται (configured) ώστε να της εκχωρηθούν: μια IP διεύθυνση, η μάσκα, καθώς και ο default gateway του δικτύου στο οποίο πρόκειται να συνδεθεί.
2. Η συσκευή συνδέεται στο δίκτυο.
3. Μέσω του σταθμού διαχείρισης δικτύων, ορίζονται τα αντικείμενα της RMON MIB v1, για τα οποία επιθυμείται η συλλογή στοιχείων. Σε αυτό το στάδιο ορίζεται και μια σειρά άλλων παραμέτρων όπως ο ρυθμός συλλογής στοιχείων (polling period).

Σημειώνεται ότι στην περίπτωση που δεν χρησιμοποιείται η Ethernet Lan probe συσκευή για την συλλογή των στοιχείων, αλλά αυτή πραγματοποιείται μέσω της εγκατάστασης RMON MIB v1 αντιπροσώπων (agents), τα βήματα 1 και 2 της παραπάνω διαδικασίας αντικαθίστανται με την εγκατάσταση των αντιπροσώπων στις κατάλληλες συσκευές.

Από τη στιγμή που εκκινείται μέχρι τη στιγμή που τερματίζεται η διαδικασία συλλογής στοιχείων, συλλέγονται και (επιλεκτικά) καταχωρούνται στοιχεία μετρήσεων σε κατάλληλα αρχεία συστήματος. Είσι για παράδειγμα στην περίπτωση που έχει οριστεί η συλλογή στοιχείων για το αντικείμενο etherStatsOctets κάθε 5 δευτερόλεπτα, τότε στο αρχείο συστήματος με το όνομα etherStatsOctets δημιουργείται μία εγγραφή κάθε 5 δευτερόλεπτα η οποία (εκτός των άλλων) περιλαμβάνει τον αριθμό των bytes που εμφανίστηκαν στο δίκτυο στο διάστημα των πέντε δευτερολέπτων. Στην περίπτωση που συλλέγονται στοιχεία για το αντικείμενο hostInPkts κάθε 10 δευτερόλεπτα, τότε δημιουργούνται τόσα αρχεία συστήματος όσα και οι hosts, με ονόματα της μορφής hostOutPkts.MAC-διεύθυνση-host. Σε κάθε ένα από αυτά τα αρχεία που αντιστοιχεί σε διαφορετικό host δημιουργείται κάθε 10 δευτερόλεπτα μία εγγραφή η οποία βασικά περιλαμβάνει τον αριθμό των πακέτων που στάλθηκαν από τον host προς όλους τους προορισμούς στο διάστημα των 10 δευτερολέπτων.

Τα αρχεία των στοιχείων μέτρησης που δημιουργούνται από την παραπάνω διαδικασία βρίσκονται σε δυαδική μορφή και απαιτείται η εκτέλεση ειδικής διεργασίας για την μετατροπή τους σε αναγνώσιμη μορφή.

## ΚΕΦΑΛΑΙΟ 6

# ΕΚΤΙΜΗΣΗ ΒΑΘΜΟΥ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ ETHERNET ΚΙΝΗΣΗΣ ΤΟΥ ΔΙΚΤΥΟΥ ΔΕΔΟΜΕΝΩΝ ΤΟΥ Ο.Π.Α.

### 1. ΕΙΣΑΓΩΓΗ

Τα Τοπικά Δίκτυα Υπολογιστών (*Local Area Networks -LANs*) έκαναν την εμφάνισή τους στα μέσα της δεκαετίας του 70. Η χρήση τους αφορούσε την διασύνδεση συσκευών επεξεργασίας δεδομένων (σταθμούς εργασίας, προσωπικούς υπολογιστές, εκτυπωτές, τερματικά, εξυπηρετητές αρχείων κλπ) σε περιβάλλον γραφείου ή τμημάτων πανεπιστημίων. Τα τοπικά δίκτυα τύπου Ethernet, αν και αποτελούν μια αρκετά παλιά τεχνολογία παραμένουν ακόμα και σήμερα αρκετά δημοφιλή. Μεταξύ των πλεονεκτημάτων τους είναι η ευκολία συντήρησης, διαχείρισης καθώς και διάρθρωσης του δικτύου (οι σταθμοί μπορούν να αποσυνδέονται ή να μετακινούνται από ένα σημείο και να επανασυνδέονται σε άλλο, χωρίς να διαταραχτεί η απρόσκοπη λειτουργία του υπόλοιπου δικτύου). Βάσει των σημερινών αναγκών, δύο από τα σημαντικότερα μειονεκτήματα της Ethernet τεχνολογίας είναι οι σχετικά χαμηλές ταχύτητες των 10 Mbit/s καθώς και η περιορισμένη φυσική εμβέλεια.

Η εμφάνιση σταθμών εργασίας υψηλής απόδοσης, οι υπερ-υπολογιστές καθώς και οι παράλληλες μηχανές έδωσαν την ώθηση για την εμφάνιση τοπικών δικτύων υπολογιστών υψηλότερης ταχύτητας (τα επονομαζόμενα *gigabit LANs*), με τουλάχιστον 100 φορές μεγαλύτερο επικοινωνιακό εύρος από τα παραδοσιακά Ethernet δίκτυα. Από την άλλη μεριά η επιτυχία και ο μεγάλος αριθμός των συμβατικών τοπικών δικτύων προκάλεσε την εξάπλωση των Μητροπολιτικών Δικτύων (*Metropolitan Area Networks - MANs*). Αυτή η τεχνολογία των δικτύων επέτρεψε την διασύνδεση τοπικών δικτύων (ακόμα και διαφορετικής τεχνολογίας) σε μια γεωγραφική περιοχή περίπου των 100 χλιομέτρων.

Η σύγχρονη τάση στις τηλεπικοινωνίες χαρακτηρίζεται από την μετακίνηση από τα δίκτυα ειδικού σκοπού (π.χ. διαφορετικά δίκτυα για τη μεταφορά φωνής, δεδομένων και κινούμενης εικόνας), προς τα δίκτυα γενικού σκοπού τα οποία είναι ανεξάρτητα υπηρεσιών και ονομάζονται *Eυρυζωνικά Ψηφιακά Δίκτυα Ολοκληρωμένων Υπηρεσιών (Broadband Integrated Services Digital Network -BISDN)*. Στα



τηλεφωνικά δίκτυα η μετάδοση σε όλη τη διάρκεια της σύνδεσης βασίζεται σε ένα (αφιερωμένο) κύκλωμα. Σε αντίθεση τα BISDN δίκτυα βασίζονται σε μια τεχνολογία μεταγωγής πακέτων η οποία επιτρέπει την μεταγωγή των δεδομένων κάθε είδους υπηρεσίας μέσω μιας κοινής ψηφιακής μορφής. Πρέπει να σημειωθεί ότι η επικοινωνία που στηρίζεται στην μετάδοση πακέτων είναι περισσότερο πολύπλοκη από αυτήν που στηρίζεται στα αφιερωμένα κυκλώματα.

Είναι φανερό ότι η κατανόηση των χαρακτηριστικών των BISDN δικτύων προϋποθέτει την μελέτη και τον χαρακτηρισμό της κίνησης των τοπικών δικτύων, όπως η Ethernet κίνηση. Με αυτό τον τρόπο είναι δυνατό να παραχθεί πολύτιμη γνώση η οποία θα βοηθήσει την λήψη αποφάσεων σχετικά με τη καταλληλότητα των μοντέλων κίνησης και θα συμβάλλει στην προσέγγιση των ζητημάτων της οικονομικότερης σχεδίασης και ελέγχου των μελλοντικών δικτύων.

## 2. ΓΕΝΙΚΗ ΣΥΝΟΨΗ ΠΕΙΡΑΜΑΤΩΝ

Προκειμένου να εκτιμηθεί ο βαθμός αυτό-ομοιότητας της κίνησης των τοπικών δικτύων του Πανεπιστημίου, πραγματοποιήθηκαν συνολικά 6 μετρήσεις σε 4 τοπικά δίκτυα του δικτύου δεδομένων. Τα αντικείμενα της RMON MIB v1 (RFC1757) για τα οποία συλλέχθηκαν στοιχεία αφορούν τον αριθμό των πακέτων και ψηφιοσυλλαβών που παρέλαβε και απέστειλε κάθε υπολογιστική συσκευή του δικτύου καθώς και τον αριθμό των πακέτων και ψηφιοσυλλαβών που εντοπίστηκαν στο δίκτυο κατά την διάρκεια των μετρήσεων. Έτσι, για κάθε μέτρηση που πραγματοποιήθηκε προέκυψαν τέσσερις χρονολογικές σειρές για κάθε μια υπολογιστική συσκευή του δικτύου (host). Κάθε μια από αυτές αφορά την κίνηση εισόδου ή εξόδου κάθε υπολογιστικής συσκευής μετρημένη σε πακέτα και ψηφιοσυλλαβές. Επίσης, για κάθε μέτρηση δημιουργήθηκαν δύο επιπλέον χρονολογικές σειρές οι οποίες αφορούν την συνολική κίνηση που εντοπίστηκε στο δίκτυο κατά την διάρκεια των μετρήσεων μετρημένη σε πακέτα και ψηφιοσυλλαβές. Ο ρυθμός συλλογής στοιχείων (polling period) για τον σχηματισμό των παραπάνω χρονολογικών σειρών ορίστηκε για τις περισσότερες μετρήσεις ανά 5 δευτερόλεπτα, ενώ για τις υπόλοιπες ανά 10 δευτερόλεπτα. Πρέπει να σημειωθεί ότι το πλήθος των παρατηρήσεων των χρονολογικών σειρών που δημιουργήθηκαν και από τις έξι μετρήσεις κυμαίνεται μεταξύ 2000 και 10000. Κατά μέσο όρο η χρονική διάρκεια συλλογής δεδομένων ήταν της τάξης των 10 ωρών.

Κάθε μια από τις χρονολογικές σειρές που προέκυψαν εκτιμήθηκε για την εμφάνιση ή όχι αυτό-ομοιότητας με τις μεθόδους των Αθροιστικών Χρονολογικών Σειρών και της R/S Στατιστικής. Σύμφωνα με τα αποτελέσματα των περισσότερων εκτιμήσεων δίδεται μια σαφής ένδειξη για την ύπαρξη αυτό-ομοιότητας, η οποία γίνεται περισσότερο έντονη καθώς αυξάνει ο βαθμός χρήσης του δικτύου. Παρατηρήθηκε επίσης, ότι για την εκτίμηση του βαθμού αυτό-ομοιότητας είναι αδιάφορο αν η κίνηση μετριέται σε πακέτα ή ψηφιοσυλλαβές. Επίσης, παρατηρήθηκε ότι δεν υπάρχει σημαντική διαφορά στο βαθμό αυτό-ομοιότητας μεταξύ των χρονολογικών σειρών εισόδου και εξόδου μιας σύγκεκριμένης υπολογιστικής συσκευής του δικτύου (host).

### 3. ΔΕΔΟΜΕΝΑ ΚΙΝΗΣΗΣ ΑΠΟ ΤΟ ΤΟΠΙΚΟ ΔΙΚΤΥΟ ΤΟΥ ΕΡΓΑΣΤΗΡΙΟΥ ΠΟΛΥΜΕΣΩΝ

#### 3.1 ΔΙΕΞΑΓΩΓΗ ΚΑΙ ΑΠΟΤΕΛΕΣΜΑΤΑ 1<sup>ΗΣ</sup> ΜΕΤΡΗΣΗΣ

Στον πίνακα που ακολουθεί συνοψίζονται οι πληροφορίες διεξαγωγής της 1<sup>ης</sup> μέτρησης από το τοπικό δίκτυο του Εργαστηρίου Πολυμέσων του Πανεπιστημίου.

Δίκτυο διεξαγωγής μετρήσεων:	LAN Εργαστηρίου Πολυμέσων
Τοποθεσία δικτύου:	Κτίριο Δεριγνύ
Τρόπος συλλογής μετρήσεων:	HP LANprobe
Ημ/νία συλλογής:	18/11/98
Χρονική διάρκεια συλλογής:	13:03:11 – 21:02:25
Ρυθμός συλλογής (polling ανά):	10 sec
Πλήθος παρατηρήσεων Χ.Σ.:	1706
RMON MIB v1 αντικείμενα:	EtherStatsOctets, etherStatsPkts, hostInOctets, hostOutOctets, hostInPkts, hostOutPkts, matrixSDOctets, matrixSDPkts
Αριθμός Χ.Σ. που εκτιμήθηκαν ως προς H:	92 <sup>1</sup>
Επίπεδα άθροισης (m) εκτιμήσεων:	200, 400, 600, ...

Παρακάτω παρατίθενται οι πίνακες των αποτελεσμάτων των εκτιμήσεων του βαθμού αυτό-ομοιότητας των χρονολογικών σειρών που προέκυψαν.

<sup>1</sup> Από τα αρχεία χρονολογικών σειρών που δημιουργήθηκαν επιλέχθηκαν για εκτίμηση της H κατά που αφορούσαν τις πιο ενεργές (active) συσκευές του δικτύου κατά τη διάρκεια των μετρήσεων.

ΜΕΘΟΔΟΣ ΕΚΤΙΜΗΣΗΣ Η HOST		<u>HostOutOctets</u>		<u>HostOutPackets</u>	
		1. ΑΘΡΟΙΣΤΙΚΩΝ ΧΡ/ΚΩΝ ΣΕΙΡΩΝ	2. R/S ΣΤΑΤΙ- ΣΤΙΚΗΣ	3. ΑΘΡΟΙΣΤΙΚΩΝ ΧΡ/ΚΩΝ ΣΕΙΡΩΝ	4. R/S ΣΤΑΤΙ- ΣΤΙΚΗΣ
1.	PC	0.53	0.82	0.708	0.855
2.	PC	0.65	0.57	0.582	0.582
3.	PC	0.65	0.666	0.779	0.78
4.	PC	0.71	0.65	0.767	0.723
5.	PC	0.61	0.611	0.597	0.538
6.	PC	0.41	0.68	0.42	0.686
7.	File Server	0.69	0.76	0.654	0.8
8.	PC	0.59	0.738	0.706	0.77
9.	PC	0.88	0.72	0.878	0.72
10.	PC	0.79	0.77	0.819	0.81
11.	PC	0.9	0.81	0.845	0.666
12.	PC	0.63	0.747	0.767	0.761
13.	PC	0.878	0.773	0.879	0.71
14.	PC	0.88	0.758	0.89	0.77
15.	Web & Mail Server	0.64	0.69	0.521	0.68
16.	PC	0.92	0.9	0.947	0.884
17.	HP Printer	0.51	0.78	0.5	0.9
18.	PC	0.061	0.32	-0.001	0.321
19.	PC	0.133	0.345	0.1003	0.347
20.	PC	0.509	0.53	-	-
<b>Μέσος όρος:</b>		<b>0.628</b>	<b>0.681</b>	<b>0.65</b>	<b>0.7</b>

Πίνακας 1.1 Οι εκτιμήσεις της Η των χρονολογικών σειρών πακέτων και ψηφιοσυλλαβών εξόδου των 20 host.

ΜΕΘΟΔΟΣ ΕΚΤΙΜΗΣΗΣ Η HOST	<u>HostInOctets</u>		<u>HostInPackets</u>	
	1. <b>ΑΘΡΟΙΣΤΙΚΩΝ ΧΡ/ΚΩΝ ΣΕΙΡΩΝ</b>	2. <b>R/S ΣΤΑΤΙ- ΣΤΙΚΗΣ</b>	3. <b>ΑΘΡΟΙΣΤΙΚΩΝ ΧΡ/ΚΩΝ ΣΕΙΡΩΝ</b>	4. <b>R/S ΣΤΑΤΙ- ΣΤΙΚΗΣ</b>
1. PC	0.71	1.01	0.7	0.85
2. PC	0.513	0.63	0.514	0.63
3. PC	0.91	0.93	0.855	0.83
4. PC	0.586	0.71	0.73	0.717
5. PC	0.58	0.6	0.601	0.54
6. PC	0.43	0.68	0.42	0.68
7. File Server	0.56	0.84	0.61	0.803
8. PC	0.69	0.73	0.72	0.77
9. PC	0.822	0.68	0.85	0.71
10. PC	0.76	0.74	0.806	0.79
11. PC	0.808	0.65	0.847	0.66
12. PC	0.706	0.71	0.69	0.746
13. PC	0.84	0.725	0.874	0.766
14. PC	0.83	0.71	0.885	0.75
15. Web & Mail Server	0.36	0.77	0.42	0.7
16. PC	0.757	0.78	0.955	0.874
17. HP Printer	0.54	0.99	0.47	0.918
18. PC	-	-	-	-
19. PC	-	-	-	-
20. PC	0.595	0.62	0.577	0.607
21. PC	0.03	0.32	0.028	0.31
22. PC	0.892	0.709	0.955	0.77
23. PC	0.683	0.72	0.68	0.713
24. PC	0.108	0.35	0.086	0.363
25. Διεύθυνση δικτύου	0.81	0.719	0.8	0.72
<b>Μέσος όρος:</b>	<b>0.631</b>	<b>0.709</b>	<b>0.655</b>	<b>0.705</b>

Πίνακας 1.2 Οι εκτιμήσεις της Η των χρονολογικών σειρών πακέτων και ψηφιοσυλλαβών εισόδου των 23 host.

ΜΕΘΟΔΟΣ ΕΚΤΙΜΗΣΗΣ Η	<u>EtherStatsOctets</u>		<u>EtherStatsPackets</u>	
	1. <b>ΑΘΡΟΙΣΤΙΚΩΝ ΧΡ/ΚΩΝ ΣΕΙΡΩΝ</b>	2. <b>R/S ΣΤΑΤΙ- ΣΤΙΚΗΣ</b>	3. <b>ΑΘΡΟΙΣΤΙΚΩΝ ΧΡ/ΚΩΝ ΣΕΙΡΩΝ</b>	4. <b>R/S ΣΤΑΤΙ- ΣΤΙΚΗΣ</b>
<b>ΕΚΤΙΜΗΣΗ Η</b>	<b>0.687</b>	<b>0.82</b>	<b>0.742</b>	<b>0.809</b>

Πίνακας 1.3 Οι εκτιμήσεις της Η των χρονολογικών σειρών πακέτων και ψηφιοσυλλαβών που εμφανίστηκαν στο δίκτυο.

## ΣΧΟΛΙΑΣΜΟΣ ΑΠΟΤΕΛΕΣΜΑΤΩΝ 1<sup>Η</sup> ΜΕΤΡΗΣΗΣ

### A. Ο ΒΑΘΜΟΣ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ ΣΕ ΣΧΕΣΗ ΜΕ ΤΟΝ ΒΑΘΜΟ ΧΡΗΣΗΣ ΤΟΥ ΔΙΚΤΥΟΥ

Οι εκτιμήσεις των Πινάκων 1.1 και 1.2 αφορούν είκοσι και εικοσιτρείς αντίστοιχα hosts από τους εβδομήντα που εντοπίστηκαν κατά τη διάρκεια της μέτρησης. Αντίθετα, οι εκτιμήσεις του Πίνακα 1.3 αφορούν το σύνολο των host του δικτύου. Συγκρίνοντας τους Πίνακες 1.1 και 1.2 με τον Πίνακα 1.3 προκύπτει ότι η μέση τιμή των εκτιμήσεων της παραμέτρου  $H$  (βαθμός αυτό-ομοιότητας) και των τεσσάρων στηλών των Πινάκων 1.1 και 1.2 (αφορούν τις δύο διαφορετικές μεθόδους εκτίμησης της  $H$  για πακέτα και ψηφισυλλαβές) είναι μικρότερη από τις αντίστοιχες εκτιμήσεις των τεσσάρων στηλών του Πίνακα 1.3. Έτσι, αφού οι χρονολογικές σειρές του Πίνακα 1.3 αφορούν το σύνολο των host του δικτύου, επιβεβαιώνεται η πρόταση ότι όσο μεγαλύτερη είναι η κίνηση του Ethernet δικτύου, τόσο μεγαλύτερη είναι η εκτίμηση της παραμέτρου  $H$  ή τόσο μεγαλύτερος είναι ο βαθμός αυτό-ομοιότητας.

### B. ΔΙΑΦΟΡΑ ΣΤΙΣ ΕΚΤΙΜΗΣΕΙΣ ΤΗΣ $H$ ΤΩΝ ΔΥΟ ΜΕΘΟΔΩΝ ΕΚΤΙΜΗΣΗΣ

Η ύπαρξη διαφοράς μεταξύ των εκτιμήσεων της  $H$  που προκύπτουν με την μέθοδο των Αθροιστικών Χρονολογικών Σειρών και της R/S Στατιστικής είναι αναμενόμενη αφού οι δύο μέθοδοι χρησιμοποιούν διαφορετικές διαδικασίες προκειμένου να εκτιμήσουν την ίδια παράμετρο. Το γεγονός της ύπαρξης διαφοράς δεν σημαίνει ότι μια από τις δύο, ή και οι δύο μέθοδοι παρέχουν ανακριβείς εκτιμήσεις. Υπενθυμίζεται ότι οι συγκεκριμένες μέθοδοι είναι ευρεστικές και χρησιμοποιούνται με σκοπό να παρέχουν μια ένδειξη για την ύπαρξη ή όχι του φαινομένου της αυτό-ομοιότητας και όχι με σκοπό την παροχή μιας μαθηματικά ακριβούς εκτίμησης του βαθμού της αυτό-ομοιότητας. Ωστόσο, συγκρίνοντας τις στήλες των Πινάκων 1.1, 1.2 και 1.3 που αφορούν τις δύο μεθόδους εκτίμησης της  $H$  (δηλαδή τις (1,2) και (3,4)), προκύπτουν τα εξής:

- Η διαφορά στις εκτιμήσεις είναι σχετικά μικρή.
- Τα αποτελέσματα μιας μεθόδου εκτίμησης δεν εμφανίζονται εν γένει μικρότερα (ή μεγαλύτερα) από αυτά της άλλης μεθόδου εκτίμησης. Με άλλα λόγια, δεν προκύπτει μια πάγια ανισοτική σχέση μεταξύ των αποτελεσμάτων η οποία να προσδιορίζεται από την μέθοδο εκτίμησης της  $H$ .
- Το γεγονός της ύπαρξης διαφοράς εκτιμήσεων δεν δημιουργεί προβλήματα «συγκρούσεων»: Όταν η εκτίμηση της μίας μεθόδου υποδεικνύει μεγάλο (ή μικρό) βαθμό αυτό-ομοιότητας, τότε αναμένεται κατ' αντιστοιχία μεγάλος (ή μικρός) βαθμός αυτό-ομοιότητας και για την εκτίμηση της άλλης μεθόδου.

## Γ. ΔΙΑΦΟΡΑ ΣΤΙΣ ΕΚΤΙΜΗΣΕΙΣ ΤΗΣ $H$ ΜΕΤΑΞΥ ΤΗΣ ΚΙΝΗΣΗΣ ΠΛΑΚΕΤΩΝ & ΨΗΦΙΟΣΥΛΛΑΒΩΝ

Η διαφορά στις εκτιμήσεις μεταξύ των χρονολογικών σειρών της κίνησης η οποία μετράται σε πακέτα και των χρονολογικών σειρών της ίδιας κίνησης η οποία μετράται σε ψηφιοσυλλαβές (bytes) οφείλεται στο γεγονός ότι τα πακέτα δεν αποτελούνται από σταθερό αριθμό ψηφιοσυλλαβών. Συγκρίνοντας στους Πίνακες 1.1, 1.2 και 1.3 τις στήλες (1,3) ή (2,4) δηλαδή τις στήλες που αφορούν ψηφιοσυλλαβές με τις στήλες που αφορούν πακέτα και οι οποίες προκύπτουν από την ίδια μέθοδο εκτίμησης, προκύπτουν τα εξής:

- Η διαφορά στις εκτιμήσεις είναι σχετικά μικρή.
- Οι εκτιμήσεις των χρονολογικών σειρών που αφορούν πακέτα δεν εμφανίζονται εν γένει μικρότερες (ή μεγαλύτερες) από αυτές που αφορούν ψηφιοσυλλαβές. Με άλλα λόγια, δεν προκύπτει μια πάγια ανισοτική σχέση μεταξύ των εκτιμήσεων η οποία να προσδιορίζεται από τον τρόπο μέτρησης της κίνησης (σε πακέτα ή ψηφιοσυλλαβές).
- Το γεγονός της ύπαρξης διαφοράς εκτιμήσεων δεν δημιουργεί προβλήματα «συγκρούσεων»: Όταν η εκτίμηση μιας συγκεκριμένης μεθόδου υποδεικνύει μεγάλο (ή μικρό) βαθμό αυτό-ομοιότητας για μια χρονολογική σειρά πακέτων, τότε αναμένεται μεγάλος (ή μικρός) βαθμός αυτό-ομοιότητας και για την αντίστοιχη χρονολογική σειρά η οποία είναι μετρημένη σε ψηφιοσυλλαβές.

## Δ. ΔΙΑΦΟΡΑ ΣΤΙΣ ΕΚΤΙΜΗΣΕΙΣ ΤΗΣ $H$ ΓΙΑ ΚΙΝΗΣΗ ΕΙΣΟΔΟΥ & ΕΞΟΔΟΥ

Οι είκοσι πρώτοι hosts του Πίνακα 1.2, είναι οι ίδιοι με τους hosts του Πίνακα 1.1. Η διαφορά έγκειται στο ότι στον Πίνακα 1.1 εκτιμάται ο βαθμός αυτό-ομοιότητας της κίνησης που δέχεται κάθε host, ενώ στον Πίνακα 1.2 εκτιμάται ο βαθμός αυτό-ομοιότητας της κίνησης που αποστέλλει κάθε host. Συγκρίνοντας για κάθε host τις εκτιμήσεις της  $H$  για την κίνηση εισόδου και εξόδου (δηλαδή τις είκοσι πρώτες γραμμές των Πινάκων 1.1 και 1.2), προκύπτει ότι:

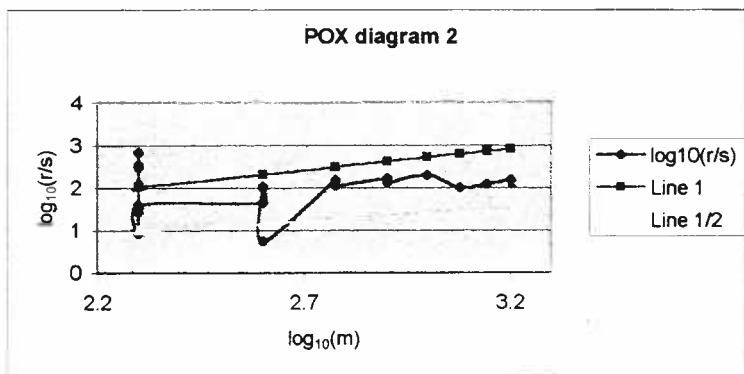
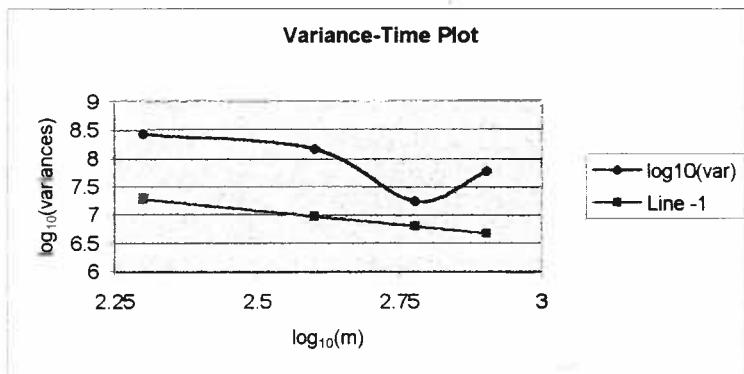
Οι εκτιμήσεις του βαθμού αυτό-ομοιότητας της κίνησης εξόδου για έναν host είναι ανάλογες των εκτιμήσεων της κίνησης εισόδου για τον ίδιο host. Η παρατήρηση ισχύει και για τις δύο μεθόδους εκτίμησης της  $H$ , και για τους δύο τρόπους έκφρασης της κίνησης (πακέτα ή ψηφιοσυλλαβές). Για παράδειγμα συγκρίνοντας τις γραμμές 11 των Πινάκων 1.1 και 1.2, προκύπτει η ομοιομορφία των εκτιμήσεων της  $H$  για την κίνηση εισόδου και εξόδου.

## E. ΕΚΤΙΜΗΣΗ ΤΟΥ ΒΑΘΜΟΥ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ ΤΗΣ ΚΙΝΗΣΗΣ ΤΟΥ ΔΙΚΤΥΟΥ

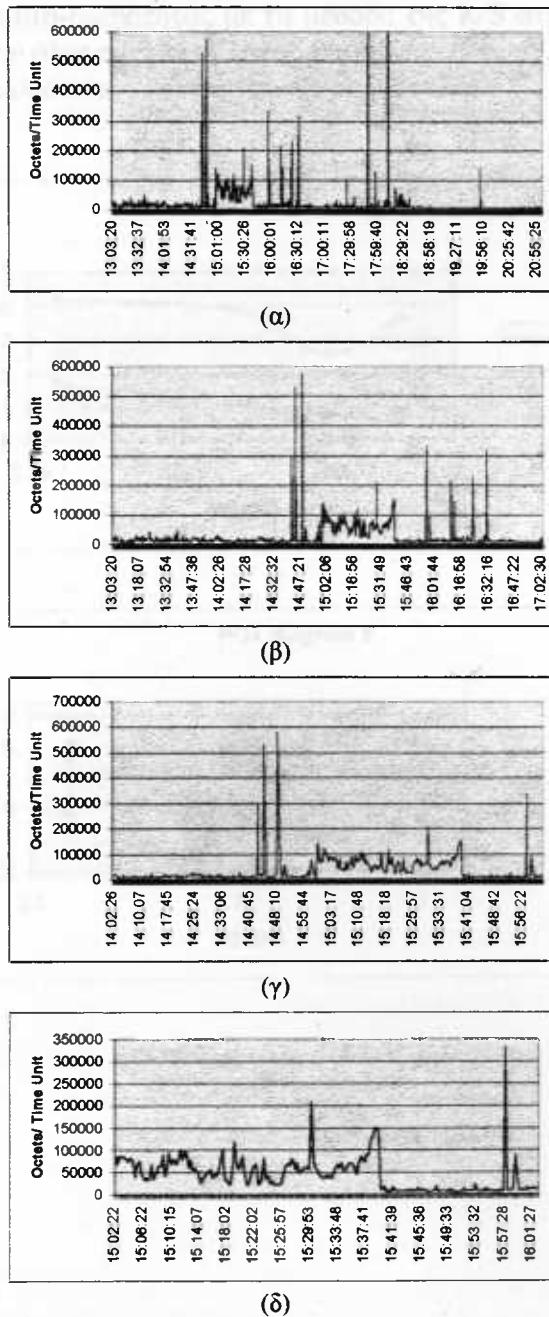
Από τον Πίνακα 1.3 προκύπτει ότι οι εκτιμήσεις του βαθμού αυτό-ομοιότητας των χρονολογικών σειρών των πακέτων και των ψηφιοσυλλαβών που εμφανίστηκαν στο δίκτυο κατά τη διάρκεια των μετρήσεων, είναι αρκετά υψηλές (αρκετά μεγαλύτερες από 0.5). Το γεγονός αυτό ισχύει και για τις δύο μεθόδους εκτίμησης και δίνει μια ένδειξη για την ύπαρξη αυτό-ομοιότητας, όπως προκύπτει και από τα παρακάτω διαγράμματα.

### E.1 ΧΡΟΝΟΛΟΓΙΚΗ ΣΕΙΡΑ ΨΗΦΙΟΣΥΛΛΑΒΩΝ

Παρακάτω παρατίθονται τα Variance-Time και POX διαγράμματα τύπου 2 της χρονολογικής σειράς *EtherStatsOctets*. Από το Variance-Time διάγραμμα φαίνεται ότι οι λογαριθμημένες διακυμάνσεις βρίσκονται πάνω από την γραμμή με κλίση -1, γεγονός που δικαιολογεί μια εκτίμηση για το  $H > 0.5$ . Από το POX διάγραμμα προκύπτει ότι οι λογαριθμημένες  $r/s$  τιμές βρίσκονται κατά πλειοψηφία μεταξύ των γραμμών με κλίση 1 και  $\frac{1}{2}$ , γεγονός που φανερώνει μεγάλο βαθμό αυτό-ομοιότητας.



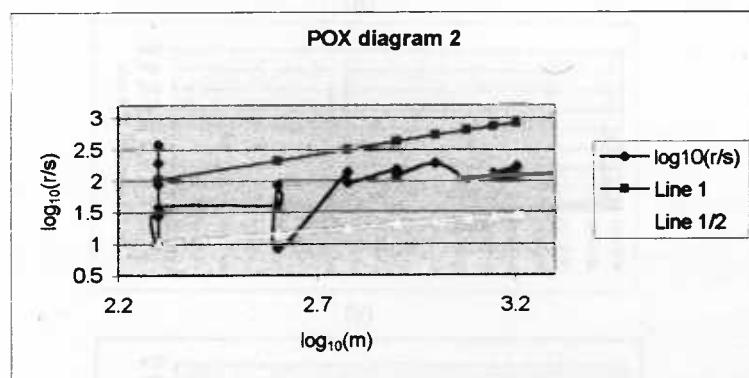
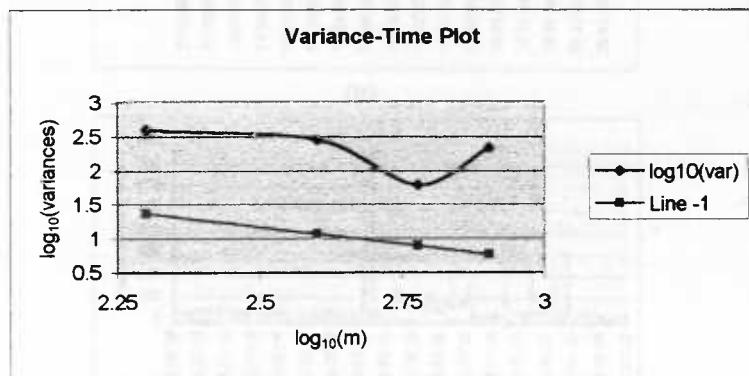
Το Σχήμα 6.1(α)-(δ) αφορά τη χρονολογική σειρά EtherStatsOctets. Στο διάγραμμα 6.1(α) παριστάνεται γραφικά το σύνολο των παρατηρήσεων της χρονολογικής σειράς. Στο διάγραμμα 6.1(β) παριστάνεται το πρώτο μισό των παρατηρήσεων του διαγράμματος 6.1(α). Στο διάγραμμα 6.1(γ) παριστάνεται γραφικά ένα τυχαίο τμήμα το οποίο αποτελείται από το ήμισυ των παρατηρήσεων του διαγράμματος 6.1 (β), ενώ στο τελευταίο διάγραμμα παριστάνεται το πρώτο μισό παρατηρήσεων του διαγράμματος 6.1 (γ). Και τα τέσσερα διαγράμματα αν και αναφέρονται σε διαφορετικές κλίμακες χρόνου, εμφανίζουν πανομοιότυπο πρότυπο (pattern) κίνησης. Το γεγονός αυτό αποτελεί μια γρήγορη «οπτική» απόδειξη σχετικά με την ύπαρξη αυτό-ομοιότητας στα δεδομένα.



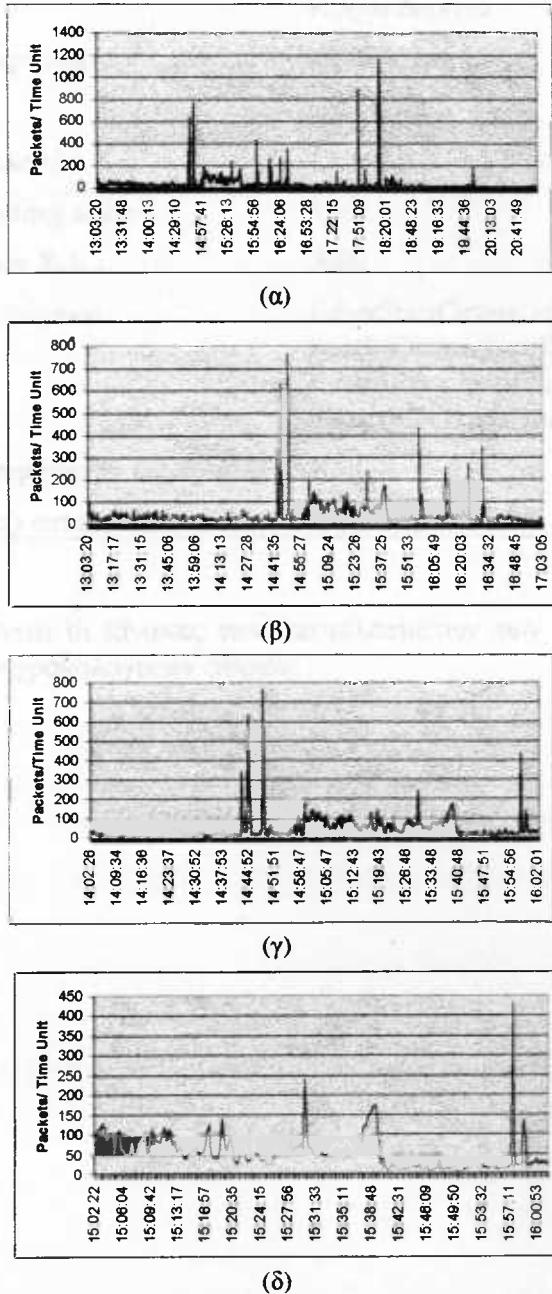
**Σχήμα 6.1** Μία «οπτική» απόδειξη για την ύπαρξη αυτό-ομοιότητας στα δεδομένα της χρονολογικής σειράς EtherStatsOctets: Και τα τέσσερα διαγράμματα που αφορούν διαφορετικές χρονικές κλίμακες της ίδιας χρονολογικής σειράς εμφανίζουν πανομοιότυπο πρότυπο (pattern) κίνησης.

## E.2 ΧΡΟΝΟΛΟΓΙΚΗ ΣΕΙΡΑ ΠΑΚΕΤΩΝ

Παρακάτω παρατίθονται τα Variance-Time και POX διαγράμματα της χρονολογικής σειράς *EtherStatsPkts*. Από το Variance-Time διάγραμμα φαίνεται ότι οι λογαριθμημένες διακυμάνσεις βρίσκονται πάνω από τη γραμμή με κλίση -1, και σε μεγαλύτερη απόσταση από αυτές του αντίστοιχου διαγράμματος της χρονολογικής σειράς *EtherStatsOctets*. Έτσι δικαιολογείται και μια μεγαλύτερη εκτίμηση του βαθμού αυτο-ομοιότητας για την χρονολογική σειρά *EtherStatsPkts* με τη μέθοδο των αθροιστικών χρονολογικών σειρών, από αυτήν της χρονολογικής σειράς *EtherStatsOctets*. Η μεγάλη εκτίμηση του βαθμού αυτο-ομοιότητας με τη μέθοδο της R/S στατιστικής φαίνεται από το POX διάγραμμα όπου όλες σχεδόν οι λογαριθμημένες r/s τιμές βρίσκονται μεταξύ των γραμμών 1 και  $\frac{1}{2}$  και μάλιστα πιο κοντά στη γραμμή με κλίση 1.



Το Σχήμα 6.2(α)-(δ) αφορά τη χρονολογική σειρά EtherStatsPkts. Το διάγραμμα 6.2(α) αφορά το σύνολο των παρατηρήσεων της. Κάθε επόμενο διάγραμμα προκύπτει από το προηγούμενο απεικονίζοντας γραφικά τυχαίο διάστημα τιμών που αντιστοιχεί στο ήμισυ των παρατηρήσεων του προηγούμενου διαγράμματος. Το γεγονός ότι και τα τέσσερα διαγράμματα, παρά το γεγονός ότι αντιτροσωπεύουν διαφορετικές κλίμακες χρόνου διατηρούν το ίδιο πρότυπο (pattern) κίνησης, αποτελεί μια οπτική απόδειξη για την ύπαρξη αυτο-ομοιότητας.



Σχήμα 6.2 Μία «οπτική» απόδειξη για την ύπαρξη αυτό-ομοιότητας στα δεδομένα της χρονολογικής σειράς EtherStatsPkts: Και τα τέσσερα διαγράμματα που αφορούν διαφορετικές χρονικές κλίμακες της ίδιας Χ.Σ. εμφανίζουν πανομοιότυπο πρότυπο (pattern) κίνησης

### 3.2 ΔΙΕΞΑΓΩΓΗ ΚΑΙ ΑΠΟΤΕΛΕΣΜΑΤΑ 2<sup>nd</sup> ΜΕΤΡΗΣΗΣ

Στον πίνακα που ακολουθεί συνοψίζονται οι πληροφορίες διεξαγωγής της 2<sup>nd</sup> μέτρησης από το τοπικό δίκτυο του Εργαστηρίου Πολυμέσων του Πανεπιστημίου.

Δίκτυο διεξαγωγής μετρήσεων:	LAN Εργαστηρίου Πολυμέσων
Τοποθεσία δικτύου:	Κτίριο Δεριγνύ
Τρόπος συλλογής μετρήσεων:	HP LANprobe
Ημ/νία συλλογής:	20/11/98
Χρονική διάρκεια συλλογής:	13:36:19 – 18:52:53
Ρυθμός συλλογής (polling ανά):	5 sec
Πλήθος παρατηρήσεων Χ.Σ.:	3080
RMON MIB v1 αντικείμενα:	EtherStatsOctets, etherStatsPkts, hostInOctets, hostOutOctets, hostInPkts, hostOutPkts, matrixSDOctets, matrixSDPkts
Αριθμός Χ.Σ. που εκτιμήθηκαν ως προς H:	130 <sup>1</sup>
Επίπεδα άθροισης (m) εκτιμήσεων:	400, 800, 1200, ...

Παρακάτω παρατίθενται οι πίνακες των αποτελεσμάτων των εκτιμήσεων του βαθμού αυτό-ομοιότητας των χρονολογικών σειρών.

<sup>1</sup> Από τα αρχεία χρονολογικών σειρών που δημιουργήθηκαν επιλέχθηκαν για εκτίμηση της H αυτά που αφορούσαν τις πιο ενεργές (active) συσκευές του δικτύου κατά τη διάρκεια των μετρήσεων.

ΜΕΘΟΔΟΣ ΕΚΤΙΜΗΣΗΣ Η ΧΡΟΝΟΛΟΓΙΚΕΣ ΣΕΙΡΕΣ	<u>HostOutOctets</u>		<u>HostOutPackets</u>	
	1. <b>ΑΘΡΟΙΣΤΙΚΩΝ ΧΡ/ΚΩΝ ΣΕΙΡΩΝ</b>	2. <b>R/S ΣΤΑΤΙ- ΣΤΙΚΗΣ</b>	3. <b>ΑΘΡΟΙΣΤΙΚΩΝ ΧΡ/ΚΩΝ ΣΕΙΡΩΝ</b>	4. <b>R/S ΣΤΑΤΙ- ΣΤΙΚΗΣ</b>
1. PC	0.54	0.76	0.708	0.794
2. PC	0.53	0.51	0.371	0.541
3. PC	0.639	0.544	0.852	0.767
4. PC	0.827	1.13	0.83	1.12
5. PC	0.747	0.683	0.716	0.665
6. PC	0.721	0.736	0.722	0.735
7. PC	0.67	0.898	0.628	0.772
8. PC	0.594	0.736	0.618	0.745
9. PC	0.768	0.72	0.745	0.741
10. PC	0.604	0.766	0.599	0.733
11. PC	0.788	0.725	0.79	0.729
12. PC	0.446	0.815	0.486	0.845
13. Router	0.964	0.88	0.908	0.864
14. PC	0.826	0.73	0.846	0.748
15. PC	0.692	0.828	0.677	0.843
16. PC	0.894	0.783	0.899	0.791
17. PC	0.795	0.866	0.787	0.826
18. PC	0.78	0.727	0.785	0.723
19. PC	0.68	0.913	0.764	0.919
20. PC	0.634	0.763	0.701	0.812
21. PC	0.854	0.92	0.88	0.938
22. PC	0.834	0.74	0.812	0.737
23. PC	0.644	0.853	0.603	0.818
24. PC	0.774	0.77	0.703	0.737
25. PC	0.859	0.824	0.889	0.822
26. File server	0.725	0.87	0.752	0.863
27. Web & mail server	0.477	0.668	0.906	0.858
28. PC	0.854	0.929	0.854	0.929
29. HP printer	0.628	0.615	0.646	0.655
30. PC	0.104	0.384	0.125	0.378
31. PC	0.094	0.368	0.159	0.388
32. PC	0.633	0.728	0.617	0.713
<b>Μέσος όρος:</b>	<b>0.675</b>	<b>0.755</b>	<b>0.699</b>	<b>0.767</b>

Πίνακας 2.1 Οι εκτιμήσεις της Η των χρονολογικών σειρών πακέτων και ψηφιοσυλλαβών εξόδου των 32 host.

<b>ΜΕΘΟΔΟΣ ΕΚΤΙΜΗΣΗΣ Η ΧΡΟΝΟΛΟΓΙΚΕΣ ΣΕΙΡΕΣ</b>		<b><i>HostInOctets</i></b>		<b><i>HostInPackets</i></b>	
		<b>ΑΘΡΟΙΣΤΙΚΩΝ ΧΡ/ΚΩΝ ΣΕΙΡΩΝ</b>	<b>R/S ΣΤΑΤΙ- ΣΤΙΚΗΣ</b>	<b>ΑΘΡΟΙΣΤΙΚΩΝ ΧΡ/ΚΩΝ ΣΕΙΡΩΝ</b>	<b>R/S ΣΤΑΤΙ- ΣΤΙΚΗΣ</b>
1.	PC	0.709	0.792	0.701	0.791
2.	PC	0.514	0.806	0.514	0.806
3.	PC	0.861	0.787	0.861	0.787
4.	PC	0.836	0.984	0.832	1.00
5.	PC	0.739	0.683	0.69	0.652
6.	PC	0.721	0.734	0.721	0.735
7.	PC	0.628	0.772	0.598	0.761
8.	PC	0.587	0.714	0.61	0.74
9.	PC	0.707	0.781	0.725	0.755
10.	PC	0.514	0.663	0.514	0.663
11.	PC	0.765	0.706	0.779	0.723
12.	PC	0.563	0.808	0.52	0.825
13.	Router	0.629	0.766	0.866	0.861
14.	PC	0.829	0.726	0.842	0.742
15.	PC	0.621	0.787	0.63	0.814
16.	PC	0.866	0.76	0.888	0.782
17.	PC	0.828	0.996	0.815	0.863
18.	PC	0.768	0.707	0.772	0.712
19.	PC	0.666	0.922	0.777	0.921
20.	PC	0.675	0.796	0.68	0.805
21.	PC	0.813	0.893	0.871	0.934
22.	PC	0.801	0.719	0.774	0.717
23.	PC	0.523	0.879	0.563	0.852
24.	PC	0.535	0.662	0.634	0.713
25.	PC	0.738	0.805	0.883	0.816
26.	File server	0.582	0.782	0.769	0.866
27.	Web & mail server	0.968	0.881	0.927	0.866
28.	PC	0.844	0.927	0.849	0.929
29.	HP printer	0.644	0.691	0.65	0.668
30.	PC	-	-	-	-
31.	PC	-	-	-	-
32.	PC	0.399	0.73	0.628	0.763
<b>Μέσος όρος:</b>		<b>0.695</b>	<b>0.762</b>	<b>0.729</b>	<b>0.795</b>

Πίνακας 2.2 Οι εκτιμήσεις της  $H$  των χρονολογικών σειρών πακέτων και ψηφιοσυλλαβών εισόδου των 30 host.

ΜΕΘΟΔΟΣ ΕΚΤΙΜΗΣΗΣ <i>H</i>	<i>EtherStatsOctets</i>		<i>EtherStatsPackets</i>	
	ΑΘΡΟΙΣΤΙΚΩΝ ΧΡ/ΚΩΝ ΣΕΙΡΩΝ	R/S ΣΤΑΤΙ- ΣΤΙΚΗΣ	ΑΘΡΟΙΣΤΙΚΩΝ ΧΡ/ΚΩΝ ΣΕΙΡΩΝ	R/S ΣΤΑΤΙ- ΣΤΙΚΗΣ
	ΕΚΤΙΜΗΣΗ <i>H</i>	0.879	0.905	0.867

Πίνακας 2.3 Οι εκτιμήσεις της *H* των χρονολογικών σειρών πακέτων και ψηφιοσυλλαβών που εμφανίστηκαν στο δίκτυο.

### ΣΧΟΛΙΑΣΜΟΣ ΑΠΟΤΕΛΕΣΜΑΤΩΝ 2ΗΣ ΜΕΤΡΗΣΗΣ

#### A. Ο ΒΑΘΜΟΣ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ ΣΕ ΣΧΕΣΗ ΜΕ ΤΟΝ ΒΑΘΜΟ ΧΡΗΣΗΣ ΤΟΥ ΔΙΚΤΥΟΥ

Οι εκτιμήσεις των Πινάκων 2.1 και 2.2 αφορούν τους 32 πιο ενεργούς από τους 70 hosts που εντοπίστηκαν κατά τη διάρκεια της μέτρησης. Αντίθετα, οι εκτιμήσεις του Πίνακα 2.3 αφορούν το σύνολο των host του δικτύου. Συγκρίνοντας τους Πίνακες 2.1 και 2.2 με τον Πίνακα 2.3 προκύπτει ότι η μέση τιμή των εκτιμήσεων της παραμέτρου *H* (βαθμός αυτό-ομοιότητας) και των τεσσάρων στηλών των Πινάκων 2.1 και 2.2 (αφορούν τις δύο διαφορετικές μεθόδους εκτίμησης της *H* για πακέτα και ψηφιοσυλλαβές) είναι μικρότερη από τις αντίστοιχες εκτιμήσεις των τεσσάρων στηλών του Πίνακα 2.3. Έτσι, αφού οι χρονολογικές σειρές του Πίνακα 2.3 αφορούν το σύνολο των host του δικτύου, επιβεβαιώνεται η πρόταση ότι όσο μεγαλύτερη είναι η κίνηση του Ethernet δικτύου, τόσο μεγαλύτερη είναι η εκτίμηση της παραμέτρου *H* ή τόσο μεγαλύτερος είναι ο βαθμός αυτό-ομοιότητας.

#### B. ΔΙΑΦΟΡΑ ΣΤΙΣ ΕΚΤΙΜΗΣΕΙΣ ΤΗΣ *H* ΤΩΝ ΔΥΟ ΜΕΘΟΔΩΝ ΕΚΤΙΜΗΣΗΣ

Συγκρίνοντας τις στήλες των Πινάκων 2.1, 2.2 και 2.3 που αφορούν τις δύο μεθόδους εκτίμησης της *H* (δηλαδή τις (1,2) και (3,4)), προκύπτουν τα εξής:

- Η διαφορά στις εκτιμήσεις είναι σχετικά μικρή.
- Οι εκτιμήσεις μιας μεθόδου εκτίμησης δεν εμφανίζονται εν γένει μικρότερες (ή μεγαλύτερες) από αυτές της άλλης μεθόδου εκτίμησης. Με άλλα λόγια, δεν προκύπτει μια πάγια ανισοτική σχέση μεταξύ των εκτιμήσεων η οποία να προσδιορίζεται από την μέθοδο εκτίμησης της *H*.
- Το γεγονός της ύπαρξης διαφοράς εκτιμήσεων δεν δημιουργεί προβλήματα «συγκρούσεων»: Όταν η εκτίμηση της μίας μεθόδου υποδεικνύει μεγάλο (ή μικρό) βαθμό αυτό-ομοιότητας, τότε αναμένεται κατ' αντιστοιχία μεγάλος (ή μικρός) βαθμός αυτό-ομοιότητας και για την εκτίμηση της άλλης μεθόδου.



## **Γ. ΔΙΑΦΟΡΑ ΣΤΙΣ ΕΚΤΙΜΗΣΕΙΣ ΤΗΣ $H$ ΜΕΤΑΞΥ ΤΗΣ ΚΙΝΗΣΗΣ ΠΑΚΕΤΩΝ & ΨΗΦΙΟΣΥΛΛΑΒΩΝ**

Συγκρίνοντας στους Πίνακες 2.1, 2.2 και 2.3 τις στήλες (1,3) ή (2,4) δηλαδή τις στήλες που αφορούν ψηφιοσυλλαβές με τις στήλες που αφορούν πακέτα και οι οποίες προκύπτουν από την ίδια μέθοδο εκτίμησης, προκύπτουν τα εξής:

- Η διαφορά στις εκτιμήσεις είναι σχετικά μικρή.
- Οι εκτιμήσεις των χρονολογικών σειρών που αφορούν πακέτα δεν εμφανίζονται εν γένει μικρότερες (ή μεγαλύτερες) από αυτές που αφορούν ψηφιοσυλλαβές. Με άλλα λόγια, δεν προκύπτει μια πάγια ανισοτική σχέση μεταξύ των εκτιμήσεων η οποία να προσδιορίζεται από τον τρόπο μέτρησης της κίνησης (σε πακέτα ή ψηφιοσυλλαβές).
- Το γεγονός της ύπαρξης διαφοράς εκτιμήσεων δεν δημιουργεί προβλήματα «συγκρούσεων»: Όταν η εκτίμηση μιας συγκεκριμένης μεθόδου υποδεικνύει μεγάλο (ή μικρό) βαθμό αυτό-ομοιότητας για μια χρονολογική σειρά πακέτων, τότε αναμένεται μεγάλος (ή μικρός) βαθμός αυτό-ομοιότητας και για την αντίστοιχη χρονολογική σειρά η οποία είναι μετρημένη σε ψηφιοσυλλαβές.

## **Δ. ΔΙΑΦΟΡΑ ΣΤΙΣ ΕΚΤΙΜΗΣΕΙΣ ΤΗΣ $H$ ΓΙΑ ΚΙΝΗΣΗ ΕΙΣΟΔΟΥ & ΕΞΟΔΟΥ**

Οι hosts του Πίνακα 2.2, είναι οι ίδιοι με τους hosts του Πίνακα 2.1. Η διαφορά έγκειται στο ότι στον Πίνακα 2.1 εκτιμάται ο βαθμός αυτό-ομοιότητας της κίνησης που δέχεται κάθε host, ενώ στον Πίνακα 2.2 εκτιμάται ο βαθμός αυτό-ομοιότητας της κίνησης που αποστέλλει κάθε host. Συγκρίνοντας για κάθε host τις εκτιμήσεις της  $H$  για την κίνηση εισόδου και εξόδου (δηλαδή τις γραμμές των Πινάκων 2.1 και 2.2), προκύπτει ότι:

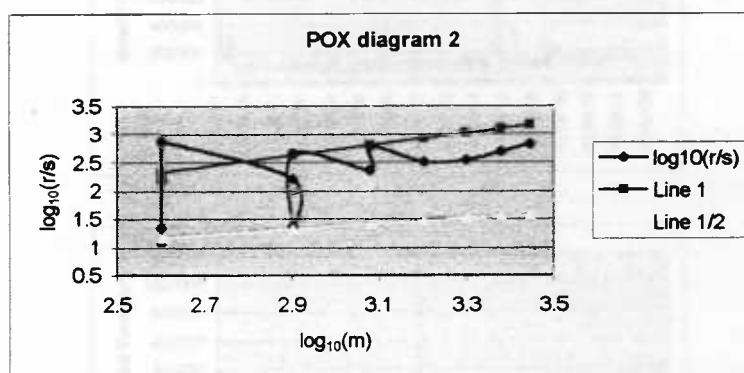
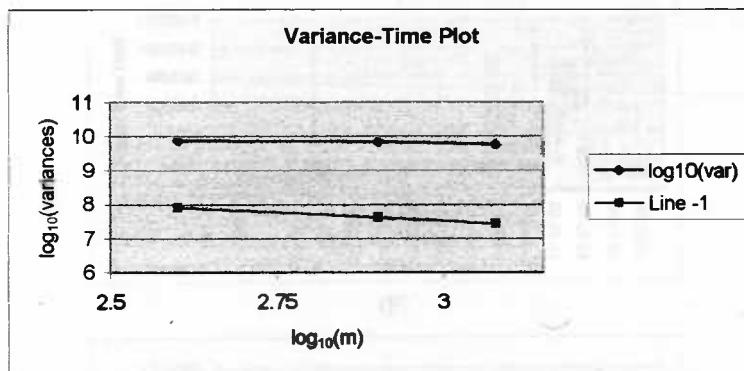
Οι εκτιμήσεις του βαθμού αυτό-ομοιότητας της κίνησης εξόδου για έναν host είναι ανάλογες των εκτιμήσεων της κίνησης εισόδου για τον ίδιο host. Η παρατήρηση ισχύει και για τις δύο μεθόδους εκτίμησης της  $H$ , και για τους δύο τρόπους έκφρασης της κίνησης (πακέτα ή ψηφιοσυλλαβές). Για παράδειγμα συγκρίνοντας τις γραμμές 16 των Πινάκων 2.1 και 2.2, προκύπτει η ομοιομορφία των εκτιμήσεων της  $H$  για την κίνηση εισόδου και εξόδου.

## E. ΕΚΤΙΜΗΣΗ ΤΟΥ ΒΑΘΜΟΥ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ ΤΗΣ ΚΙΝΗΣΗΣ ΤΟΥ ΔΙΚΤΥΟΥ

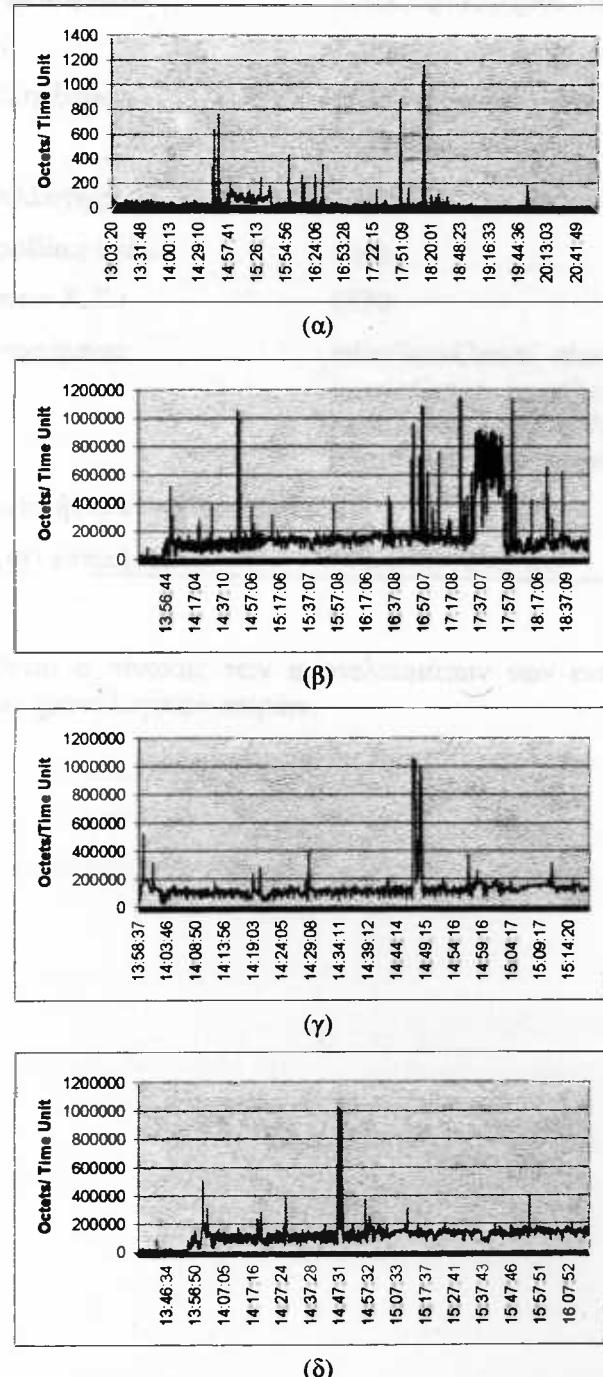
Από τον Πίνακα 2.3 προκύπτει ότι οι εκτιμήσεις του βαθμού αυτό-ομοιότητας των χρονολογικών σειρών των πακέτων και των ψηφιοσυλλαβών που εμφανίστηκαν στο δίκτυο κατά τη διάρκεια των μετρήσεων, είναι αρκετά υψηλές (αρκετά μεγαλύτερες από 0.5). Το γεγονός αυτό ισχύει και για τις δύο μεθόδους εκτίμησης και δίνει μια ένδειξη για την ύπαρξη αυτό-ομοιότητας, όπως προκύπτει και από τα παρακάτω διαγράμματα.

### E.1 ΧΡΟΝΟΛΟΓΙΚΗ ΣΕΙΡΑ ΨΗΦΙΟΣΥΛΛΑΒΩΝ

Παρακάτω παρατίθονται τα Variance-Time και POX διαγράμματα της χρονολογικής σειράς *EtherStatsOctets*. Από το Variance-Time διάγραμμα φαίνεται ότι οι λογαριθμημένες διακυμάνσεις βρίσκονται αρκετά πάνω από τη γραμμή με κλίση -1, γεγονός που δικαιολογεί μια αρκετά μεγάλη εκτίμηση για την *H*. Από το POX διάγραμμα τύπου 2 προκύπτει ότι οι λογαριθμημένες *r/s* τιμές βρίσκονται κατά πλειοψηφία μεταξύ των γραμμών με κλίση 1 και  $\frac{1}{2}$ , γεγονός που φανερώνει μεγάλο βαθμό αυτό-ομοιότητας.



Το Σχήμα 6.3(α)-(δ) αφορά τη χρονολογική σειρά EtherStatsPkts. Το διάγραμμα (α) αφορά το σύνολο των παρατηρήσεων της. Κάθε επόμενο διάγραμμα προκύπτει από το προηγούμενο απεικονίζοντας γραφικά ένα τυχαίο διάστημα τιμών που αντιστοιχεί στο ήμισυ των παρατηρήσεων του προηγούμενου διαγράμματος. Το γεγονός ότι και τα τέσσερα διαγράμματα, παρά το γεγονός ότι αντιπροσωπεύουν διαφορετικές κλίμακες χρόνου διατηρούν το ίδιο πρότυπο (pattern) κίνησης, αποτελεί μια οπτική απόδειξη για την ύπαρξη αυτό-ομοιότητας.



Σχήμα 6.3 Μία «οπτική» απόδειξη για την ύπαρξη αυτό-ομοιότητας στα δεδομένα της χρονολογικής σειράς EtherStatsPkts. Και τα τέσσερα διαγράμματα που αφορούν διαφορετικές χρονικές κλίμακες της ίδιας χρονολογικής σειράς εμφανίζουν πανομοιότυπο πρότυπο (pattern) κίνησης.

### 3.3 ΔΙΕΞΑΓΩΓΗ ΚΑΙ ΑΠΟΤΕΛΕΣΜΑΤΑ 3<sup>ΗΣ</sup> ΜΕΤΡΗΣΗΣ

Στον πίνακα που ακολουθεί συνοψίζονται οι πληροφορίες διεξαγωγής της 3<sup>ης</sup> μέτρησης από το τοπικό δίκτυο του Εργαστηρίου Πολυμέσων του Πανεπιστημίου.

Δίκτυο διεξαγωγής μετρήσεων:	LAN Εργαστηρίου Πολυμέσων
Τοποθεσία δικτύου:	Κτίριο Δεριγνύ
Τρόπος συλλογής μετρήσεων:	HP LANprobe
Ημ/νία συλλογής:	25/11/98
Χρονική διάρκεια συλλογής:	12:28:01 – 21:44:00
Ρυθμός συλλογής (polling ανά):	5 sec
Πλήθος παρατηρήσεων Χ.Σ.:	6570
RMON MIB v1 αντικείμενα:	etherStatsOctets, etherStatsPkts, hostInOctets, hostOutOctets, hostInPkts, hostOutPkts, matrixSDOctets, matrixSDPkts
Αριθμός Χ.Σ. που εκτιμήθηκαν ως προς Η:	4
Επίπεδα άθροισης (m) εκτιμήσεων:	400, 800, 1200, ...

Παρακάτω παρατίθεται ο πίνακας των αποτελεσμάτων των εκτιμήσεων του βαθμού αυτό-ομοιότητας των χρονολογικών σειρών.

Συγχρόνως με την έκθεση 2.8 της απήντη (1.3) ή (2.4) παρέχεται η στήλη για ανάλυση της αυτό-ομοιότητας με τις απόλειτες αποστολές παρατηρήσεων από την παρατηρητική πλατφόρμα της Εθνικής Τεχνολογίας.

- Η παρατηρητική πλατφόρμα διαθέτει δύο είδη αναλυτικών προβολών: Ένα για αποστολές παρατηρήσεων. Όσεν η επεξιπτώτη μεταποιούμενη, μεθόδους παρατηρήσεων μετάλλητη ή μη μετάλλητη παρατηρησητική τιμή παρατηρήσεων, τόσο για αποστολές παρατηρήσεων όσο και για την παρατηρησητική τιμή παρατηρήσεων από την αναλυτική πλατφόρμα.
- Η παρατηρητική πλατφόρμα διαθέτει δύο είδη αναλυτικών προβολών: Ένα για αποστολές παρατηρήσεων. Όσεν η επεξιπτώτη μεταποιούμενη, μεθόδους παρατηρήσεων μετάλλητη ή μη μετάλλητη παρατηρησητική τιμή παρατηρήσεων, τόσο για αποστολές παρατηρήσεων όσο και για την παρατηρησητική τιμή παρατηρήσεων από την αναλυτική πλατφόρμα.

ΜΕΘΟΔΟΣ ΕΚΤΙΜΗΣΗΣ Η	<u>EtherStatsOctets</u>		<u>EtherStatsPackets</u>	
	1. <b>ΑΘΡΟΙΣΤΙΚΩΝ ΧΡ/ΚΩΝ ΣΕΙΡΩΝ</b>	2. <b>R/S ΣΤΑΤΙ- ΣΤΙΚΗΣ</b>	3. <b>ΑΘΡΟΙΣΤΙΚΩΝ ΧΡ/ΚΩΝ ΣΕΙΡΩΝ</b>	4. <b>R/S ΣΤΑΤΙ- ΣΤΙΚΗΣ</b>
<b>ΕΚΤΙΜΗΣΗ Η</b>	0.728	0.801	0.743	0.784

Πίνακας 3.1 Οι εκτιμήσεις της Η των χρονολογικών σειρών πακέτων και ψηφιοσυλλαβών που εμφανίστηκαν στο δίκτυο.

### **ΣΧΟΛΙΑΣΜΟΣ ΑΠΟΤΕΛΕΣΜΑΤΩΝ 3ΗΣ ΜΕΤΡΗΣΗΣ**

#### **A. ΔΙΑΦΟΡΑ ΣΤΙΣ ΕΚΤΙΜΗΣΕΙΣ ΤΗΣ Η ΤΩΝ ΔΥΟ ΜΕΘΟΔΩΝ ΕΚΤΙΜΗΣΗΣ**

Συγκρίνοντας τις στήλες του Πίνακα 3.1 που αφορούν τις δύο μεθόδους εκτίμησης της Η (δηλαδή τις (1,2) και (3,4)), προκύπτουν τα εξής:

- Η διαφορά στις εκτιμήσεις είναι σχετικά μικρή.
- Το γεγονός της ύπαρξης διαφοράς εκτιμήσεων δεν δημιουργεί προβλήματα «συγκρούσεων»: Όταν η εκτίμηση της μίας μεθόδου υποδεικνύει μεγάλο (ή μικρό) βαθμό αυτό-ομοιότητας, τότε αναμένεται κατ' αντίστοιχία μεγάλος (ή μικρός) βαθμός αυτό-ομοιότητας και για την εκτίμηση της άλλης μεθόδου.

#### **B. ΔΙΑΦΟΡΑ ΣΤΙΣ ΕΚΤΙΜΗΣΕΙΣ ΤΗΣ Η ΜΕΤΑΞΥ ΤΗΣ ΚΙΝΗΣΗΣ ΠΑΚΕΤΩΝ & ΨΗΦΙΟΣΥΛΛΑΒΩΝ**

Συγκρίνοντας στον Πίνακα 3.1 τις στήλες (1,3) ή (2,4) δηλαδή τις στήλες που αφορούν ψηφιοσυλλαβές με τις στήλες που αφορούν πακέτα και οι οποίες προκύπτουν από την ίδια μέθοδο εκτίμησης, προκύπτουν τα εξής:

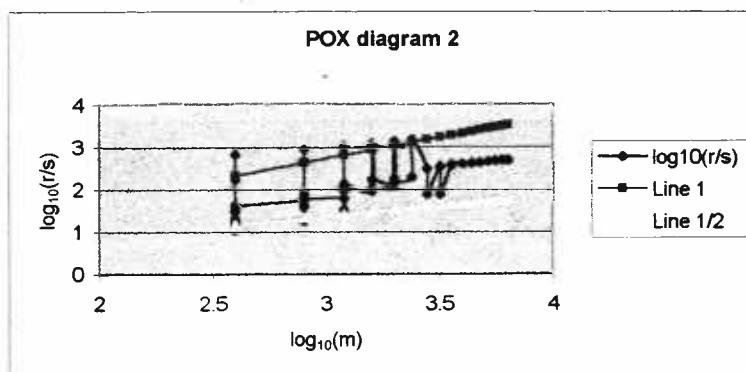
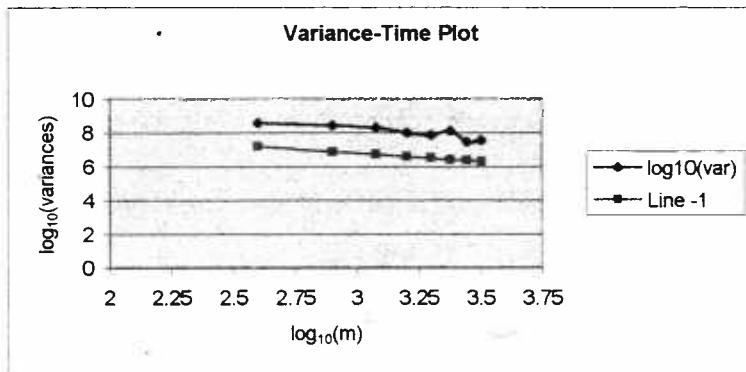
- Η διαφορά στις εκτιμήσεις είναι σχετικά μικρή.
- Το γεγονός της ύπαρξης διαφοράς εκτιμήσεων δεν δημιουργεί προβλήματα «συγκρούσεων»: Όταν η εκτίμηση μιας συγκεκριμένης μεθόδου υποδεικνύει μεγάλο (ή μικρό) βαθμό αυτό-ομοιότητας για μια χρονολογική σειρά πακέτων, τότε αναμένεται μεγάλος (ή μικρός) βαθμός αυτό-ομοιότητας και για την αντίστοιχη χρονολογική σειρά η οποία είναι μετρημένη σε ψηφιοσυλλαβές.

## Γ. ΕΚΤΙΜΗΣΗ ΤΟΥ ΒΑΘΜΟΥ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ ΤΗΣ ΚΙΝΗΣΗΣ ΤΟΥ ΔΙΚΤΥΟΥ

Από τον Πίνακα 3.1 προκύπτει ότι οι εκτιμήσεις του βαθμού αυτό-ομοιότητας των χρονολογικών σειρών των πακέτων και των ψηφιοσυλλαβών που εμφανίστηκαν στο δίκτυο κατά τη διάρκεια των μετρήσεων, είναι αρκετά υψηλές (αρκετά μεγαλύτερες από 0.5). Το γεγονός αυτό ισχύει και για τις δύο μεθόδους εκτίμησης και δίνει μια ένδειξη για την ύπαρξη αυτό-ομοιότητας, όπως προκύπτει και από τα παρακάτω διαγράμματα.

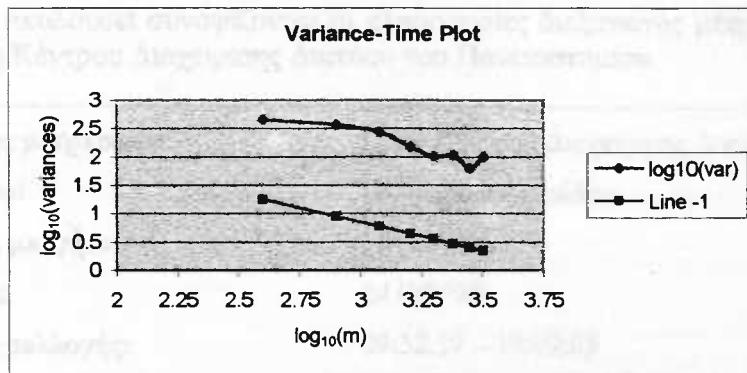
### Γ.1 ΧΡΟΝΟΛΟΓΙΚΗ ΣΕΙΡΑ ΨΗΦΙΟΣΥΛΛΑΒΩΝ

Η ύπαρξη αρκετά μεγάλου βαθμού αυτό-ομοιότητας για τη χρονολογική σειρά *EtherStatsOctets* επιβεβαιώνεται από τα παρακάτω διαγράμματα. Στο Variance-Time διάγραμμα οι λογαριθμημένες διακυμάνσεις έχουν αρκετή απόσταση από την γραμμή με κλίση  $-1$ . Στο ROX διάγραμμα τύπου 2 το μεγαλύτερο μέρος των λογαριθμημένων  $r/s$  τιμών βρίσκεται μεταξύ των γραμμών με κλίση  $1$  και  $\frac{1}{2}$ .

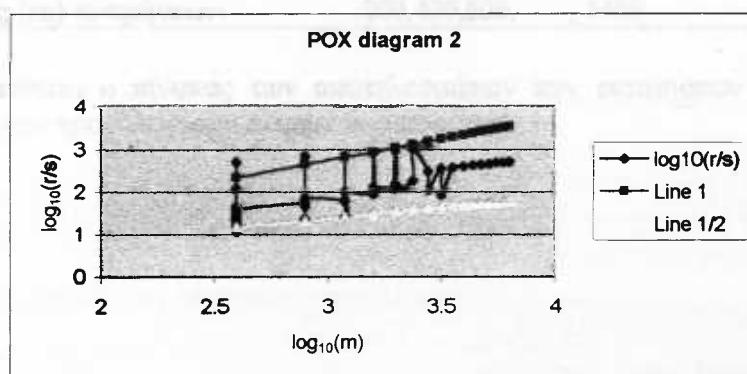


## Γ.2 ΧΡΟΝΟΛΟΓΙΚΗ ΣΕΙΡΑ ΠΑΚΕΤΩΝ

Παρακάτω στο Variance-Time διάγραμμα της χρονολογικής σειράς *EtherStatsPackets* φαίνεται ότι οι λογαριθμημένες διακυμάνσεις των αθροιστικών χρονολογικών σειρών



βρίσκονται σε αρκετά μεγάλη απόσταση πάνω από τη γραμμή με κλίση -1, γεγονός που φανερώνει την ύπαρξη υψηλού βαθμού αυτό-ομοιότητας της χρονολογικής σειράς. Στο POX διάγραμμα τύπου 2 το μεγαλύτερο μέρος των λογαριθμημένων r/s τιμών βρίσκεται μεταξύ των γραμμών με κλίση 1<sup>st</sup> και 1/2, γεγονός που αποδεικνύει την ύπαρξη αυτό-ομοιότητας.



#### 4. ΔΕΔΟΜΕΝΑ ΚΙΝΗΣΗΣ ΑΠΟ ΤΟ ΤΟΠΙΚΟ ΔΙΚΤΥΟ ΤΟΥ ΚΕΝΤΡΟΥ ΔΙΑΧΕΙΡΙΣΗΣ ΔΙΚΤΥΟΥ (NOC)

Στον πίνακα που ακολουθεί συνοψίζονται οι πληροφορίες διεξαγωγής μέτρησης από το τοπικό δίκτυο του Κέντρου Διαχείρισης Δικτύου του Πανεπιστημίου.

Δίκτυο διεξαγωγής μετρήσεων:	LAN Κέντρου Διαχείρισης Δικτύων
Τοποθεσία δικτύου:	Πτέρυγα Αντωνιάδου
Τρόπος συλλογής μετρήσεων:	HP LANprobe
Ημ/νία συλλογής:	24/09/98
Χρονική διάρκεια συλλογής:	09:32:59 – 19:09:03
Ρυθμός συλλογής (polling ανά):	10 sec
Πλήθος παρατηρήσεων Χ.Σ.:	3543
RMON MIBv1 αντικείμενα:	matrixSDOctets, matrixSDPkts
Αριθμός Χ.Σ. που συλλέχθηκαν:	28
Αριθμός Χ.Σ. που εκτιμήθηκαν ως προς H:	$10^1$
Επίπεδα άθροισης (m) εκτιμήσεων:	200,400,600, . . . , 3400

Παρακάτω παρατίθεται ο πίνακας των αποτελεσμάτων των εκτιμήσεων του βαθμού αυτό-ομοιότητας των χρονολογικών σειρών που προέκυψαν.

Παρακάτω παρατίθεται ο πίνακας των αποτελεσμάτων των εκτιμήσεων του βαθμού αυτό-ομοιότητας των χρονολογικών σειρών που προέκυψαν.

- Η διαφορά στην ακτινίδια των προσοτήνων
- Οι παραπόμπες μεταξύ αυτών μετρητών δεν εμπλέκονται στην ρυθμίση της μεγαλύτερης ταύτισης της γέννησης μεταβολής περιόδου. Με άλλα λόγια, δεν προσέβαλλε με κάποια αναστολή στην μεταβολή της περιόδου η αύξηση της προσδικητικής από την μετέδοτη περιόδου της H.
- Το γεγονός της αυτέρης διαφορής περιβάλλεται στην προστασία από την αντιστοίχη περιόδου. Εάν τη περιόδου της αύλης μεταβολής να διαλέγεται από την περιόδο παραμετρίους στη συνέχεια, τότε αυτή η περιόδος προστασίας θα γίνεται περιορισμένη, καθώς τα περιόδους παραμετρίους καταγράφονται σε απόσταση από την περιόδο παραμετρίου που έχει οριστεί. Εγκαρδικό είναι η περιόδος παραμετρίου που έχει οριστεί, καθώς η περιόδος παραμετρίου που έχει οριστεί θα προστασίει την διάφορη περιόδο παραμετρίου που έχει οριστεί, καθώς η περιόδος παραμετρίου που έχει οριστεί θα προστασίει την διάφορη περιόδο παραμετρίου που έχει οριστεί.

<sup>1</sup> Κάθε ένα από τα 28 άρχεια χρονολογικών σειρών που δημιουργήθηκαν αφορούν διαφορετικό ζεύγος πηγής-προορισμού. Εκτελώντας τις κατάλληλες αθροίσεις προκύπτουν εύκολα τα συνολικά στοιχεία κίνησης που αφορούν όλους τους προορισμούς κάθε πηγής. Έτσι προκύπτουν οι 10 χρονολογικές σειρές, μια για κάθε διαφορετική πηγή.

**ΑΠΟΤΕΛΕΣΜΑΤΑ ΕΚΤΙΜΗΣΕΩΝ ΤΗΣ  $H$**

HOST	ΜΕΘΟΔΟΣ ΕΚΤΙΜΗΣΗΣ $H$	HostOutOctets	
		1. ΑΘΡΟΙΣΤΙΚΩΝ ΧΡ/ΚΩΝ ΣΕΙΡΩΝ	2. R/S ΣΤΑΤΙΣΤΙΚΗΣ
1.	Host I	0.578	0.516
2.	Power Hub Default Gateway	0.025	0.38
3.	Cisco Δρομολογητής σύνδεσης με το Internet	0.29	0.3
4.	HP Σταθμός Διαχείρισης Δικτύου	0.465	0.513
5.	PC	0.63	0.768
6.	PC	0.58	0.68
7.	PC	0.19	0.45
8.	PC	0.72	0.7
9.	PC	0.57	0.7
10.	PC	0.83	0.67

**ΣΧΟΛΙΑΣΜΟΣ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΕΚΤΙΜΗΣΗΣ ΤΗΣ  $H$**

**Α. ΔΙΑΦΟΡΑ ΣΤΙΣ ΕΚΤΙΜΗΣΕΙΣ ΤΗΣ  $H$  ΤΩΝ ΔΥΟ ΜΕΘΟΔΩΝ ΕΚΤΙΜΗΣΗΣ**

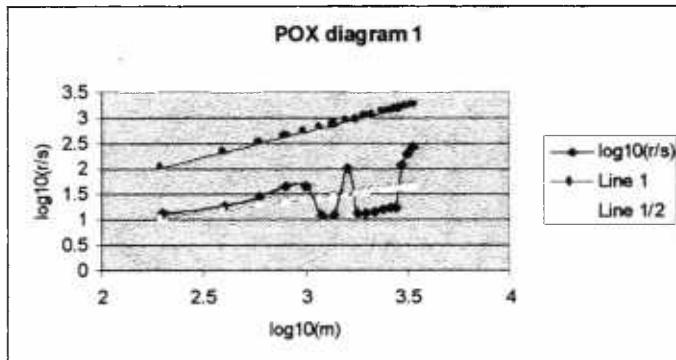
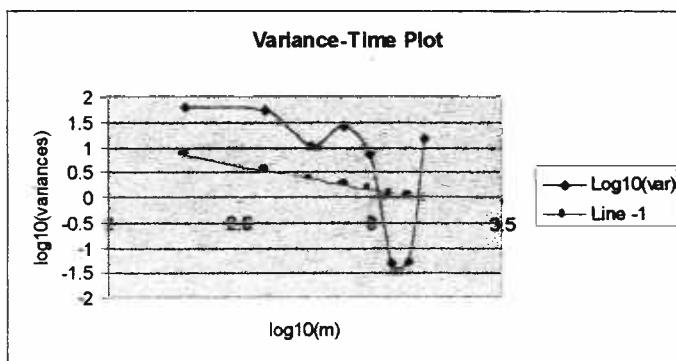
Συγκρίνοντας τις στήλες (1,2) του Πίνακα 1 οι οποίες αφορούν τις δύο μεθόδους εκτίμησης της  $H$ , προκύπτουν τα εξής:

- Η διαφορά στις εκτιμήσεις είναι σχετικά μικρή.
- Οι εκτιμήσεις μιας μεθόδου εκτίμησης δεν εμφανίζονται εν γένει μικρότερες (ή μεγαλύτερες) από αυτές της άλλης μεθόδου εκτίμησης. Με άλλα λόγια, δεν προκύπτει μια πάγια ανισοτική σχέση μεταξύ των εκτιμήσεων η οποία να προσδιορίζεται από την μέθοδο εκτίμησης της  $H$ .
- Το γεγονός της ύπαρξης διαφοράς εκτιμήσεων δεν δημιουργεί προβλήματα «συγκρούσεων». Όταν η εκτίμηση της μίας μεθόδου υποδεικνύει μεγάλο (ή μικρό) βαθμό αυτό-ομοιότητας, τότε αναμένεται κατ' αντιστοιχία μεγάλος (ή μικρός) βαθμός αυτό-ομοιότητας και για την εκτίμηση της άλλης μεθόδου. Για παράδειγμα μπορεί η διαφορά των δύο εκτιμήσεων της γραμμής 2 του Πίνακα 1, να «φαίνεται» μεγάλη, όμως και οι δύο τιμές των εκτιμήσεων φανερώνουν έλλειψη αυτό-ομοιότητας.

## B. ΕΚΤΙΜΗΣΗ ΤΟΥ ΒΑΘΜΟΥ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ

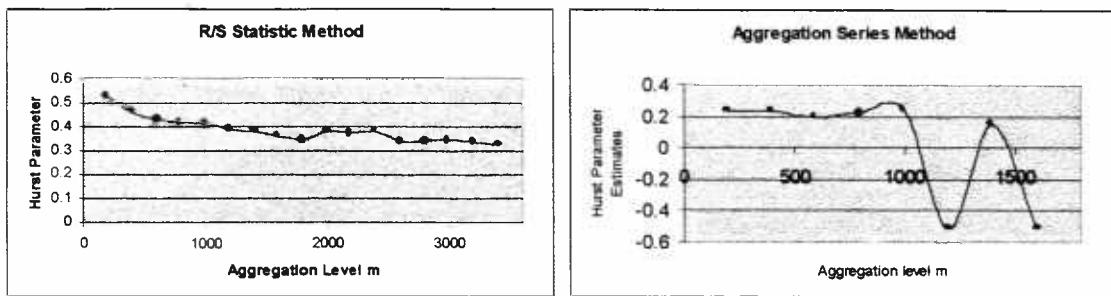
Από τον Πίνακα 1 προκύπτει ότι για μερικούς host υπάρχει ένδειξη αυτό-ομοιότητας στις χρονολογικές σειρές των ψηφιοσυλλαβών εξόδου. Συγκεκριμένα, οι εκτιμήσεις της παραμέτρου  $H$  για τις χρονολογικές σειρές ψηφιοσυλλαβών εξόδου των host υπ' αριθμόν 2,3,4 και 7 δεν ξεπερνούν την τιμή 0.5 και συνεπώς δεν χαρακτηρίζονται από την ύπαρξη αυτό-ομοιότητας. Για τις υπόλοιπες χρονολογικές σειρές υπάρχει ένδειξη ύπαρξης αυτό-ομοιότητας. Παρακάτω παρατίθενται για κάθε host τα διαγράμματα που προκύπτουν από τις δύο μεθόδους εκτίμησης του βαθμού της αυτό-ομοιότητας (παράμετρος  $H$ ).

### B.1 ΧΡΟΝΟΛΟΓΙΚΗ ΣΕΙΡΑ ΨΗΦΙΟΣΥΛΛΑΒΩΝ ΕΞΟΔΟΥ ΓΙΑ ΤΟΝ HOST 1



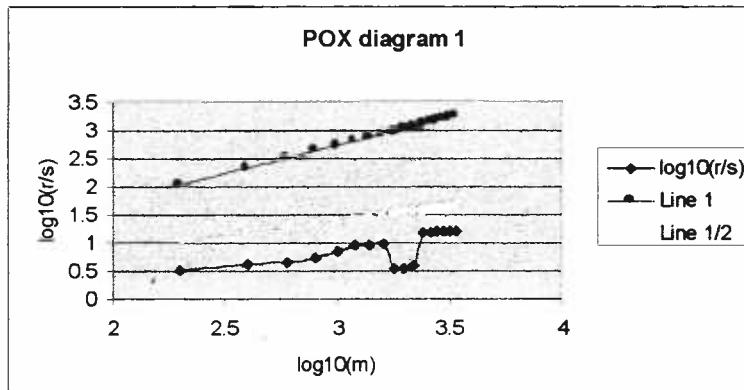
Ο σχετικά χαμηλός βαθμός αυτό-ομοιότητας επιβεβαιώνεται από το Variance-Time διάγραμμα από το γεγονός ότι για ορισμένα επίπεδα άθροισης οι λογαριθμημένες διακυμάνσεις βρίσκονται αρκετά πιο κάτω από τη γραμμή με κλίση  $-1$ . Επίσης, από το POX διάγραμμα τύπου 1 φαίνεται ότι για αρκετά επίπεδα άθροισης, οι λογαριθμημένες  $r/s$  τιμές βρίσκονται κάτω από την γραμμή με κλίση  $\frac{1}{2}$ .

## B.2 ΧΡΟΝΟΛΟΓΙΚΗ ΣΕΙΡΑ ΨΗΦΙΟΣΥΛΛΑΒΩΝ ΕΞΟΔΟΥ ΓΙΑ ΤΟΝ HOST2



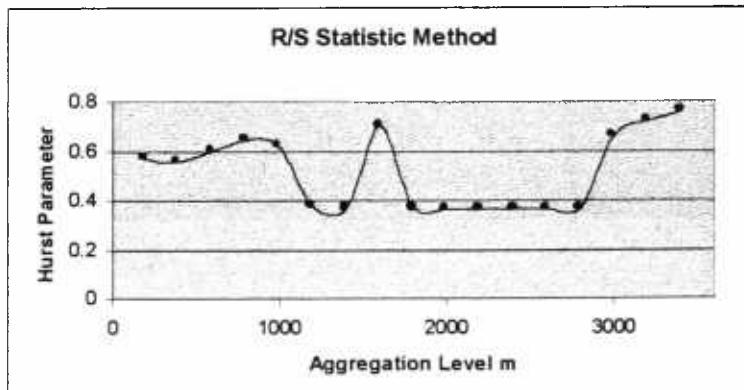
Στα παραπάνω διαγράμματα απεικονίζονται οι εκτιμήσεις της παραμέτρου  $H$  για διάφορα επίπεδα άθροισης  $m$ . Επομένως, παρατηρείται ότι και στις δύο μεθόδους εκτιμησης για κανένα επίπεδο άθροισης δεν ισχύει  $H > 0.5$ . Το γεγονός αποδεικνύει την ανυπαρξία αυτό-ομοιότητας των χρονολογικών σειρών.

## B.3 ΧΡΟΝΟΛΟΓΙΚΗ ΣΕΙΡΑ ΨΗΦΙΟΣΥΛΛΑΒΩΝ ΕΞΟΔΟΥ ΓΙΑ ΤΟΝ HOST3



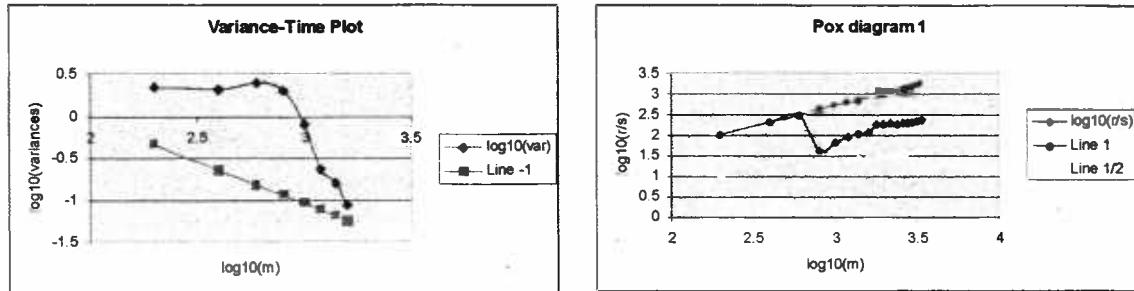
Από το παραπάνω POX διάγραμμα τύπου 1 προκύπτει ότι οι εκτιμήσεις της παραμέτρου  $H$  για όλα τα εξεταζόμενα επίπεδα άθροισης  $m$ , βρίσκονται κάτω από τη γραμμή με κλίση  $\frac{1}{2}$ , γεγονός που φανερώνει την έλλειψη αυτό-ομοιότητας.

## B.4 ΧΡΟΝΟΛΟΓΙΚΗ ΣΕΙΡΑ ΨΗΦΙΟΣΥΛΛΑΒΩΝ ΕΞΟΔΟΥ ΓΙΑ ΤΟΝ HOST4



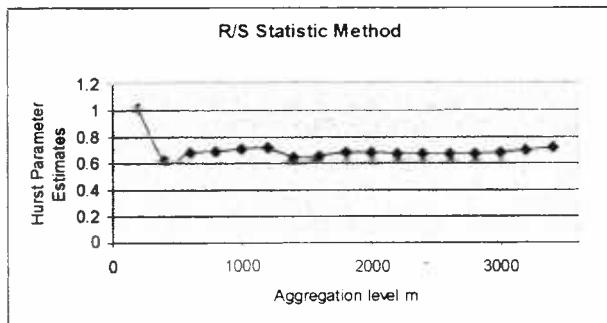
Από το παραπάνω διάγραμμα προκύπτει ότι οι εκτιμήσεις της παραμέτρου  $H$  για διάφορα επίπεδα άθροισης βρίσκονται στο διάστημα [0.38, 0.77], με μέσο όρο τιμών γύρω στο 0.5. Βάσει αυτών των στοιχείων προκύπτει ότι δεν μπορούμε να αποφανθούμε σχετικά με την ύπαρξη αυτό-ομοιότητας στα δεδομένα.

#### B.5 ΧΡΟΝΟΛΟΓΙΚΗ ΣΕΙΡΑ ΨΗΦΙΟΣΥΛΛΑΒΩΝ ΕΞΟΔΟΥ ΓΙΑ ΤΟΝ HOST5



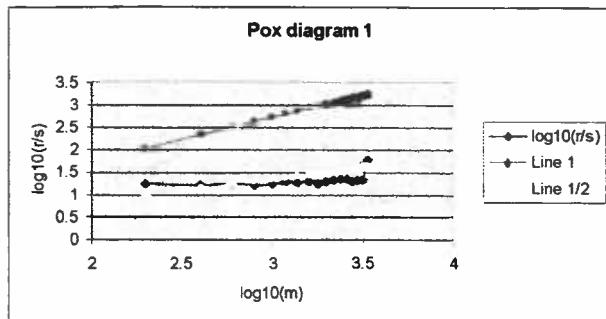
Από τα παραπάνω διαγράμματα προκύπτει η ύπαρξη αυτό-ομοιότητας. Στο Variance-Time διάγραμμα όλες οι τιμές των λογαριθμημένων διακυμάνσεων των αθροιστικών χρονολογικών σειρών βρίσκονται πάνω από τη γραμμή με κλίση -1. Στο Pox διάγραμμα τύπου 1, οι λογαριθμημένες τιμές των r/s βρίσκονται μεταξύ των γραμμών με κλίση 1 και  $\frac{1}{2}$ .

#### B.6 ΧΡΟΝΟΛΟΓΙΚΗ ΣΕΙΡΑ ΨΗΦΙΟΣΥΛΛΑΒΩΝ ΕΞΟΔΟΥ ΓΙΑ ΤΟΝ HOST6



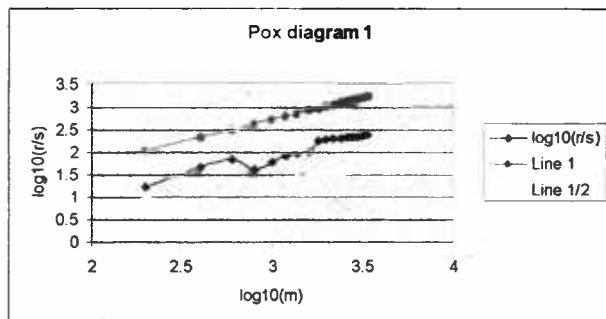
Από το παραπάνω διάγραμμα φαίνεται ότι καθώς αυξάνουν τα επίπεδα άθροισης, οι αντίστοιχες εκτιμήσεις της παραμέτρου  $H$  σταθεροποιούνται λίγο πάνω από την τιμή 0.6. Γεγονός το οποίο δίνει μια ένδειξη για την ύπαρξη αυτό-ομοιότητας.

### B.7 ΧΡΟΝΟΛΟΓΙΚΗ ΣΕΙΡΑ ΨΗΦΙΟΣΥΛΛΑΒΩΝ ΕΞΟΔΟΥ ΓΙΑ ΤΟΝ HOST7



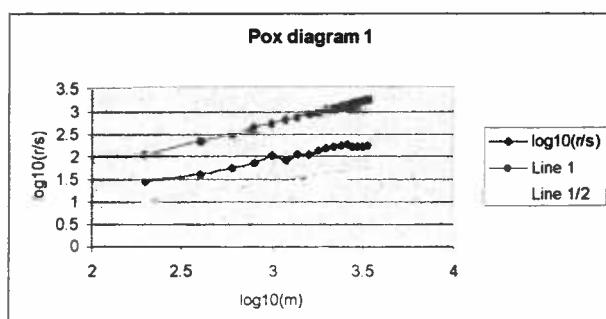
Το παραπάνω POX διάγραμμα τύπου 1 δείχνει την έλλειψη αυτό-ομοιότητας, αφού για τα περισσότερα επίπεδα άθροισης οι λογαριθμημένες τιμές των r/s βρίσκονται κάτω από την γραμμή με κλίση  $\frac{1}{2}$ .

### B.8 ΧΡΟΝΟΛΟΓΙΚΗ ΣΕΙΡΑ ΨΗΦΙΟΣΥΛΛΑΒΩΝ ΕΞΟΔΟΥ ΓΙΑ ΤΟΝ HOST8



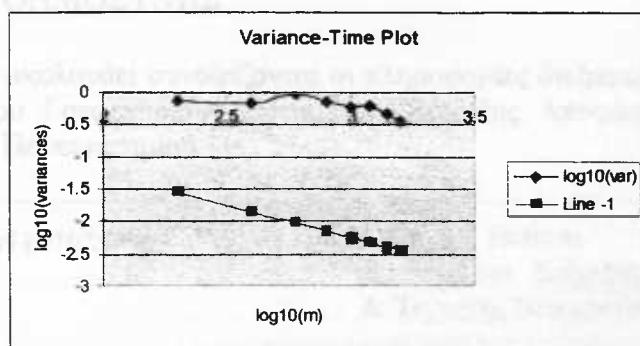
Από το παραπάνω POX διάγραμμα φαίνεται ότι οι περισσότερες λογαριθμημένες τιμές των r/s βρίσκονται μεταξύ των γραμμών με κλίση 1 και  $\frac{1}{2}$  και σε ίση απόσταση από αυτές, γεγονός που δικαιολογεί μια εκτίμηση της  $H$ , ίση με 0.7.

### B.9 ΧΡΟΝΟΛΟΓΙΚΗ ΣΕΙΡΑ ΨΗΦΙΟΣΥΛΛΑΒΩΝ ΕΞΟΔΟΥ ΓΙΑ ΤΟΝ HOST9



Από το παραπάνω POX διάγραμμα φαίνεται ότι η γραμμή που σχηματίζουν οι λογαριθμημένες τιμές των r/s, βρίσκεται μεταξύ των γραμμών με κλίση 1 και  $\frac{1}{2}$ , ενώ είναι σχεδόν παράλληλη με αυτές.

### B.10 ΧΡΟΝΟΛΟΓΙΚΗ ΣΕΙΡΑ ΨΗΦΙΟΣΥΛΛΑΒΩΝ ΕΞΟΔΟΥ ΓΙΑ ΤΟΝ HOST10



Η μεγάλη εκτίμηση της παραμέτρου της  $H$  σύμφωνα με τη μέθοδο των αθροιστικών χρονολογικών σειρών, αποδεικνύεται και από το Variance-Time διάγραμμα στο οποίο οι τιμές των διακυμάνσεων έχουν μεγάλη απόσταση από την γραμμή με κλίση  $-1$ . Ο υψηλός βαθμός αυτό-ομοιότητας εξηγείται αφού στο μεγαλύτερο μέρος τους τα δεδομένα έχουν μηδενικές τιμές.

Χρονολογική Σειρά	Host		Non-Host	
	Αυτό-Ομοιότητα	Χρήση Διφρύ	Αυτό-Ομοιότητα	Χρήση Διφρύ
1. Host	0.75	0.38	0.49	0.39
2. Host	0.7	0.35	0.48	0.38
3. Host	0.7	0.35	0.48	0.38

\* Από τις πρώτες γραπτές σημειώσεις των αναπορηθείσας απειλήσεων που δέχθηκαν οι δύο παραπομπές, από την παραδοσιακή ιστορία της Ελλάδας, η οποία αποδεικνύεται ότι η απειλή προέρχεται από την Κύπρο.

## 5. ΔΕΔΟΜΕΝΑ ΚΙΝΗΣΗΣ ΑΠΟ ΤΟ ΤΟΠΙΚΟ ΔΙΚΤΥΟ HELRUN ΤΟΥ ΕΡΓΑΣΤΗΡΙΟΥ ΣΥΣΤΗΜΑΤΩΝ ΣΤΗΡΙΞΗΣ ΑΠΟΦΑΣΕΩΝ & ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ

Στον πίνακα που ακολουθεί συνοψίζονται οι πληροφορίες διεξαγωγής μέτρησης στο τοπικό δίκτυο του Εργαστηρίου Συστημάτων Στήριξης Αποφάσεων & Τεχνητής Νοημοσύνης του Πανεπιστημίου.

<b>Δίκτυο διεξαγωγής μετρήσεων:</b>	LAN Helrun Εργαστηρίου Συστημάτων Στήριξης Αποφάσεων & Τεχνητής Νοημοσύνης
<b>Τοποθεσία δικτύου</b>	Πτέρυγα Αντωνιάδου
<b>Τρόπος συλλογής μετρήσεων:</b>	HP LANprobe
<b>Ημ/νία συλλογής:</b>	19/10/98 – 20/10/98
<b>Χρονική διάρκεια συλλογής:</b>	11:31:24 – 17:51:07
<b>Ρυθμός συλλογής (Polling ανά):</b>	10 sec
<b>Πλήθος παρατηρήσεων Χ.Σ.:</b>	10874
<b>RMON MIB v1 αντικείμενα:</b>	HostInOctets, hostOutOctets, hostInPkts, hostOutPkts, matrixSDOctets, matrixSDPkts
<b>Αριθμός Χ.Σ. που εκτιμήθηκαν ως προς Η:</b>	12 <sup>1</sup>
<b>Επίπεδα άθροισης (m) εκτιμήσεων:</b>	400, 800, 1200, . . . , 10400

Παρακάτω παραθέτονται οι πίνακες των αποτελεσμάτων των εκτιμήσεων του βαθμού αυτό-ομοιότητας των χρονολογικών σειρών.

ΜΕΘΟΔΟΣ ΕΚΤΙΜΗΣΗΣ Η	<i>HostOutOctets</i>		<i>HostOutPackets</i>	
	1. <i>ΑΘΡΟΙΣΤΙΚΩΝ ΧΡ/ΚΩΝ ΣΕΙΡΩΝ</i>	2. <i>R/S ΣΤΑΤΙ- ΣΤΙΚΗΣ</i>	3. <i>ΑΘΡΟΙΣΤΙΚΩΝ ΧΡ/ΚΩΝ ΣΕΙΡΩΝ</i>	4. <i>R/S ΣΤΑΤΙ- ΣΤΙΚΗΣ</i>
HOST				
1. Host1	0.75	0.652	0.59	0.538
2. Host2	0.7	0.73	0.67	0.74
3. Host3	-	-	-	-

Πίνακας 1 Οι εκτιμήσεις της Η των χρονολογικών σειρών πακέτων και ψηφιοσυλλαβών εξόδου των 2 host.

<sup>1</sup> Από τα αρχεία χρονολογικών σειρών που δημιουργήθηκαν επιλέχθηκαν για εκτίμηση του βαθμού αυτό-ομοιότητας αυτά που αφορούσαν τις πιο ενεργές (active) συσκευές του δικτύου κατά τη διάρκεια των μετρήσεων.

ΜΕΘΟΔΟΣ ΕΚΤΙΜΗΣΗΣ Η	<u>HostInOctets</u>		<u>HostInPackets</u>	
	1. <b>ΑΘΡΟΙΣΤΙΚΩΝ ΧΡ/ΚΩΝ ΣΕΙΡΩΝ</b>	2. <b>R/S ΣΤΑΤΙ- ΣΤΙΚΗΣ</b>	3. <b>ΑΘΡΟΙΣΤΙΚΩΝ ΧΡ/ΚΩΝ ΣΕΙΡΩΝ</b>	4. <b>R/S ΣΤΑΤΙ- ΣΤΙΚΗΣ</b>
<b>HOST</b>				
1. Host1	0.53	0.82	0.708	0.855
2. Host2	0.65	0.57	0.582	0.582
3. Host3	0.65	0.666	0.779	0.78

Πίνακας 2 Οι εκτιμήσεις της Η των χρονολογικών σειρών πακέτων και ψηφιοσυλλαβών εισόδου των 3 host.

### **ΣΧΟΛΙΑΣΜΟΣ ΑΠΟΤΕΛΕΣΜΑΤΩΝ**

#### **A. ΔΙΑΦΟΡΑ ΣΤΙΣ ΕΚΤΙΜΗΣΕΙΣ ΤΗΣ Η ΤΩΝ ΔΥΟ ΜΕΘΟΔΩΝ ΕΚΤΙΜΗΣΗΣ**

Συγκρίνοντας τις στήλες των Πινάκων 1 και 2 που αφορούν τις δύο μεθόδους εκτιμήσης της Η (δηλαδή τις (1,2) και (3,4)), προκύπτουν τα εξής:

- Η διαφορά στις εκτιμήσεις είναι σχετικά μικρή.
- Οι εκτιμήσεις μιας μεθόδου εκτιμήσης δεν εμφανίζονται εν γένει μικρότερες (ή μεγαλύτερες) από αυτές της άλλης μεθόδου εκτιμήσης. Με άλλα λόγια, δεν προκύπτει μια πάγια ανισοτική σχέση μεταξύ των εκτιμήσεων η οποία να προσδιορίζεται από την μέθοδο εκτιμήσης της Η.
- Το γεγονός της ύπαρξης διαφοράς εκτιμήσεων δεν δημιουργεί προβλήματα «συγκρούσεων». Όταν η εκτίμηση της μίας μεθόδου υποδεικνύει μεγάλο (ή μικρό) βαθμό αυτό-ομοιότητας, τότε αναμένεται κατ' αντιστοιχία μεγάλος (ή μικρός) βαθμός αυτό-ομοιότητας και για την εκτίμηση της άλλης μεθόδου.

#### **B. ΔΙΑΦΟΡΑ ΣΤΙΣ ΕΚΤΙΜΗΣΕΙΣ ΤΗΣ Η ΓΙΑ ΚΙΝΗΣΗ ΕΙΣΟΔΟΥ & ΕΞΟΔΟΥ**

Συγκρίνοντας για κάθε host τις εκτιμήσεις της Η για την κίνηση εισόδου και εξόδου (δηλαδή τις δύο πρώτες γραμμές των Πινάκων 1 και 2), προκύπτει ότι:

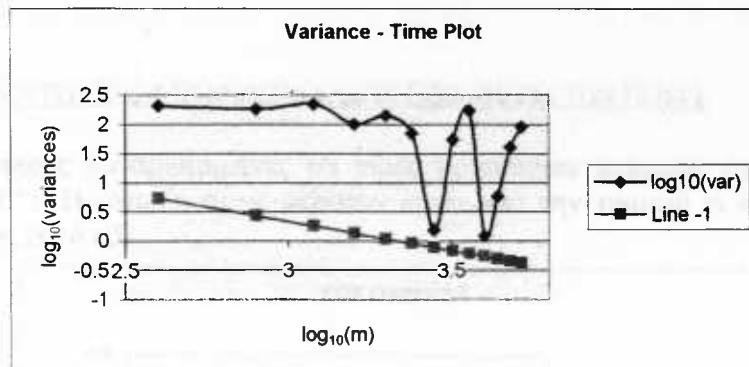
Οι εκτιμήσεις του βαθμού αυτό-ομοιότητας της κίνησης εξόδου για έναν host είναι ανάλογες των εκτιμήσεων της κίνησης εισόδου για τον ίδιο host. Η παρατήρηση ισχύει και για τις δύο μεθόδους εκτιμήσης της Η, και για τους δύο τρόπους έκφρασης της κίνησης (πακέτα ή ψηφιοσυλλαβές).

## Γ. ΕΚΤΙΜΗΣΗ ΤΟΥ ΒΑΘΜΟΥ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ ΤΗΣ ΚΙΝΗΣΗΣ ΤΟΥ ΔΙΚΤΥΟΥ

Από τους Πίνακες 1 και 2 προκύπτει ότι οι εκτιμήσεις του βαθμού αυτό-ομοιότητας των χρονολογικών σειρών των πακέτων και των ψηφιοσυλλαβών των τριών host είναι αρκετά μεγαλύτερες από 0.5. Το γεγονός αυτό ισχύει και για τις δύο μεθόδους εκτίμησης και δίνει μια ένδειξη για την ύπαρξη αυτό-ομοιότητας, όπως προκύπτει και από τα παρακάτω διαγράμματα.

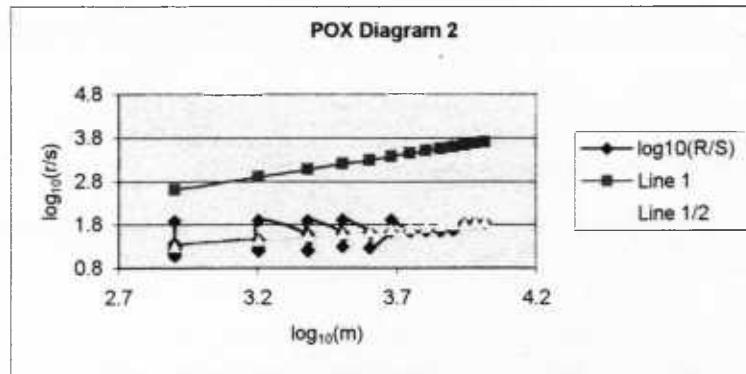
### Γ.1 ΧΡΟΝΟΛΟΓΙΚΗ ΣΕΙΡΑ ΨΗΦΙΟΣΥΛΛΑΒΩΝ ΕΙΣΟΔΟΥ ΓΙΑ ΤΟΝ HOST1

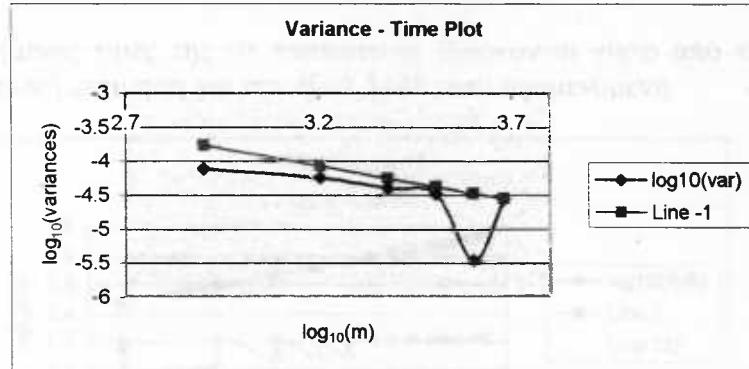
Οι διακυμάνσεις στο Variance-Time διάγραμμα βρίσκονται αρκετά πάνω από την γραμμή με κλίση -1, στην οποία αντιστοιχεί  $H=0.5$ . Έτσι το σχήμα δικαιολογεί μια τιμή  $H=0.75$ .



### Γ.2 ΧΡΟΝΟΛΟΓΙΚΗ ΣΕΙΡΑ ΠΑΚΕΤΩΝ ΕΙΣΟΔΟΥ ΓΙΑ ΤΟΝ HOST1

Οι λογαριθμημένες τιμές της r/s στατιστικές βρίσκονται γύρω από τη γραμμή με κλίση  $\frac{1}{2}$ , με τις μισές από αυτές να βρίσκονται πάνω και τις άλλες μισές να βρίσκονται κάτω από τη γραμμή στην οποία αντιστοιχεί τιμή για την  $H=0.5$ . Η τελική εκτίμηση για την  $H=0.4945$  είναι αναμενόμενη.

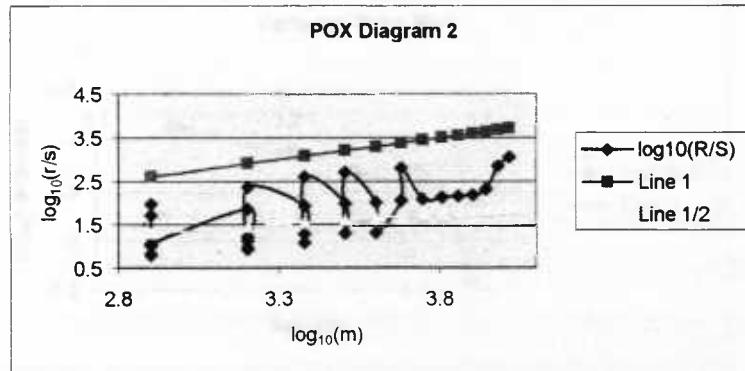




Η χαμηλή εκτίμηση της  $H$  επιβεβαιώνεται και από το Variance-Time διάγραμμα, αφού οι λογαριθμημένες διακυμάνσεις βρίσκονται σχεδόν πάνω στην γραμμή με κλίση  $-1$  ή και κάτω από αυτή.

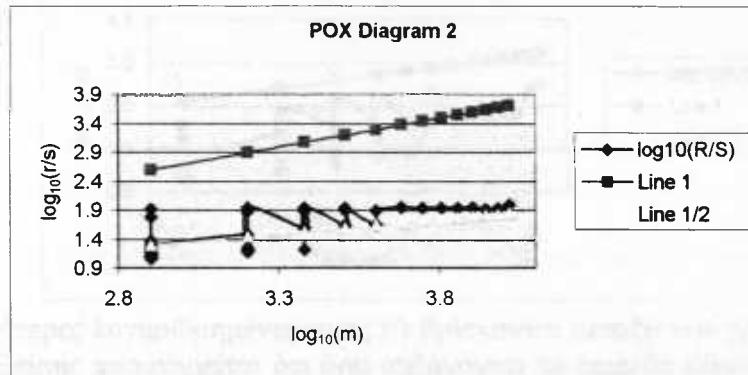
### Γ.3 ΧΡΟΝΟΛΟΓΙΚΗ ΣΕΙΡΑ ΨΗΦΙΟΣΥΛΛΑΒΩΝ ΕΞΟΔΟΥ ΓΙΑ ΤΟΝ HOST1

Οι περισσότερες λογαριθμημένες  $r/s$  τιμές βρίσκονται ανάμεσα στις γραμμές με κλίση 1 και  $\frac{1}{2}$ . Η ύπαρξη όμως μερικών κάτω από την γραμμή  $\frac{1}{2}$  δικαιολογεί την εκτίμηση της  $H=0.65$ .

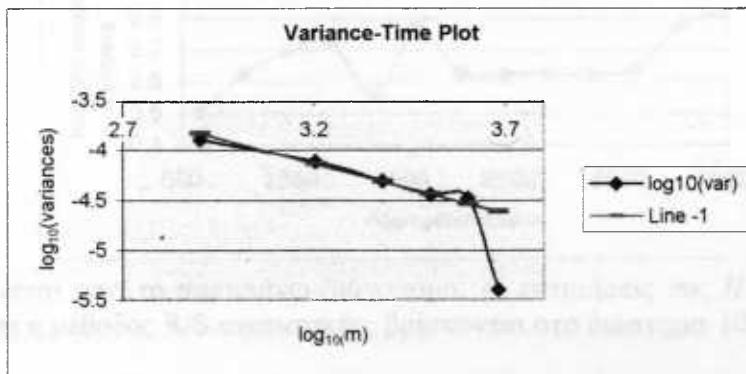


#### Γ.4 ΧΡΟΝΟΛΟΓΙΚΗ ΣΕΙΡΑ ΠΑΚΕΤΩΝ ΕΞΟΔΟΥ ΓΙΑ ΤΟΝ HOST1

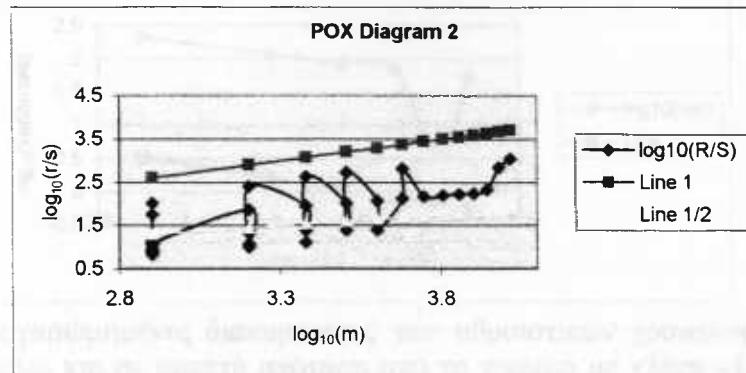
Οι λογαριθμημένες τιμές της r/s στατιστικής βρίσκονται γύρω από τη γραμμή με κλίση  $\frac{1}{2}$ . Η τελική εκτίμηση για την  $H=0.5498$  είναι αναμενόμενη.



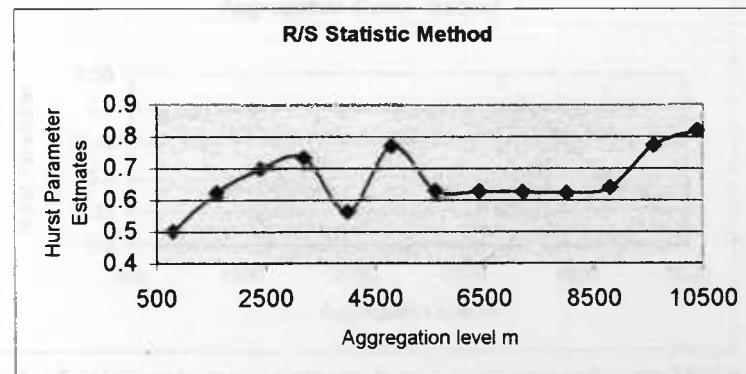
Οι περισσότερες τιμές των λογαριθμημένων διακυμάνσεων συμπίπτουν με τη γραμμή με κλίση  $-1$ , ενώ η τιμή της λογαριθμημένης διακύμανσης που αντιστοιχεί στο τελευταίο επίπεδο άθροισης  $m=4800$ , βρίσκεται κάτω από αυτή. Η εικόνα αυτή δικαιολογεί μια εκτίμηση της  $H$  κάτω από 0.5.



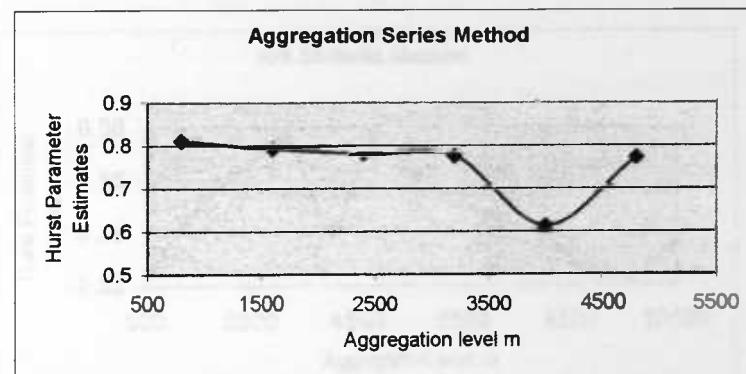
Γ.5 ΧΡΟΝΟΛΟΓΙΚΗ ΣΕΙΡΑ ΨΗΦΙΟΣΥΛΛΑΒΩΝ ΕΙΣΟΔΟΥ ΓΙΑ ΤΟΝ HOST2



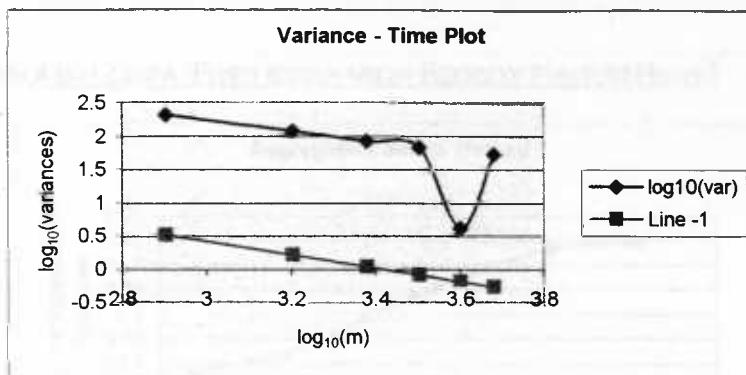
Οι περισσότερες λογαριθμημένες τιμές  $r/s$  βρίσκονται μεταξύ των γραμμών με κλίση 1 και  $1/2$ . Επίσης παρατηρείται ότι όσο αυξάνονται τα επίπεδα άθροισης, οι τιμές  $r/s$  «πλησιάζουν» τη γραμμή με κλίση 1. Με αυτόν τον τρόπο επιβεβαιώνεται η ύπαρξη αυτό-ομοιότητας στα δεδομένα.



Όπως φαίνεται από το παραπάνω διάγραμμα, οι εκτιμήσεις της  $H$  παραμέτρου τις οποίες δίνει η μέθοδος R/S στατιστικής, βρίσκονται στο διάστημα [0.5,0.8].

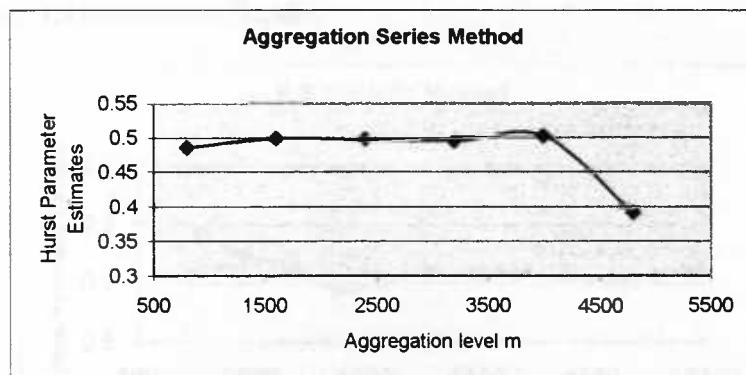


Οι εκτιμήσεις της  $H$  με τη μέθοδο των αθροιστικών χρονολογικών σειρών, βρίσκονται στο διάστημα [0.6,0.8].

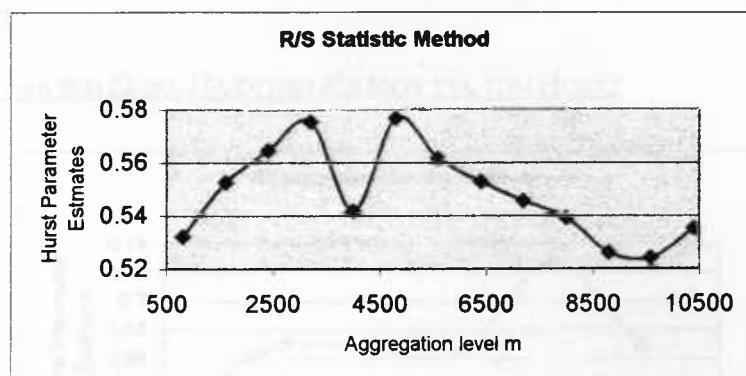


Επίσης, οι λογαριθμημένες διακυμάνσεις των αθροιστικών χρονολογικών σειρών, βρίσκονται πάνω και σε αρκετή απόταση από τη γραμμή με κλίση -1, γεγονός που αποτελεί ένδειξη αυτό-ομοιότητας.

#### Γ.6 ΧΡΟΝΟΛΟΓΙΚΗ ΣΕΙΡΑ ΠΑΚΕΤΩΝ ΕΙΣΟΔΟΥ ΓΙΑ ΤΟΝ HOST2

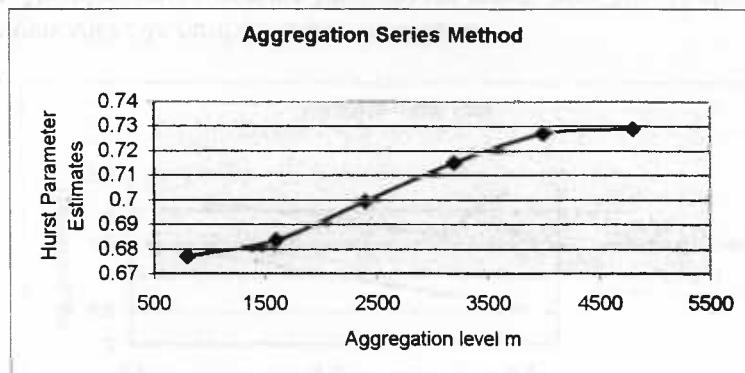


Για όλα τα επίπεδα άθροισης  $m$ , οι εκτιμήσεις της παραμέτρου Hurst με τη μέθοδο των αθροιστικών χρονολογικών σειρών, βρίσκονται κοντά ή κάτω από την τιμή 0.5. Γεγονός που δεν αποτελεί ένδειξη ύπαρξης αυτό-ομοιότητας.

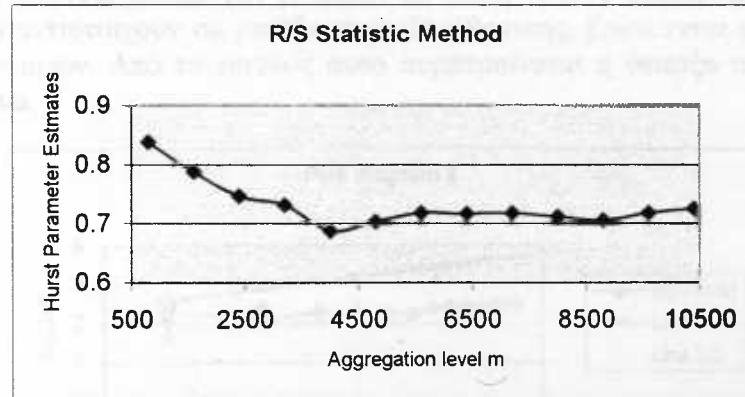


Για όλα τα επίπεδα άθροισης  $m$ , οι εκτιμήσεις της παραμέτρου Hurst με τη μέθοδο της r/s στατιστικής, βρίσκονται ανάμεσα στο διάστημα [0.52,0.58].

### Γ.7 ΧΡΟΝΟΛΟΓΙΚΗ ΣΕΙΡΑ ΨΗΦΙΟΣΥΛΛΑΒΩΝ ΕΞΟΔΟΥ ΓΙΑ ΤΟΝ HOST2

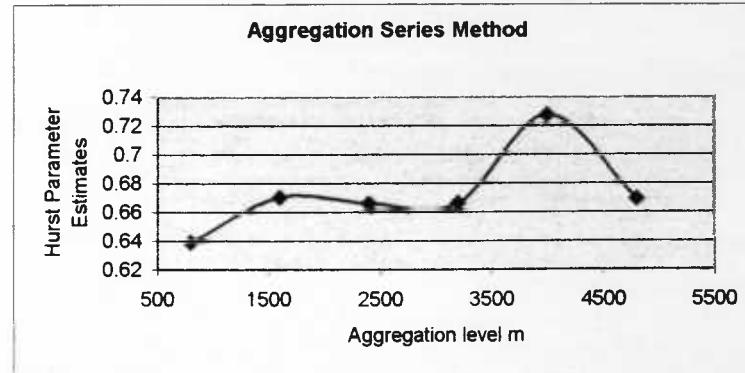


Οι εκτιμήσεις της παραμέτρου  $H$ , βρίσκονται στο διάστημα  $[0.675, 0.73]$ , ενώ καθώς αυξάνεται το επίπεδο άθροισης, αυξάνονται και οι τιμές των εκτιμήσεων  $H$ , γεγονός που δίνει μια ένδειξη για την ύπαρξη αυτό-ομοιότητας σύμφωνα με την μέθοδο των Αθροιστικών Χρονολογικών Σειρών.



Για όλα τα επίπεδα άθροισης οι εκτιμήσεις της παραμέτρου  $H$ , σταθερποιούνται γύρω από την τιμή 0.73, από όπου προκύπτει και η εκτίμηση της  $H$ , με την μέθοδο της R/S Στατιστικής.

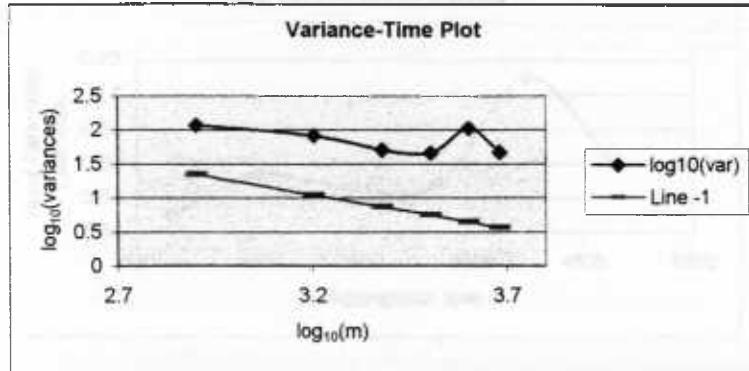
### Γ.8 ΧΡΟΝΟΛΟΓΙΚΗ ΣΕΙΡΑ ΠΑΚΕΤΩΝ ΕΞΟΔΟΥ ΓΙΑ ΤΟΝ HOST2



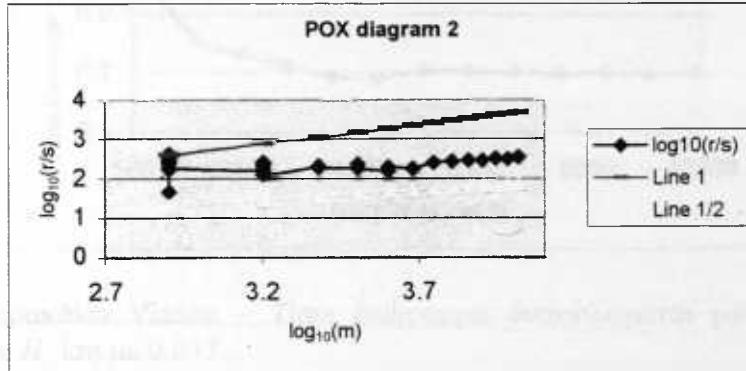
Από το παραπάνω διάγραμμα φαίνεται ότι οι εκτιμήσεις της παραμέτρου  $H$  για τα συγκεκριμένα επίπεδα άθροισης, βρίσκονται στο διάστημα  $[0.64, 0.73]$ .

### Γ.9 ΧΡΟΝΟΛΟΓΙΚΗ ΣΕΙΡΑ ΨΗΦΙΟΣΥΛΛΑΒΩΝ ΕΙΣΟΔΟΥ ΓΙΑ ΤΟΝ HOST3

Από το παρακάτω διάγραμμα φαίνεται ότι οι λογαριθμημένες διακυμάνσεις των αθροιστικών χρονολογικών σειρών βρίσκονται πάνω από την γραμμή με κλίση -1, γεγονός που δεικνύει την ύπαρξη αυτό-ομοιότητας.

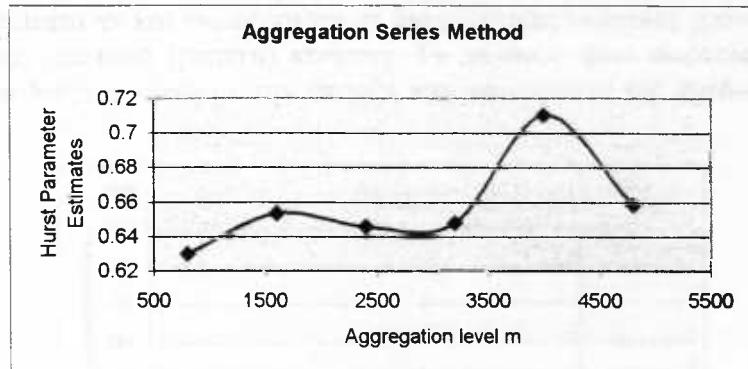


Επίσης, στο POX διάγραμμα τύπου 2, το μεγαλύτερο πλήθος των λογαριθμημένων τιμών r/s βρίσκονται μεταξύ των γραμμών με κλίση 1 και ½. Ειδικότερα, όλες οι τιμές των r/s που αντιστοιχούν σε μεγάλα επίπεδα αθροιστης, βρίσκονται μεταξύ των δύο οριακών γραμμών. Από το γεγονός αυτό συμπεραίνεται η ύπαρξη αυτό-ομοιότητας στα δεδομένα.

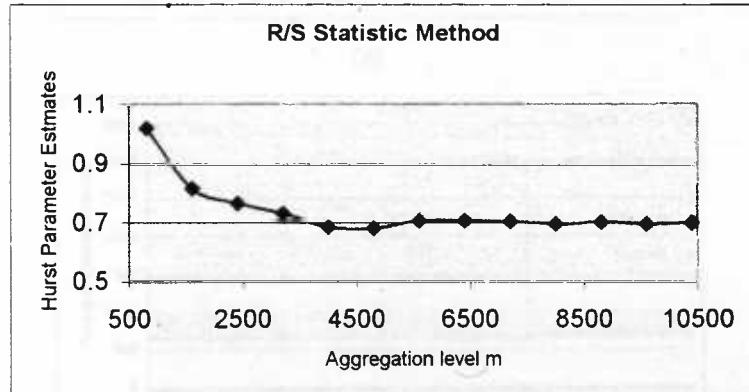


#### Γ.10 ΧΡΟΝΟΛΟΓΙΚΗ ΣΕΙΡΑ ΠΑΚΕΤΩΝ ΕΙΣΟΔΟΥ ΓΙΑ τΟN HOST3

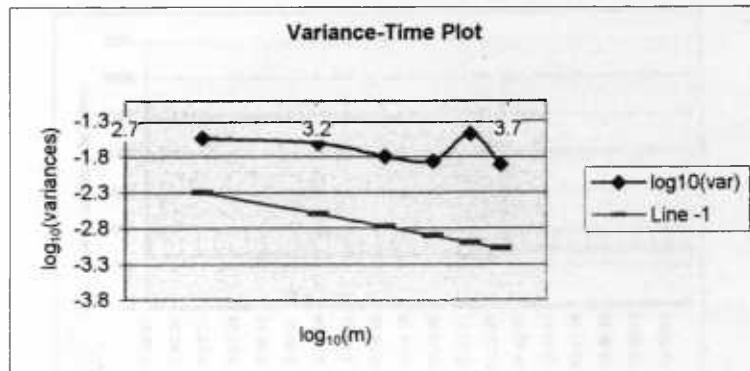
Όπως φαίνεται στο παρακάτω διάγραμμα, οι εκτιμήσεις της παραμέτρου  $H$ , για τα συγκεκριμένα επίπεδα άθροισης κυμαίνονται στο διάστημα [0.63, 0.71].



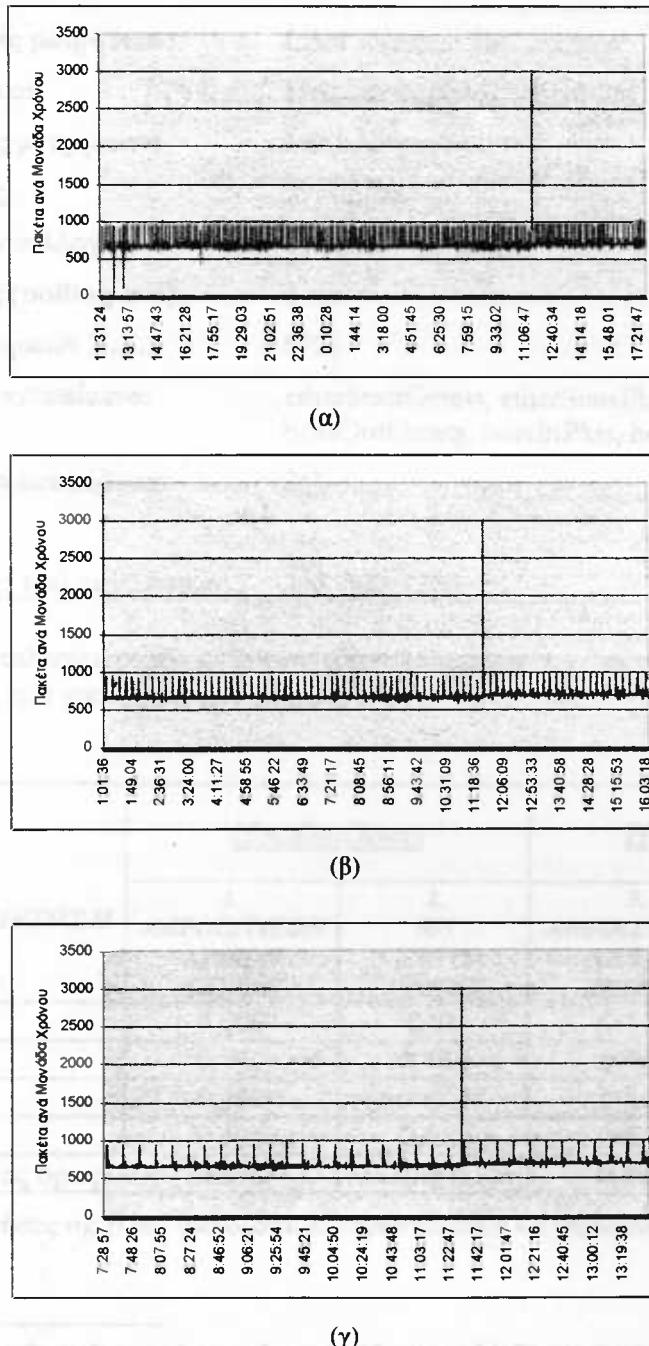
Για μεγάλα επίπεδα άθροισης, οι εκτιμήσεις της παραμέτρου  $H$  σταθεροποιούνται γύρω από την τιμή 0.72.



Από το παρακάνω Viance – Time διάγραμμα δικαιολογείται μια εκτίμηση της παραμέτρου  $H$ , ίση με 0.657.



Το Σχήμα 6.4 (α)-(γ) αφορά τη χρονολογική σειρά των ψηφιοσυλλαβών εισόδου (input bytes) του host1. Στο διάγραμμα 6.4(α) παριστάνεται γραφικά το σύνολο των παρατηρήσεων που συλλέχθηκαν για τη συγκεκριμένη Χ.Σ. Στο διάγραμμα 6.4(β) παριστάνεται το ήμισυ των παρατηρήσεων του διαγράμματος 6.4(α), ενώ στο διάγραμμα 6.4(γ) παριστάνεται γραφικά το ήμισυ του διαγράμματος 6.4(β). Τα τρία αυτά διαγράμματα αν και αναφέρονται σε διαφορετικές κλίμακες χρόνου, εμφανίζουν πανομοιότυπο πρότυπο (pattern) κίνησης. Το γεγονός αυτό αποτελεί μια γρήγορη «οπτική» απόδειξη σχετικά με την ύπαρξη του φαινομένου της αυτό-ομοιότητας στα δεδομένα.



**Σχήμα 6.4** Μία "οπτική" απόδειξη για την ύπαρξη αυτό-ομοιότητας: Και τα τρία διαγράμματα που αφορούν διαφορετικές χρονικές κλίμακες της ίδιας Χ.Σ. εμφανίζουν πανομοιότυπο πρότυπο (pattern) κίνησης.

## 6. ΔΕΔΟΜΕΝΑ ΚΙΝΗΣΗΣ ΑΠΟ ΤΟ ΤΟΠΙΚΟ ΔΙΚΤΥΟ ΤΟΥ ΚΕΝΤΡΟΥ ΥΠΟΛΟΓΙΣΤΩΝ

Στον πίνακα που ακολουθεί συνοψίζονται οι πληροφορίες διεξαγωγής μέτρησης στο τοπικό δίκτυο του Κέντρου Υπολογιστών του Πανεπιστημίου.

Δίκτυο διεξαγωγής μετρήσεων:	LAN Κέντρου Υπολογιστών
Τοποθεσία διεργασίας:	Πτέρυγα Δεριγνύ
Τρόπος συλλογής μετρήσεων:	HP LANprobe
Ημ/νία συλλογής:	21/12/98
Χρονική διάρκεια συλλογής:	13:03:11 – 21:02:25
Ρυθμός συλλογής (polling ανά):	5 sec
Πλήθος παρατηρήσεων Χ.Σ.:	5723
RMON MIB v1 αντικείμενα:	etherStatsOctets, etherStatsPkts, hostInOctets, hostOutOctets, hostInPkts, hostOutPkts
Αριθμός Χ.Σ. που εκτιμήθηκαν	26 <sup>1</sup>
ως προς Η:	
Επίπεδα άθροισης (m) εκτιμήσεων:	400, 800, 1200, ...

Παρακάτω παρατίθενται οι πίνακες των αποτελεσμάτων των εκτιμήσεων του βαθμού αυτό-ομοιότητας των χρονολογικών σειρών.

ΜΕΘΟΔΟΣ ΕΚΤΙΜΗΣΗΣ Η HOST	<u>HostOutOctets</u>		<u>HostOutPackets</u>	
	1. <u>ΑΘΡΟΙΣΤΙΚΩΝ ΧΡ/ΚΩΝ ΣΕΙΡΩΝ</u>	2. <u>R/S ΣΤΑΤΙ- ΣΤΙΚΗΣ</u>	3. <u>ΑΘΡΟΙΣΤΙΚΩΝ ΧΡ/ΚΩΝ ΣΕΙΡΩΝ</u>	4. <u>R/S ΣΤΑΤΙ- ΣΤΙΚΗΣ</u>
1. PC	0.009	0.25	0.117	0.278
2. PC	0.557	0.685	0.401	0.569
3. PC	0.352	0.487	0.394	0.51
4. PC	0.596	0.619	0.572	0.603
Μέσος όρος:	0.378	0.51	0.371	0.49

Πίνακας 1 Οι εκτιμήσεις της Η των χρονολογικών σειρών πακέτων και ψηφιοσυλλαβών εξόδου των 4 host.

<sup>1</sup> Από τα αρχεία χρονολογικών σειρών που δημιουργήθηκαν επιλέχθηκαν για εκτίμηση της Η αυτά που αφορούσαν τις πιο ενεργές (active) συσκευές του δικτύου κατά τη διάρκεια των μετρήσεων.

ΜΕΘΟΔΟΣ ΕΚΤΙΜΗΣΗΣ Η HOST	<u>HostInOctets</u>		<u>HostInPackets</u>	
	1. <b>ΑΘΡΟΙΣΤΙΚΩΝ ΧΡ/ΚΩΝ ΣΕΙΡΩΝ</b>	2. <b>R/S ΣΤΑΤΙ- ΣΤΙΚΗΣ</b>	3. <b>ΑΘΡΟΙΣΤΙΚΩΝ ΧΡ/ΚΩΝ ΣΕΙΡΩΝ</b>	4. <b>R/S ΣΤΑΤΙ- ΣΤΙΚΗΣ</b>
2. PC	0.577	0.686	0.526	0.659
4. PC	0.57	0.685	0.422	0.576
<b>Μέσος όρος:</b>	<b>0.573</b>	<b>0.685</b>	<b>0.474</b>	<b>0.617</b>

Πίνακας 2 Οι εκτιμήσεις της  $H$  των χρονολογικών σειρών πακέτων και ψηφιοσυλλαβών εισόδου των 2 host.

ΜΕΘΟΔΟΣ ΕΚΤΙΜΗΣΗΣ Η	<u>EtherStatsOctets</u>		<u>EtherStatsPackets</u>	
	1. <b>ΑΘΡΟΙΣΤΙΚΩΝ ΧΡ/ΚΩΝ ΣΕΙΡΩΝ</b>	2. <b>R/S ΣΤΑΤΙ- ΣΤΙΚΗΣ</b>	3. <b>ΑΘΡΟΙΣΤΙΚΩΝ ΧΡ/ΚΩΝ ΣΕΙΡΩΝ</b>	4. <b>R/S ΣΤΑΤΙ- ΣΤΙΚΗΣ</b>
<b>ΕΚΤΙΜΗΣΗ Η</b>	0.588	0.617	0.566	0.598

Πίνακας 3 Οι εκτιμήσεις της  $H$  των χρονολογικών σειρών πακέτων και ψηφιοσυλλαβών που εμφανίστηκαν στο δίκτυο.

### ΣΧΟΛΙΑΣΜΟΣ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

#### A. Ο ΒΑΘΜΟΣ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ ΣΕ ΣΧΕΣΗ ΜΕ ΤΟΝ ΒΑΘΜΟ ΧΡΗΣΗΣ ΤΟΥ ΔΙΚΤΥΟΥ

Οι εκτιμήσεις των Πινάκων 1 και 2 αφορούν τέσσερις και δύο αντίστοιχα hosts από τους εβδομηντατρείς που εντοπίστηκαν κατά τη διάρκεια της μέτρησης. Αντίθετα, οι εκτιμήσεις του Πίνακα 3 αφορούν το σύνολο των host του δικτύου. Συγκρίνοντας τους Πίνακες 1 και 2 με τον Πίνακα 3 προκύπτει ότι η μέση τιμή των εκτιμήσεων της παραμέτρου  $H$  (βαθμός αυτό-ομοιότητας) και των τεσσάρων στηλών του Πίνακα 1 (αφορούν τις δύο διαφορετικές μεθόδους εκτίμησης της  $H$  για πακέτα και ψηφιοσυλλαβές) είναι μικρότερη από τις αντίστοιχες εκτιμήσεις των τεσσάρων στηλών του Πίνακα 3. Το ίδιο ισχύει για τις στήλες 1 και 2 του Πίνακα 1. Έτσι, αφού οι χρονολογικές σειρές του Πίνακα 3 αφορούν το σύνολο των host του δικτύου, επιβεβαιώνεται η πρόταση ότι όσο μεγαλύτερη είναι η κίνηση του Ethernet δικτύου, τόσο μεγαλύτερη είναι η εκτίμηση της παραμέτρου  $H$  ή τόσο μεγαλύτερος είναι ο βαθμός αυτό-ομοιότητας.

## Β. ΔΙΑΦΟΡΑ ΣΤΙΣ ΕΚΤΙΜΗΣΕΙΣ ΤΗΣ $H$ ΤΩΝ ΔΥΟ ΜΕΘΟΔΩΝ ΕΚΤΙΜΗΣΗΣ

Συγκρίνοντας τις στήλες των Πινάκων 1, 2 και 3 που αφορούν τις δύο μεθόδους εκτίμησης της  $H$  (δηλαδή τις (1,2) και (3,4)), προκύπτουν τα εξής:

- Η διαφορά στις εκτιμήσεις είναι σχετικά μικρή.
- Οι εκτιμήσεις μιας μεθόδου εκτίμησης δεν εμφανίζονται εν γένει μικρότερες (ή μεγαλύτερες) από αυτές της άλλης μεθόδου εκτίμησης. Με άλλα λόγια, δεν προκύπτει μια πάγια ανισοτική σχέση μεταξύ των εκτιμήσεων η οποία να προσδιορίζεται από την μέθοδο εκτίμησης της  $H$ .
- Το γεγονός της ύπαρξης διαφοράς εκτιμήσεων δεν δημιουργεί προβλήματα «συγκρούσεων»: Όταν η εκτίμηση της μίας μεθόδου υποδεικνύει μεγάλο (ή μικρό) βαθμό αυτό-ομοιότητας, τότε αναμένεται κατ' αντιστοιχία μεγάλος (ή μικρός) βαθμός αυτό-ομοιότητας και για την εκτίμηση της άλλης μεθόδου.

## Γ. ΔΙΑΦΟΡΑ ΣΤΙΣ ΕΚΤΙΜΗΣΕΙΣ ΤΗΣ $H$ ΜΕΤΑΞΥ ΤΗΣ ΚΙΝΗΣΗΣ ΠΑΚΕΤΩΝ & ΨΗΦΙΟΣΥΛΛΑΒΩΝ

Συγκρίνοντας στους Πίνακες 1, 2 και 3 τις στήλες (1,3) ή (2,4) δηλαδή τις στήλες που αφορούν ψηφιοσυλλαβές με τις στήλες που αφορούν πακέτα και οι οποίες προκύπτουν από την ίδια μέθοδο εκτίμησης, προκύπτουν τα εξής:

- Η διαφορά στις εκτιμήσεις είναι σχετικά μικρή.
- Οι εκτιμήσεις των χρονολογικών σειρών που αφορούν πακέτα δεν εμφανίζονται εν γένει μικρότερες (ή μεγαλύτερες) από αυτές που αφορούν ψηφιοσυλλαβές. Με άλλα λόγια, δεν προκύπτει μια πάγια ανισοτική σχέση μεταξύ των εκτιμήσεων η οποία να προσδιορίζεται από τον τρόπο μέτρησης της κίνησης (σε πακέτα ή ψηφιοσυλλαβές).
- Το γεγονός της ύπαρξης διαφοράς εκτιμήσεων δεν δημιουργεί προβλήματα «συγκρούσεων»: Όταν η εκτίμηση μιας συγκεκριμένης μεθόδου υποδεικνύει μεγάλο (ή μικρό) βαθμό αυτό-ομοιότητας για μια χρονολογική σειρά πακέτων, τότε αναμένεται μεγάλος (ή μικρός) βαθμός αυτό-ομοιότητας και για την αντίστοιχη χρονολογική σειρά η οποία είναι μετρημένη σε ψηφιοσυλλαβές.

## Δ. ΔΙΑΦΟΡΑ ΣΤΙΣ ΕΚΤΙΜΗΣΕΙΣ ΤΗΣ $H$ ΓΙΑ ΚΙΝΗΣΗ ΕΙΣΟΔΟΥ & ΕΞΟΔΟΥ

Συγκρίνοντας για κάθε host τις εκτιμήσεις της  $H$  για την κίνηση εισόδου και εξόδου (δηλαδή τις γραμμές υπ' αριθμόν 2 και 4 των Πινάκων 1 και 2), προκύπτει ότι:

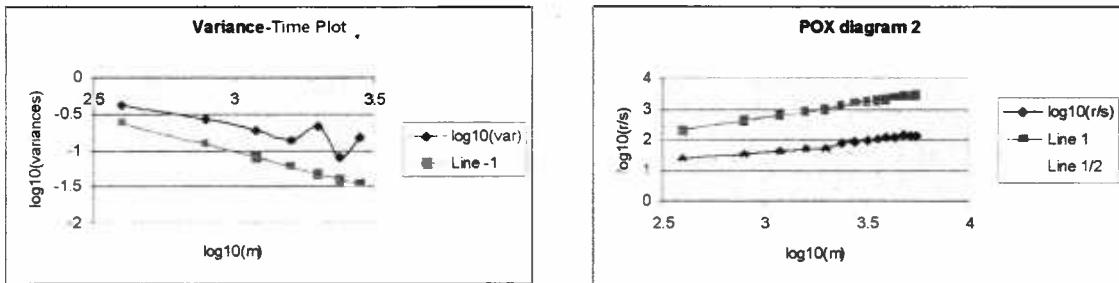
Οι εκτιμήσεις του βαθμού αυτό-ομοιότητας της κίνησης εξόδου για έναν host είναι ανάλογες των εκτιμήσεων της κίνησης εισόδου για τον ίδιο host. Η παρατήρηση ισχύει και για τις δύο μεθόδους εκτίμησης της  $H$ , και για τους δύο τρόπους έκφρασης της κίνησης (πακέτα ή ψηφιοσυλλαβές).

## E. ΕΚΤΙΜΗΣΗ ΤΟΥ ΒΑΘΜΟΥ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ ΤΗΣ ΚΙΝΗΣΗΣ ΤΟΥ ΔΙΚΤΥΟΥ

Από τον Πίνακα 3 προκύπτει ότι οι εκτιμήσεις του βαθμού αυτό-ομοιότητας των χρονολογικών σειρών των πακέτων και των ψηφιοσυλλαβών που εμφανίστηκαν στο δίκτυο κατά τη διάρκεια των μετρήσεων, κυμαίνονται γύρω στο 0.5. Το γεγονός αυτό ισχύει και για τις δύο μεθόδους εκτίμησης και δεν δίνει μια ξεκάθαρη ένδειξη για την ύπαρξη αυτό-ομοιότητας, όπως προκύπτει και από τα παρακάτω διαγράμματα.

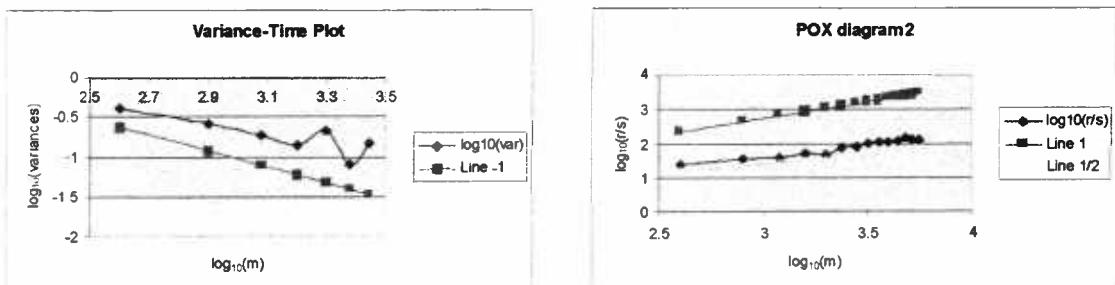
### E.1 ΧΡΟΝΟΛΟΓΙΚΗ ΣΕΙΡΑ ΨΗΦΙΟΣΥΛΛΑΒΩΝ

Ο σχετικά χαμηλός βαθμός της αυτό-ομοιότητας, φαίνεται από τα παρακάτω διαγράμματα. Στο διάγραμμα Διακύμανσης-Χρόνου οι λογαριθμημένες διακυμάνσεις βρίσκονται σε μικρή απόσταση από την γραμμή με κλίση -1, ενώ στο POX διάγραμμα οι λογαριθμημένες τιμές των  $r/s$  βρίσκονται πλησιέστερα στη γραμμή με κλίση  $\frac{1}{2}$ , γεγονός που αποτελεί ένδειξη για χαμηλό βαθμό αυτό-ομοιότητας.



### E.2 ΧΡΟΝΟΛΟΓΙΚΗ ΣΕΙΡΑ ΠΑΚΕΤΩΝ

Η ομοιότητα των δύο επόμενων και των δύο προηγούμενων διαγραμμάτων αποδεικνύει την ομοιόμορφία των εκτιμήσεων της  $H$  για την κίνηση πακέτων και ψηφιοσυλλαβών. Στο παρακάτω διάγραμμα Διακύμανσης-Χρόνου οι λογαριθμημένες διακυμάνσεις βρίσκονται σε μικρή απόσταση από τη γραμμή με κλίση -1, ενώ στο POX διάγραμμα οι λογαριθμημένες  $r/s$  τιμές βρίσκονται κοντά στη γραμμή με κλίση  $\frac{1}{2}$ .



## ΚΕΦΑΛΑΙΟ 7

# ΣΥΛΛΟΓΗ & ΕΚΤΙΜΗΣΗ ΒΑΘΜΟΥ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ ΕΞΩΤΕΡΙΚΗΣ ΚΙΝΗΣΗΣ ΤΟΥ ΔΙΚΤΥΟΥ ΔΕΔΟΜΕΝΩΝ ΤΟΥ Ο.Π.Α.

### 1. ΕΙΣΑΓΩΓΗ

Στα πλαίσια της προσπάθειας χαρακτηρισμού της εξωτερικής κίνησης του δικτύου δεδομένων του Πανεπιστημίου σε σχέση με το φαινόμενο της αυτό-ομοιότητας, συλλέχθηκαν δεδομένα κίνησης από το δρομολογητή της εταιρείας Cisco ο οποίος συνδέει το δίκτυο του Πανεπιστημίου με το Internet. Στο κεφάλαιο αυτό περιγράφονται οι βασικές τεχνολογίες συλλογής δεδομένων κίνησης από τον δρομολογητή ενώ παρουσιάζονται τα αποτελέσματα των εκτιμήσεων του βαθμού αυτό-ομοιότητας των δειγμάτων κίνησης που συλλέχθηκαν.

### 2. Η ΤΕΧΝΟΛΟΓΙΑ NETFLOW

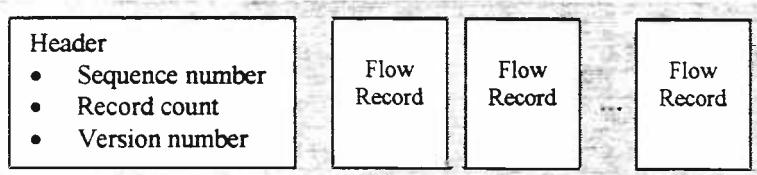
Η συλλογή δεδομένων κίνησης από δρομολογητές της εταιρείας Cisco βασίζεται στην τεχνολογία NetFlow. Βάσει αυτής της τεχνολογίας εντοπίζεται το είδος και η προτεραιότητα της κίνησης, παρέχοντας κυρίως τη δυνατότητα κοστολόγησης ανάλογα με την ποιότητα υπηρεσίας (QoS), αλλά και την υποστήριξη ενός συνόλου άλλων εφαρμογών που σχετίζονται με τον σχεδιασμό και την παρακολούθηση δικτύου.

Τα στοιχεία κίνησης συλλέγονται σε επίπεδο ροής. Ως ροή δικτύου ορίζεται μια μονής κατεύθυνσης ακολουθία πακέτων μεταξύ συγκεκριμένου ζεύγους πηγής και προορισμού. Η πηγή και ο προορισμός κάθε ροής προσδιορίζονται μοναδικά από τις IP διευθύνσεις τους και τους αριθμούς των θυρών (port numbers) επιπέδου μεταφοράς. Επίσης, για τον μοναδικό προσδιορισμό των ροών χρησιμοποιούνται ο τύπος του IP πρωτοκόλλου, ο τύπος υπηρεσίας (ToS) καθώς και τα αναγνωριστικά των διεπαφών.

Κάθε ροή στον δρομολογητή έχει μια διαδικασία δημιουργίας και λήξης όπως ορίζεται από το λογισμικό του δρομολογητή. Στο αρχικό πακέτο κάθε ροής δημιουργείται μια καταχώρηση στην κρυφή μνήμη ροών (flow cash). Όταν

λαμβάνεται ένα πακέτο που δεν μπορεί να ενταχθεί σε καμία ροή που βρίσκεται ήδη στην κρυφή μνήμη, δημιουργείται μια νέα καταχώρηση ροής. Η κατάσταση και η ταυτότητα μιας ροής διατηρείται όσο υπάρχει η κίνηση ενεργών πακέτων αυτής της ροής. Όταν η κίνηση για την δεδομένη ροή σταματήσει, η ροή εκπνέει. Η εκπνοή της ροής καθορίζεται από ευρετική μέθοδο που βασίζεται στο πρωτόκολλο, στο τελευταίο πακέτο που λήφθηκε και στο χρόνο λήψης του.

Οι ροές που εκπνέουν ομαδοποιούνται σε “NetFlow Export” UDP datagrams προκειμένου να εξαχθούν από την συσκευή. Το NetFlow Export datagram αποτελείται από μια επικεφαλίδα και μια ακολουθία από εγγραφές ροών, όπως φαίνεται στο Σχήμα 7.1.



Σχήμα 7.1 Η μορφή του NetFlow Export Datagram

Τα περιεχόμενα των εγγραφών των ροών (Flow Records), έχουν ως εξής:

- |   |                                |
|---|--------------------------------|
| 1. Source IP address                    | 11. End of flow timestamp      |
| 2. Destination IP address               | 12. IP protocol                |
| 3. Source TCP/UDP application port      | 13. Type of Service (ToS) byte |
| 4. Destination TCP/UDP application port | 14. TCP Flags                  |
| 5. Next hop router IP address           | 15. Source AS number           |
| 6. Input physical interface index       | 16. Destination AS number      |
| 7. Output physical interface index      | 17. Source subnet mask         |
| 8. Packet count for this flow           | 18. Destination subnet mask    |
| 9. Byte count for this flow             | 19. Flags                      |
| 10. Start of flow timestamp             | 20. Shortcut router IP address |

Η συμβατική μεταγωγή πακέτων στο επίπεδο δικτύου χειρίζεται τα εισερχόμενα πακέτα ανεξάρτητα, με ξεχωριστές σειριακές εργασίες για την μεταγωγή, την ασφάλεια, τις υπηρεσίες και τις μετρήσεις κίνησης. Με την μεταγωγή NetFlow, αυτές οι διαδικασίες εφαρμόζονται μόνο στο πρώτο πακέτο της ροής. Οι πληροφορίες του πρώτου πακέτου χρησιμοποιούνται για να δημιουργήσουν μια καταχώρηση στην κρυφή μνήμη του NetFlow. Τα επόμενα πακέτα της ροής αντιμετωπίζονται μέσω μιας διαδικασίας κατά την οποία η μεταγωγή, η συλλογή λογιστικών δεδομένων, και οι υπόλοιπες υπηρεσίες πραγματοποιούνται ταυτόχρονα.

Με αυτόν τον τρόπο το NetFlow εκμεταλλεύεται τη φύση της κίνησης, δηλαδή τη μορφή που έχουν οι ροές, έτσι ώστε:

- Να παρέχει συλλογή λεπτομερών λογιστικών δεδομένων με την ελάχιστη επίπτωση στην επίδοση του δρομολογητή. Οι μετρήσεις του NetFlow αποθηκεύονται και ενημερώνονται στην κρυφή του μνήμη ως μια ακέραια συνιστώσα της διαδικασίας μεταγωγής.

- Να επεξεργάζεται αποδοτικά τις λίστες προσπέλασης για το φιλτράρισμα των πακέτων και τις υπηρεσίες ασφάλειας, καθώς μόνο το πρώτο πακέτο υπόκειται σε έλεγχο ως προς τις λίστες προσπέλασης.

Υπάρχουν δύο τρόποι εξαγωγής δεδομένων από έναν δρομολογητή της εταιρείας Cisco στον οποίο έχει ενεργοποιηθεί το NetFlow:

- Με την εντολή *show ip cashe flow*. Η έξοδος αυτής της εντολής παρέχει μια συνοπτική αναφορά για τις τιμές των μεγεθών που έχουν συλλεχθεί μέχρι τη στιγμή που εκδόθηκε η εντολή, ένα πίνακα που δίνει συνοπτική αναφορά στατιστικών για κάθε υπηρεσία (π.χ. TCP-Telnet, TCP-FTP, UDP-DNS, κ.λπ.) καθώς και ένα πίνακα με τις ροές που ήταν ενεργές τη στιγμή που εκδόθηκε η εντολή.
- Μέσω του μηχανισμού εξαγωγής δεδομένων του NetFlow: Σε κάθε διεπαφή (interface) του δρομολογητή εκτελούμε την εντολή *ip route-cashe flow*. Η εντολή αυτή ενεργοποιεί την διαδικασία τήρησης λογιστικών πληροφοριών στην συγκεκριμένη διεπαφή. Αντίθετα, η εντολή *no ip route-cashe flow* απενεργοποιηθεί την διαδικασία αυτή. Για να ξεκινήσει η διαδικασία μεταφοράς των δεδομένων, αρκεί να εκτελεστεί η εντολή *ip flow-export <IP address> <UDP port>*, όπου *<IP address>* και *<UDP port>* είναι η IP διεύθυνση και το UDP port του host στον οποίο θα εξαχθούν τα δεδομένα των ροών.

### 3. Η ΕΦΑΡΜΟΓΗ ΣΥΛΛΟΓΗΣ ΔΕΔΟΜΕΝΩΝ ΚΙΝΗΣΗΣ CFLOWD

Τα NetFlow Export πακέτα που περιγράφτηκαν παραπάνω, αποστέλλονται συνήθως σε ένα σταθμό εργασίας ο οποίος καθορίζεται από τον χρήστη και στον οποίο τρέχει μια εφαρμογή λήψης και επεξεργασίας των NetFlow Export πακέτων. Τέτοιες εφαρμογές αποτελούν ο FlowCollector της εταιρείας Cisco καθώς και το cflowd. Στην παράγραφο αυτή παρουσιάζονται η λειτουργία και οι δυνατότητες της εφαρμογής cflowd.

Το cflowd έχει τη δυνατότητα να αποθηκεύει για κάθε δρομολογητή από τον οποίο λαμβάνει δεδομένα ροών, τα εξής δεδομένα:

1. Συνολικός αριθμός ροών που παραλήφθηκαν από τον δρομολογητή (flows received)
2. Συνολικός αριθμός ροών που χάθηκαν από τον δρομολογητή (missed flows)
3. Συνολικός αριθμός πακέτων και ψηφιοσυλλαβών (bytes) κάθε IP πρωτοκόλλου (protocol traffic)
4. Συνολικός αριθμός πακέτων και ψηφιοσυλλαβών κάθε UDP/TCP port number (port traffic)
5. Συνολικός αριθμός πακέτων και ψηφιοσυλλαβών που στέλνονται από το ένα δίκτυο στο άλλο (net traffic matrix)
6. Συνολικός αριθμός πακέτων και ψηφιοσυλλαβών που στέλνονται από το ένα αυτόνομο σύστημα στο άλλο (AS traffic matrix)
7. Ανεπεξέργαστα (raw) δεδομένα ροών σε δυαδική μορφή (raw flows)

### 3.1 ΡΥΘΜΙΣΗ ΤΟΥ CFLOWD

Το cflowd λαμβάνει όλες τις πληροφορίες ρύθμισής του (configuration) από ένα αρχείο που συνήθως ονομάζεται cflowd.conf. Στο αρχείο αυτό υπάρχουν τρία είδη εγγραφών που πρέπει να ρυθμιστούν:

- Η εγγραφή OPTIONS, βάσει της οποίας καθορίζονται γενικές επιλογές, όπως οι κατάλογοι μέσα στους οποίους θα τοποθετηθούν τα δεδομένα που λαμβάνονται.
- Η εγγραφή CISCOEXPORTER, βάσει της οποίας καθορίζονται πληροφορίες που αφορούν τη συσκευή Cisco η οποία αναμένεται να στείλει δεδομένα ροών στο cflowd. Τέτοιες πληροφορίες περιλαμβάνουν την IP διεύθυνση της Cisco συσκευής, το port number στο οποίο η Cisco συσκευή θα στέλνει πακέτα, κ.α.
- Η εγγραφή COLLECTOR, βάσει της οποίας προσδιορίζεται η IP διεύθυνση του υπολογιστή στον οποίο τρέχει η εφαρμογή cflowd.

### 3.2 ΟΝΟΜΑΤΟΔΟΣΙΑ ΑΡΧΕΙΩΝ ΤΟΥ CFLOWD

Τα αρχεία δεδομένων που δημιουργούνται από το cflowd ακολουθούν μια σύμβαση ονοματοδοσίας η οποία αποτελείται από την IP διεύθυνση της συσκευής Cisco από την οποία προέρχονται τα δεδομένα, τον τύπο των δεδομένων, και το εάν η όχι το αρχείο περιλαμβάνει δεδομένα τα οποία το cflowd επεξεργάζεται εκείνη τη στιγμή. Για παράδειγμα:

195.251.252.126.port.cur

IP address    port    current  
                data    data

195.251.252.126.protocol.prev

IP address    protocol previous  
                data    data

Επίσης, το όνομα κάθε αρχείου ανεπεξέργαστων (raw) δεδομένων ροών είναι της μορφής:

195.251.252.126.flows.x

IP address    flow  
                data

όπου x ένας ακέραιος αριθμός ο οποίος προσδιορίζει ένα συγκεκριμένο αρχείο ανεπεξέργαστων δεδομένων.



### 3.3 ΔΙΑΔΙΚΑΣΙΕΣ ΤΟΥ CFLOWD

Το cflowd περιλαμβάνει έναν αριθμό διαδικασιών οι οποίες επιτρέπουν την εμφάνιση των περιεχομένων των αρχείων δεδομένων που δημιουργούνται. Αυτές είναι:

#### 3.3.1 cfdfows

Η διαδικασία cfdfows εμφανίζει τα περιεχόμενα των αρχείων των ανεπεξέργαστων (raw) ροών δεδομένων στην προκαθορισμένη ροή εξόδου σε αναγνώσιμη μορφή. Παράδειγμα:

```
% cfdfows 195.251.252.126.flows
FLOW TABLE
Index:          0xff
Cisco_vers:    5
SysUpTime:     84219715
Unix_secs:     843858164 (09/27/96  17:02:44)
Unix_nsecs:    179
Flow_seq:      1470
Num_entries:   30
FLOW ENTRY
Index:          0xffff
Srcaddr:        192.152.1.5
Dstaddr:        204.100.7.6
Descriptor:     0x30
Input_if:       0
Output_if:      0
Srcport:        16896
Dstport:        1092
Pkts:           25
Bytes:          232
Nexthop:        195.176.1.5
Starttime:      843858164
Endtime:        843958184
Protocol:       6
Tos:             0
Src_as:          200
Dst_as:          600
Src_mask:        24
Dst_mask:        24
FLOW ENTRY
**
```

### 3.3.2 cfdports

Η διαδικασία cfdports εμφανίζει τα περιεχόμενα των αρχείων που αφορούν αριθμούς θυρών (port numbers). Παράδειγμα:

```
% cfdports 195.251.252.126.port.prev
```

Port Table

-----  
Num ports: 16282

Port	input		output	
	packets	bytes	packets	bytes
5000	65	3883	6216	258773
80	626338	321896224	603370	35887064
3251	237	12110	217	45387
1152	812	54217	761	489290
1597	218	14928	281	124925
123	4313	327788	4331	329156

...

### 3.3.3 cfdprotos

Η διαδικασία cfdprotos εμφανίζει τα περιεχόμενα των αρχείων που περιλαμβάνουν δεδομένα πρωτοκόλλων (protocol traffic data). Παράδειγμα:

```
% cfdprotos 195.251.252.126.protocol.prev
```

Protocol Table

Protocol	packets	bytes
1	11890	785297
2	1112	156101
4	23288	6553054
6	1927473	549201762
17	210765	31545429
47	393	144750

...

### 3.3.4 cfdnets

Η διαδικασία cfdnets εμφανίζει τα περιεχόμενα των αρχείων που αφορούν δεδομένα κίνησης μεταξύ δικτύων. Παράδειγμα:

```
% cfdnets 195.251.252.126.netmatrix.prev
```

Net Matrix

Number of net entries: 557529

SRCCNET/MASK->	DSTNET/MASK	PKTS	BYTES
7.248.232.97/32 ->	164.58.0.0/16	1	80
204.162.96.0/22 ->	199.174.128.0/17	49	13201
192.216.50.8/32 ->	205.228.184.0/24	6	382
206.214.0.0/15 ->	204.123.0.0/16	2215	395727

### 3.3.5 cfdases

Η διαδικασία cfdases εμφανίζει τα περιεχόμενα των αρχείων που αφορούν δεδομένα κίνησης μεταξύ αυτόνομων συστημάτων. Παράδειγμα:

```
% cfdases 195.251.252.126.asmatrix.prev  
As Matrix
```

SrcAS->	dstAS	PKTS	BYTES
0 ->	0	19094	3561039
0 ->	200	11050	18801282
200 ->	4183	113438	34522979
0 ->	33	13966	872894

## 3.4 ΜΕΤΑΤΡΟΠΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΚΙΝΗΣΗΣ ΣΕ ΣΤΟΙΧΕΙΑ ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ

Όπως θα έχει ήδη γίνει κατανοητό από προηγούμενα κεφάλαια το φαινόμενο της αυτό-ομοιότητας χαρακτηρίζει διαδικασίες οι οποίες εξελίσσονται στο χρόνο (ή το χώρο). Ετσι, απαραίτητη προϋπόθεση προκειμένου να εκτιμηθεί ο βαθμός αυτό-ομοιότητας ενός δείγματος δεδομένων κίνησης αποτελεί το δείγμα να έχει τη μορφή μιας χρονολογικής σειράς, ώστε να είναι γνωστός ο τρόπος εξέλιξης των δεδομένων κίνησης στο χρόνο.

Όπως φαίνεται από τις παραπάνω παραγράφους η εφαρμογή λήψης δεδομένων ροών από τον Cisco δρομολογητή, cflowd δεν παρέχει στοιχεία κίνησης σε μορφή χρονολογικών σειρών. Για τον λόγο αυτό υλοποιήθηκε μια μακροεντολή στο πακέτο επεξεργασίας λογιστικών φύλλων Excel, η οποία βάσει των στοιχείων κίνησης που παρέχει το cflowd δημιουργεί την αντίστοιχη χρονολογική σειρά. Σημειώνεται ότι ο κώδικας της μακροεντολής παρατίθεται στο Παράρτημα 3.

Το cflowd μέσω της διαδικασίας cfdflows δίνει μεταξύ άλλων τις εξής πληροφορίες για κάθε ροή α) χρόνος έναρξης, β) χρόνος λήξης, γ) πλήθος πακέτων, δ) πλήθος ψηφιοσυλλαβών. Προκειμένου να παραχθεί η χρονολογική σειρά η οποία θα αφορά τα παραπάνω στοιχεία όλων των ροών, ακολουθείται η εξής διαδικασία:

- Ο αύξων αριθμός, ο χρόνος έναρξης, ο χρόνος λήξης, το πλήθος των πακέτων και το πλήθος των ψηφιοσυλλαβών κάθε ροής, εισάγονται στα κελιά ενός φύλλου εργασίας του Excel, ως εξής:

Flow	Packets	Bytes	Start Time	End Time	Time Series	Packets	Bytes
1.	25	232	164	184			
2.	18	188	175	193			
3.	33	243	186	196			
4.	17	212	189	192			
5.	22	250	200	208			

Σημειώνεται ότι οι ροές θα πρέπει να εισάγονται στις γραμμές του φύλλου εργασίας σε τρόπο ώστε η στήλη *Start Time* να είναι ταξινομημένη σε αύξουσα σειρά. Επίσης, αναμένεται ότι τα δεδομένα θα αρχίσουν να εισάγονται στις θέσεις που δείχνει ο παραπάνω πίνακας, αρχίζοντας από το κελί (1,1) (εκεί τοποθετείται ο τίτλος *Flow*), χωρίς κενά μεταξύ τους.

- Εκτελείται η μακροεντολή *ExportTimeSeries()*, η οποία εισάγει δεδομένα στις στήλες με τίτλο *Time Series*, *Packets* και *Bytes*. Αυτές οι τρεις στήλες αποτελούν τις δύο χρονολογικές σειρές των πακέτων και των ψηφιοσυλλαβών οι οποίες αντιστοιχούν στα δεδομένα των ροών εισόδου. Ήτοι, μετά την εκτέλεση της μακροεντολής, το φύλλο εργασίας που περιγράφεται παραπάνω θα έχει τη μορφή:

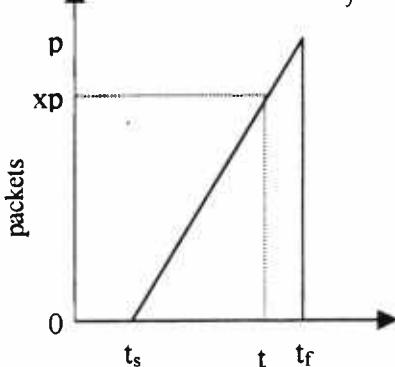
<i>Flow</i>	<i>Packets</i>	<i>Bytes</i>	<i>Start Time</i>	<i>End Time</i>	<i>Time Series</i>	<i>Packets</i>	<i>Bytes</i>
1.	25	232	164	184	164	0	0
2.	18	188	175	193	166	2.5	23.2
3.	33	243	186	196	168	5	46.4
4.	17	212	189	192	170	7.5	69.6
5.	22	250	200	208	172	10	92.8
					174	12.5	116
					176	16	149.644
					178	20.5	193.733
					180	25	237.822
					182	29.5	326
					184	34	114.888
					186	11	184.377

### ΜΗΧΑΝΙΣΜΟΣ ΕΞΑΓΩΓΗΣ ΧΡΟΝΟΛΟΓΙΚΗΣ ΣΕΙΡΑΣ

Όσον αφορά τον μηχανισμό εξαγωγής της χρονολογικής σειράς ισχύουν τα εξής: Υποθέτουμε ότι στον χρόνο έναρξης κάθε ροής  $t_s$  αντιστοιχούν 0 πακέτα και 0 bytes. Επίσης, υποθέτουμε ότι στον χρόνο λήξης κάθε ροής  $t_f$  αντιστοιχούν  $p$  το πλήθος πακέτα και  $b$  το πλήθος bytes, όπου  $p$  και  $b$  το πλήθος των πακέτων και των bytes της κάθε ροής. Τέλος, υποθέτουμε ότι ο αριθμός των πακέτων  $xp$  που αντιστοιχούν στο

χρόνο  $t$  όπου  $t_s < t < t_f$ , είναι  $xp = \frac{p*t}{t_f - t_s}$ . Αντίστοιχα, για τον αριθμό των bytes που

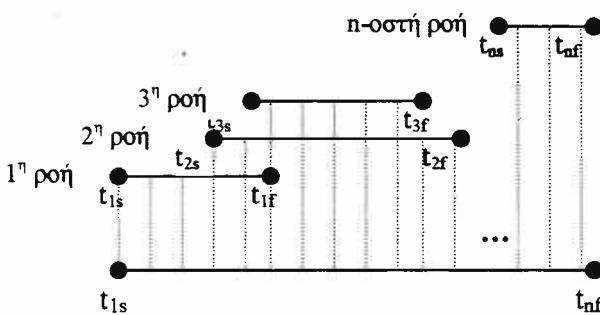
αντιστοιχούν στο χρόνο  $t$  όπου  $t_s < t < t_f$ , ισχύει  $xb = \frac{b*t}{t_f - t_s}$ . Σχηματικά έχουμε:



Σχήμα 7.2 Δημιουργία χρονολογικών στοιχείων ροής βάσει των στοιχείων του αριθμού

Με αυτόν τον τρόπο, έχοντας τους αριθμούς των πακέτων και των bytes που αντιστοιχούν στους χρόνους έναρξης και λήξης μιας ροής, μπορούμε να βρούμε να βρούμε τους αριθμούς των πακέτων και των bytes που αντιστοιχούν σε οποιαδήποτε χρονική στιγμή μεταξύ των χρόνων έναρξης και λήξης. Έτσι, βρίσκοντας τον αριθμό των πακέτων και των bytes στους χρόνους  $t_s$ ,  $t_s + \tau$ ,  $t_s + 2\tau$ , ...,  $t_s + n\tau = t_f$ , προκύπτει η χρονολογική σειρά των πακέτων και των bytes για τη συγκεκριμένη ροή.

Στην πραγματικότητα δεν ενδιαφερόμαστε για την εξαγωγή της χρονολογικής σειράς μιας συγκεκριμένης ροής, αλλά για τον υπολογισμό μιας χρονολογικής σειράς η οποία θα είναι αντιτροσωπευτική για το σύνολο των ροών. Η τελευταία προκύπτει ως εξής: Ας υποθέσουμε ότι έχουμε ή το πλήθος ροές ταξινομημένες ως προς το χρόνο έναρξης τους, και  $t_{is}$ ,  $t_{if}$ , οι χρόνοι έναρξης και λήξης της  $i$  ροής, όπου  $1 \leq i \leq n$ . Έτσι, από το χρονική στιγμή έναρξης της πρώτης ροής  $t_{is}$ , μέχρι τη χρονική στιγμή λήξης της τελευταίας ροής  $t_{if}$ , μπορούν να συμβούν επικαλύψεις στη διάρκεια ορισμένων ροών. Ένα παράδειγμα επικαλύψεων φαίνεται στο παρακάτω σχήμα:



Σχήμα 7.3 Δημιουργία χρονολογικών στοιχείων ροών λαμβάνοντας υπόψη τις χρονικές επικαλύψεις που παρουσιάζουν.

Λαμβάνοντας υπ' όψη τις χρονικές επικαλύψεις των ροών, η χρονολογική σειρά που αντιστοιχεί στο σύνολο των ροών προκύπτει λαμβάνοντας σε κάθε χρονικό σημείο της σειράς το άθροισμα του αριθμού των πακέτων (ή των bytes) κάθε μιας από τις ροές που επικαλύπτονται χρονικά στο συγκεκριμένο σημείο.

#### 4. Η ΕΦΑΡΜΟΓΗ ΣΥΛΛΟΓΗΣ ΔΕΔΟΜΕΝΩΝ ΚΙΝΗΣΗΣ FLOWCOLLECTOR

Η εφαρμογή FlowCollector παρέχει τη δυνατότητα συλλογής δεδομένων ροής από ένα πλήθος συσκευών στις οποίες έχει ενεργοποιηθεί η εφαρμογή NetFlow. Αυτό επιτυγχάνεται ορίζοντας τον FlowCollector να «ακούει» σε συγκεκριμένες UDP θύρες (ports) προκειμένου να παραλάβει και να επεξεργαστεί τα εξαγόμενα datagrams της NetFlow εφαρμογής. Συγκεκριμένα ο FlowCollector αποτελεί μια UNIX εφαρμογή η οποία:

- Παραλαμβάνει τα datagrams που εξάγονται από τις συσκευές στις οποίες έχει ενεργοποιηθεί το NetFlow.
- Επεξεργάζεται τα δεδομένα που παραλαμβάνει ώστε να μπορεί μέσα από κατάλληλες ομαδοποιήσεις να εμφανίσει διαφορετικές μορφές τους.

- Αποθηκεύει ανεπεξέργαστες πληροφορίες ροών για την επεξεργασία τους από άλλες εφαρμογές.

Ο FlowCollector δίνει τη δυνατότητα επιλεκτικής μείωσης του όγκου των δεδομένων που συλλέγονται προκειμένου αυτά να μπορούν να προσαρμοστούν στις απαιτήσεις των εφαρμογών που θα τα χρησιμοποιήσουν. Αυτό επιτυγχάνεται μέσα από τους μηχανισμούς των φίλτρων και των σχημάτων ομαδοποίησης.

### Φιλτράρισμα Δεδομένων

Ο μηχανισμός των φίλτρων δίνει τη δυνατότητα στον χρήστη να προσδιορίσει ποια πεδία ροών θα γίνουν αποδεκτά και ποια πεδία θα απορριφθούν για περαιτέρω επεξεργασία και αποθήκευση. Ο FlowCollector παρέχει τη δυνατότητα φιλτραρίσματος στα ακόλουθα πεδία ροών:

Source IP address  
Destination IP address  
Next hop router IP address  
Source TCP/UDP application port  
Destination TCP/UDP application port  
Input physical interface index  
Output physical interface index  
Source AS number  
Destination AS number  
Type of Service (ToS) byte  
Export datagram source IP address

### Σχήματα Ομαδοποίησης

Τα σχήματα ομαδοποίησης που παρέχει ο FlowCollector επιτρέπουν τη μείωση του όγκου των δεδομένων που εξάγονται. Αυτό επιτυγχάνεται μέσω της αποθήκευσης μιας μορφής «περιλήψης» των δεδομένων των ροών ανάλογα με το σχήμα της ομαδοποίησης. Ενδεικτικά, ένα παράδειγμα σχήματος ομαδοποίησης αποτελεί το Source node για το οποίο αποθηκεύεται μια γραμμή για κάθε IP διεύθυνση πηγής η οποία περιλαμβάνει την διεύθυνση πηγής, το συνολικό αριθμό πακέτων και ψηφιοσυλλαβών που στάλθηκαν καθώς και τον συνολικό αριθμό ροών.

### Ο μηχανισμός των Threads

Ο FlowCollector παρέχει έναν μηχανισμό προσδιορισμού πολιτικών επεξεργασίας και αποθήκευσης δεδομένων για κάθε θύρα συλλογής δεδομένων, ο οποίος καλείται *threads*. Ο ορισμός ενός τέτοιου μηχανισμού πραγματοποιείται με τη βοήθεια φίλτρων και σχημάτων ομαδοποίησης ενώ παράλληλα ορίζονται και παράμετροι όπως μονοπάτια καταλόγων αποθήκευσης και χρονικά διαστήματα συλλογής. Πολλά threads μπορούν να οριστούν για κάθε θύρα προκειμένου να ικανοποιηθούν οι ανάγκες διαφορετικών εφαρμογών (χρέωση, σχεδιασμός δικτύου κλπ.)



## 5. ΔΙΑΔΙΚΑΣΙΑ ΣΥΛΛΟΓΗΣ ΔΕΔΟΜΕΝΩΝ ΕΞΩΤΕΡΙΚΗΣ ΚΙΝΗΣΗΣ ΑΠΟ ΤΟ ΔΙΚΤΥΟ ΔΕΔΟΜΕΝΩΝ ΤΟΥ Ο.Π.Α.

Στα πλαίσια της παρούσας εργασίας και προκειμένου να συλλεχθούν δεδομένα κίνησης από το δρομολογητή της εταιρείας Cisco ο οποίος συνδέει το δίκτυο του Πανεπιστημίου με το Internet, ακολουθήθηκε η εξής διαδικασία:

1. Η εφαρμογή NetFlow ενεργοποιείται στο Cisco δρομολογητή σύνδεσης με το Internet.
2. Η εφαρμογή FlowCollector εγκαθίσταται σε σταθμό εργασίας με λειτουργικό σύστημα UNIX.
3. Στον Cisco δρομολογητή ορίζεται η εξαγωγή των datagrams της εφαρμογής NetFlow στην IP διεύθυνση του σταθμού εργασίας στον οποίο «τρέχει» η εφαρμογή FlowCollector και στο port στο οποίο «ακούει».
4. Εκτελείται η κατάλληλη διαδικασία για την αποθήκευση των δεδομένων κίνησης που συλλέγονται σε ένα ASCII αρχείο.

### 5.1 ΕΞΑΓΩΜΕΝΟ ΑΡΧΕΙΟ ΔΕΔΟΜΕΝΩΝ ΚΙΝΗΣΗΣ

Στο σημείο αυτό περιγράφεται η μορφή και ο τρόπος κατασκευής του αρχείου των εξωτερικών δεδομένων κίνησης το οποίο εξάγεται από την FlowCollector εφαρμογή με την εκτέλεση κατάλληλης διαδικασίας και το οποίο χρησιμοποιήθηκε στην παρούσα εργασία.

Κάθε 1 sec αποθηκεύονται στο ev λόγω αρχείο σε ξεχωριστές γραμμές τα στοιχεία κάθε μιας από τις ροές που εντοπίστηκαν κατά τη διάρκεια του δευτερολέπτου. Τα στοιχεία αυτά περιλαμβάνουν την IP διεύθυνση πηγής (Src-IP), την IP διεύθυνση προορισμού (DstIP), το port πηγής (SPort), το port προορισμού (DPort), τη διεπαφή εισόδου (InInt), τη διεπαφή εξόδου (OutInt), το χρησιμοποιούμενο πρωτόκολλο (Prot) καθώς και τον συνολικό αριθμό πακέτων (Pkts) και bytes (Octets) που μεταφέρθηκαν. Σε περίπτωση που για το χρονικό διάστημα του ενός δευτερολέπτου δεν εντοπιστούν ροές, δεν πραγματοποιείται καμία εγγραφή στο αρχείο. Παρακάτω παραθέτονται ενδεικτικά οι πρώτες γραμμές ενός τέτοιου αρχείου αποθήκευσης δεδομένων κίνησης:

Listening on 9991

Flowstat 1 [13] from 195.251.255.153, Fri Jan 8 13:56:51 1999								
SrcIP	DstIP	SPort	DPort	InInt	OutInt	Prot	Pkts	Octets
209.1.112.230	195.251.248.26	80	1117	160	32	6	2	80
209.1.112.251	195.251.248.239	80	1085	160	32	6	4	265
209.1.112.252	195.251.248.239	80	1082	160	32	6	6	636
209.1.112.251	195.251.248.239	80	1084	160	32	6	3	225
209.1.112.251	195.251.248.239	80	1083	160	32	6	3	225
209.1.112.161	195.251.248.237	80	1090	160	32	6	2	864
209.1.112.252	195.251.248.219	80	1466	160	32	6	6	4040
198.3.99.211	195.251.248.81	80	1120	160	32	6	7	3380
198.3.99.70	195.251.248.81	80	1121	160	32	6	13	14485
195.167.100.39	195.251.248.74	80	1047	160	32	6	3	120
195.167.100.39	195.251.248.74	80	1048	160	32	6	3	120
195.167.100.39	195.251.248.74	80	1049	160	32	6	3	120



207.211.106.40	195.251.251.181	80	1286	160	32	6	4	794
Flowstat 2 [19] from 195.251.255.153, Fri Jan 8 13:56:52 1999								
SrcIP	DstIP	SPort	DPort	InInt	OutInt	Prot	Pkts	Octets
207.82.198.150	195.251.255.142	53	772	160	32	17	3	1419
204.162.80.139	195.251.248.71	80	1404	160	32	6	3	288
139.91.1.1	195.251.248.21	53	1468	160	32	17	2	358
199.95.207.68	195.251.255.142	53	2772	160	32	17	2	162
194.177.210.210	195.251.255.66	53	9768	160	32	17	2	486
194.177.210.210	195.251.255.66	53	9770	160	32	17	2	546

## 5.2 ΜΕΤΑΤΡΟΠΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΚΙΝΗΣΗΣ ΣΕ ΣΤΟΙΧΕΙΑ ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ

Όπως έχει ήδη αναφερθεί βασική προϋπόθεση προκειμένου να εκτιμηθεί ο βαθμός αυτό-ομοιότητας ενός δείγματος δεδομένων κίνησης είναι το δείγμα των δεδομένων κίνησης να έχει τη μορφή χρονοσειράς. Για το σκοπό αυτό υλοποιήθηκε μακροεντολή στο πακέτο επεξεργασίας λογιστικών φύλλων Excel, η οποία βάσει των δεδομένων κίνησης που παρέχει ο FlowCollector κατασκευάζει χρονολογικές σειρές. Οι χρονολογικές σειρές που κατασκευάζονται είναι τέσσερις και αφορούν τα πακέτα και τις ψηφιοσυλλαβές της συνολικής κίνησης καθώς και της κίνησης του WEB (προσδιορίζεται από το well known port 80).

Από το αρχείο δεδομένων εξωτερικής κίνησης του FlowCollector προκύπτει ότι σε κάθε δευτερόλεπτο είτε αντιστοιχεί μια (ή περισσότερες) γραμμές ροών είτε δεν αντιστοιχεί καμία γραμμή ροής. Η τελευταία περίπτωση συμβαίνει όταν κατά τη διάρκεια του δευτερολέπτου δεν εντοπίστηκε κίνηση και συνεπώς στο αρχείο δεδομένων δεν αναγράφονται ούτε στοιχεία ροών κίνησης γι' αυτό το δευτερόλεπτο ούτε η σφραγίδα χρόνου γι' αυτό το δευτερόλεπτο.

Έχοντας αυτά υπόψη, η βασική ιδέα υλοποίησης της μακροεντολής εξαγωγής χρονοσειρών είναι η εξής: Κάθε παρατήρηση των χρονολογικών σειρών που δημιουργούνται αντιστοιχεί σε διάστημα χρόνου ενός δευτερολέπτου. Σε περίπτωση που για το συγκεκριμένο δευτερόλεπτο υπάρχουν στοιχεία ροών τότε αθροίζονται τα πακέτα και οι ψηφιοσυλλαβές όλων των ροών που αντιστοιχούν σ' αυτό. Ξεχωριστά αθροίζονται τα πακέτα και οι ψηφιοσυλλαβές εκείνων των ροών για τις οποίες το port πηγής ισούται με 80 (WEB κίνηση). Κάθε ένα από τα αθροίσματα των πακέτων και των ψηφιοσυλλαβών της συνολικής και της WEB κίνησης αποτελούν μια παρατήρηση για κάθε μια από τις αντίστοιχες χρονολογικές σειρές των πακέτων και των ψηφιοσυλλαβών της συνολικής και της WEB κίνησης που κατασκευάζονται. Σε περίπτωση που για κάποιο δευτερόλεπτο δεν υπάρχουν στοιχεία ροών, αυτό σημαίνει ότι δεν εντοπίστηκε κίνηση κατά τη διάρκεια του, κατά τη συνέπεια οι παρατηρήσεις των τεσσάρων χρονολογικών σειρών που του αντιστοιχούν παίρνουν την τιμή μηδέν. Σημειώνεται ότι ο κώδικας της μακροεντολής παραθέτεται στο Παράτημα 4.

## 6. ΔΕΔΟΜΕΝΑ ΕΞΩΤΕΡΙΚΗΣ ΚΙΝΗΣΗΣ ΕΙΣΟΔΟΥ ΤΟΥ Ο.Π.Α.

### 6.1 ΔΙΕΞΑΓΩΓΗ ΚΑΙ ΕΚΤΙΜΗΣΕΙΣ 1<sup>ΗΣ</sup> ΜΕΤΡΗΣΗΣ

Τρόπος συλλογής μετρήσεων:	NETFLOW/FLOWCOLLECTOR
Ημ/νία συλλογής:	4/1/99
Ρυθμός συλλογής (polling ανά):	1 sec
Επίπεδα άθροισης (m) εκτιμήσεων:	400, 800, 1200, ...

Για την καλύτερη επεξεργασία των δεδομένων κίνησης της 1ης μέτρησης, αυτά διασπάστηκαν σε έξι υπό-σύνολα δεδομένων. Ο παρακάτω Πίνακας περιλαμβάνει τις εκτιμήσεις της παραμέτρου  $H$  (βαθμός αυτό-ομοιότητας) για κάθε ένα από τα έξι σύνολα δεδομένων κίνησης και για κάθε μια από τις δύο χρησιμοποιούμενες μεθόδους εκτίμησης της παραμέτρου  $H$  (*Αθροιστικών Χρονολογικών Σειρών* και *R/S Στατιστικής*). Επιπλέον, για κάθε ένα από τα υπό-σύνολα δεδομένων εκτιμάται η παράμετρος  $H$  της WEB κίνησης που αντιστοιχεί σε κάθε υπο-σύνολο δεδομένων.

ΧΡΟΝΙΚΗ ΔΙΑΡΚΕΙΑ ΣΥΛΛΟΓΗΣ	ΠΛΗΘΟΣ ΠΑΡΑΤΗΡΗΣΕΩΝ	Octets		Packets	
		1. ΑΦΡ/ΚΩΝ ΧΡ/ΚΩΝ ΣΕΙΡΩΝ	2. R/S ΣΤΑΤΙΣΤΙΚΗΣ	3. ΑΦΡ/ΚΩΝ ΧΡ/ΚΩΝ ΣΕΙΡΩΝ	4. R/S ΣΤΑΤΙΣΤΙΚΗΣ
1. 17:15:41 – 19:00:14	6273	0.384	0.548	0.3	0.564
	Web κίνηση	0.535	0.534	0.52	0.562
2. 19:00:01 - 22:29:59	12598	0.432	0.511	0.466	0.543
	Web κίνηση	0.662	0.59	0.644	0.594
3. 22:30:08 – 01:59:59	19793	0.379	0.496	0.484	0.586
	Web κίνηση	0.479	0.601	0.433	0.614
4. 02:00:00 – 09:29:59	27180	0.486	0.582	0.565	0.582
	Web κίνηση	0.658	0.577	0.653	0.563
5. 09:30:00 – 11:59:59	9060	0.598	0.639	0.591	0.634
	Web κίνηση	0.603	0.692	0.603	0.665
6. 12:00:00 – 14:02:00	7284	0.392	0.557	0.358	0.563
	Web κίνηση	0.632	0.64	0.578	0.639

Πίνακας 1 Οι εκτιμήσεις της  $H$  των χρονολογικών σειρών πακέτων και ψηφιοσυλλαβών εισόδου της εξωτερικής κίνησης του δικτύου δεδομένων του Ο.Π.Α.

## 6.2 ΔΙΕΞΑΓΩΓΗ ΚΑΙ ΕΚΤΙΜΗΣΕΙΣ 2<sup>ΗΣ</sup> ΜΕΤΡΗΣΗΣ

Τρόπος συλλογής μετρήσεων:	NETFLOW/FLOWCOLLECTOR
Ημ/νία συλλογής:	7/1/99
Ρυθμός συλλογής (polling ανά):	1 sec
Επίπεδα άθροισης (m) εκτιμήσεων:	800, 1600, 2400, ...

Ο πίνακας που ακολουθεί περιλαμβάνει τις εκτιμήσεις της παραμέτρου  $H$  (βαθμός αυτό-ομοιότητας) για το σύνολο των δεδομένων κίνησης που προέκυψε από τη 2<sup>η</sup> μέτρηση χρησιμοποιώντας κάθε μια από τις δύο μεθόδους εκτίμησης της παραμέτρου  $H$  (Αθροιστικών Χρονολογικών Σειρών και R/S Στατιστικής). Επιπλέον, εκτιμάται η παράμετρος  $H$  της Web κίνησης αντιστοιχεί σ' αυτό το δείγμα της κίνησης.

ΧΡΟΝΙΚΗ ΔΙΑΡΚΕΙΑ ΣΥΛΛΟΓΗΣ	ΠΛΗΘΟΣ ΠΑΡΑΤΗΡΗΣΕΩΝ	Octets		Packets	
		1. ΑΘΡ/ΚΩΝ ΧΡ/ΚΩΝ ΣΕΙΡΩΝ	2. R/S ΣΤΑΤΙΣΤΙΚΗΣ	3. ΑΘΡ/ΚΩΝ ΧΡ/ΚΩΝ ΣΕΙΡΩΝ	4. R/S ΣΤΑΤΙΣΤΙΚΗΣ
15:13:18 – 19:38:14	8672	0.433	0.589	0.48	0.583
	Web κίνηση	0.638	0.601	0.68	0.638

Πίνακας 2 Οι εκτιμήσεις της  $H$  των χρονολογικών σειρών πακέτων και ψηφιοσύλλαβών εισόδου της εξωτερικής κίνησης του δικτύου δεδομένων του Ο.Π.Α..

## 6.3 ΔΙΕΞΑΓΩΓΗ ΚΑΙ ΕΚΤΙΜΗΣΕΙΣ 3<sup>ΗΣ</sup> ΜΕΤΡΗΣΗΣ

Τρόπος συλλογής μετρήσεων:	NETFLOW/FLOWCOLLECTOR
Ημ/νία συλλογής:	8/1/99
Ρυθμός συλλογής (polling ανά):	1 sec
Επίπεδα άθροισης (m) εκτιμήσεων:	400, 800, 1200, 2400, ...

Ο πίνακας που ακολουθεί περιλαμβάνει τις εκτιμήσεις της παραμέτρου  $H$  (βαθμός αυτό-ομοιότητας) για το σύνολο των δεδομένων κίνησης που προέκυψε από τη 3<sup>η</sup> μέτρηση χρησιμοποιώντας κάθε μια από τις δύο μεθόδους εκτίμησης της παραμέτρου  $H$  (Αθροιστικών Χρονολογικών Σειρών και R/S Στατιστικής). Επιπλέον, εκτιμάται η παράμετρος  $H$  της Web κίνησης αντιστοιχεί σ' αυτό το δείγμα της κίνησης.

		<i>Octets</i>		<i>Packets</i>	
ΧΡΟΝΙΚΗ ΔΙΑΡΚΕΙΑ ΣΥΛΛΟΓΗΣ	ΠΛΗΘΟΣ ΠΑΡΑΤΗΡΗΣΕΩΝ	1. <i>ΑΘΡ/ΚΩΝ ΧΡ/ΚΩΝ ΣΕΙΡΩΝ</i>	2. <i>R/S ΣΤΑΤΙΣΤΙΚΗΣ</i>	3. <i>ΑΘΡ/ΚΩΝ ΧΡ/ΚΩΝ ΣΕΙΡΩΝ</i>	4. <i>R/S ΣΤΑΤΙΣΤΙΚΗΣ</i>
15:13:18 – 16:33:50	5880	0.325	0.518	0.489	0.579
	Web κίνηση	0.537	0.56	0.551	0.592

Πίνακας 2 Οι εκτιμήσεις της  $H$  των χρονολογικών σειρών πακέτων και ψηφιοσυλλαβών εισόδου της εξωτερικής κίνησης του δικτύου δεδομένων του Ο.Π.Α.

#### 6.4 ΣΧΟΛΙΑΣΜΟΣ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΜΕΤΡΗΣΕΩΝ

##### A. ΣΥΓΚΡΙΣΗ MAIL ΚΑΙ ΣΥΝΟΛΙΚΗΣ ΚΙΝΗΣΗΣ ΩΣ ΠΡΟΣ ΤΟΝ ΒΑΘΜΟ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ

Από τους παραπάνω πίνακες και για το σύνολο των μετρήσεων προκύπτει ότι ο βαθμός αυτό-ομοιότητας της Web κίνησης είναι μεγαλύτερος από τον βαθμό αυτό-ομοιότητας που αντιστοιχεί στη συνολική κίνηση.

##### B. ΔΙΑΦΟΡΑ ΣΤΙΣ ΕΚΤΙΜΗΣΕΙΣ ΤΗΣ $H$ ΤΩΝ ΔΥΟ ΜΕΘΟΔΩΝ ΕΚΤΙΜΗΣΗΣ

Υπενθυμίζεται ότι η ύπαρξη διαφοράς μεταξύ των εκτιμήσεων της  $H$  που προκύπτουν με την μέθοδο των Αθροιστικών Χρονολογικών Σειρών και της R/S Στατιστικής είναι αναμενόμενη αφού οι δύο μέθοδοι χρησιμοποιούν διαφορετικές διαδικασίες προκειμένου να εκτιμήσουν την ίδια παράμετρο. Ωστόσο, συγκρίνοντας τις στήλες (1,2) και (3,4) των παραπάνω πινάκων οι οποίες αφορούν τις δύο μεθόδους εκτιμήσης της  $H$ , προκύπτουν τα εξής:

- Σε απόλυτες τιμές η διαφορά στις εκτιμήσεις είναι σχετικά μεγάλη, όμως;
- Στις περιπτώσεις που διαφορά είναι σημαντική τότε ως πιο έγκυρη εκτίμηση πρέπει να θεωρείται αυτή που προκύπτει από την μέθοδο της R/S στατιστικής, αφού για τον υπολογισμό της λαμβάνονται υπόψη περισσότερα επίπεδα άθροισης.
- Οι εκτιμήσεις μιας μεθόδου εκτίμησης δεν εμφανίζονται εν γένει μικρότερες (ή μεγαλύτερες) από αυτές της άλλης μεθόδου εκτίμησης. Με άλλα λόγια, δεν προκύπτει μια πάγια ανισοτική σχέση μεταξύ των εκτιμήσεων η οποία να προσδιορίζεται από την μέθοδο εκτίμησης της  $H$ .

##### Γ. ΔΙΑΦΟΡΑ ΣΤΙΣ ΕΚΤΙΜΗΣΕΙΣ ΤΗΣ $H$ ΜΕΤΑΞΥ ΤΗΣ ΚΙΝΗΣΗΣ ΠΑΚΕΤΩΝ & ΨΗΦΙΟΣΥΛΛΑΒΩΝ

Η διαφορά στις εκτιμήσεις μεταξύ των χρονολογικών σειρών της κίνησης η οποία μετράται σε πακέτα και των χρονολογικών σειρών της ίδιας κίνησης η οποία μετράται σε ψηφιοσυλλαβές (bytes) οφείλεται στο γεγονός ότι τα πακέτα δεν

αποτελούνται από σταθερό αριθμό ψηφιοσυλλαβών. Συγκρίνοντας τις στήλες (1,3) ή (2,4) στους παραπάνω πίνακες, δηλαδή τις στήλες που αφορούν ψηφιοσυλλαβές με τις στήλες που αφορούν πακέτα και οι οποίες προκύπτουν από την ίδια μέθοδο εκτίμησης, προκύπτουν τα εξής:

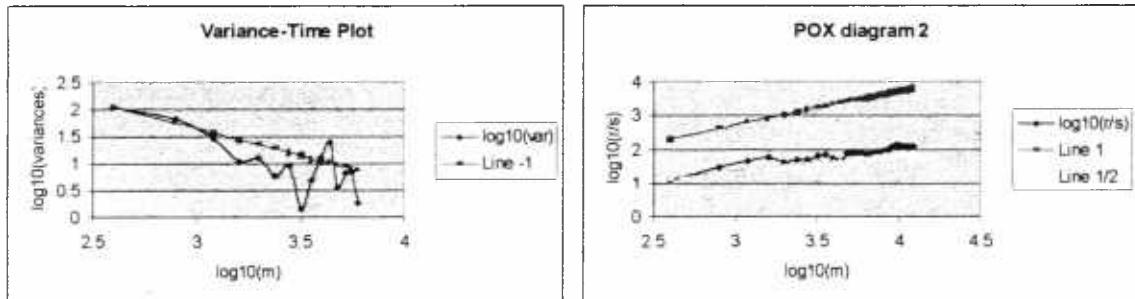
- Η διαφορά στις εκτιμήσεις είναι σχετικά μικρή.
- Οι εκτιμήσεις των χρονολογικών σειρών που αφορούν πακέτα δεν εμφανίζονται εν γένει μικρότερες (ή μεγαλύτερες) από αυτές που αφορούν ψηφιοσυλλαβές. Με άλλα λόγια, δεν προκύπτει μια πάγια ανισοτική σχέση μεταξύ των εκτιμήσεων η οποία να προσδιορίζεται από τον τρόπο μέτρησης της κίνησης (σε πακέτα ή ψηφιοσυλλαβές).
- Το γεγονός της ύπαρξης διαφοράς εκτιμήσεων δεν δημιουργεί προβλήματα «συγκρούσεων»: Όταν η εκτίμηση μιας συγκεκριμένης μεθόδου υποδεικνύει μεγάλο (ή μικρό) βαθμό αυτό-ομοιότητας για μια χρονολογική σειρά πακέτων, τότε αναμένεται μεγάλος (ή μικρός) βαθμός αυτό-ομοιότητας και για την αντίστοιχη χρονολογική σειρά η οποία είναι μετρημένη σε ψηφιοσυλλαβές.

#### Δ. ΕΚΤΙΜΗΣΗ ΤΟΥ ΒΑΘΜΟΥ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ ΤΗΣ ΚΙΝΗΣΗΣ ΤΟΥ ΔΙΚΤΥΟΥ

Από τους παραπάνω πίνακες προκύπτει ότι οι εκτιμήσεις του βαθμού αυτό-ομοιότητας της εξωτερικής κίνησης εισόδου δεν παρέχουν μια σαφή ένδειξη για την ύπαρξη αυτό-ομοιότητας αφού ο βαθμός της τελευταίας για το σύνολο των περιπτώσεων κυμαίνεται γύρω στο 0.5. Παρακάτω παραθέτονται διαγράμματα ενδεικτικά του βαθμού αυτό-ομοιότητας του 2<sup>ου</sup> και 6<sup>ου</sup> συνόλου της 1<sup>ης</sup> μέτρησης τα οποία εξετάστηκαν παραπάνω.

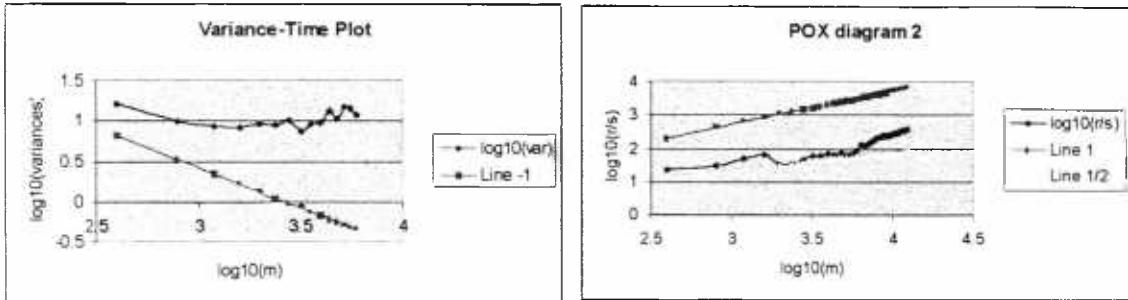
#### 2Ο ΣΥΝΟΛΟ ΔΕΔΟΜΕΝΩΝ ΕΞΩΤΕΡΙΚΗΣ ΚΙΝΗΣΗΣ 1<sup>ΗΣ</sup> ΜΕΤΡΗΣΗΣ (ΣΥΝΟΛΙΚΗ ΚΙΝΗΣΗ - ΠΑΚΕΤΑ)

Τα *Variance-Time* και *POX* διαγράμματα που ακολουθούν δεν παρέχουν ένδειξη για ύπαρξη αυτό-ομοιότητας αφού στο *Variance-Time* διάγραμμα οι λογαριθμημένες διακυμάνσεις βρίσκονται στην πλειοψηφία τους κάτω από τη γραμμή με κλίση -1, ενώ στο *POX* διάγραμμα οι λογαριθμημένες τιμές των r/s βρίσκονται κοντά στη γραμμή με κλίση  $\frac{1}{2}$ .



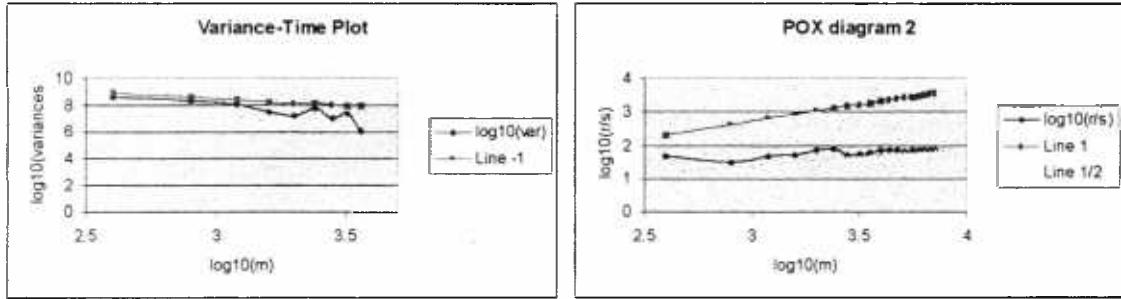
## 2Ο ΣΥΝΟΛΟ ΔΕΔΟΜΕΝΩΝ ΕΞΩΤΕΡΙΚΗΣ ΚΙΝΗΣΗΣ 1<sup>ΗΣ</sup> ΜΕΤΡΗΣΗΣ (WEB ΚΙΝΗΣΗ - ΠΑΚΕΤΑ)

Συγκρίνοντας τα δύο προηγούμενα και τα δύο επόμενα διαγράμματα προκύπτει ότι ο βαθμός αυτό-ομοιότητας της Web κίνησης εμφανίζεται μεγαλύτερος σε σχέση με αυτόν της αντίστοιχης συνολικής κίνησης.



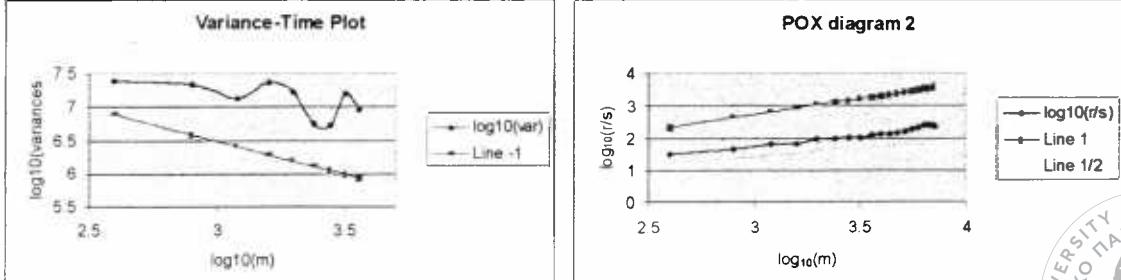
## 6Ο ΣΥΝΟΛΟ ΔΕΔΟΜΕΝΩΝ ΕΞΩΤΕΡΙΚΗΣ ΚΙΝΗΣΗΣ 1<sup>ΗΣ</sup> ΜΕΤΡΗΣΗΣ (ΣΥΝΟΛΙΚΗ ΚΙΝΗΣΗ - BYTES)

Το παρακάτω *Variance-Time* διάγραμμα δείχνει ανυπαρξία αυτό-ομοιότητας, αφού το σύνολο των λογαριθμημένων διακυμάνσεων βρίσκεται κάτω από τη γραμμή με κλίση  $-1$ . Στο POX διάγραμμα οι λογαριθμημένες τιμές των  $r/s$  βρίσκονται πολύ κοντά στην γραμμή με κλίση  $1$ . Κατά συνέπεια δεν προκύπτει ένδειξη για ύπαρξη αυτό-ομοιότητας στα δεδομένα.



## 6Ο ΣΥΝΟΛΟ ΔΕΔΟΜΕΝΩΝ ΕΞΩΤΕΡΙΚΗΣ ΚΙΝΗΣΗΣ 1<sup>ΗΣ</sup> ΜΕΤΡΗΣΗΣ (WEB ΚΙΝΗΣΗ - BYTES)

Τα παρακάτω διαγράμματα δίνουν μια σαφή ένδειξη για την ύπαρξη αυτό-ομοιότητας στα δεδομένα της Web κίνησης. Στο *Variance-Time* διάγραμμα οι λογαριθμημένες διακυμάνσεις βρίσκονται σε μεγάλη απόσταση από τη γραμμή με κλίση  $-1$ , ενώ στο POX διάγραμμα οι λογαριθμημένες τιμές των  $r/s$  βρίσκονται σε αρκετή απόσταση από τη γραμμή με κλίση  $1/2$ .



## ΚΕΦΑΛΑΙΟ 8

### ΣΥΜΠΕΡΑΣΜΑΤΑ & ΠΡΟΤΑΣΕΙΣ

#### 1. ΣΥΜΠΕΡΑΣΜΑΤΑ ΕΡΕΥΝΑΣ

Στα πλαίσια της προσπάθειας χαρακτηρισμού της κίνησης του δικτύου δεδομένων του Πανεπιστημίου σε σχέση με το φαινόμενο της αυτό-ομοιότητας, συλλέχθηκαν πραγματικά δεδομένα κίνησης από τοπικά δίκτυα του Πανεπιστημίου καθώς και δεδομένα εξωτερικής κίνησης εισόδου από το δρομολογητή που συνδέει το δίκτυο δεδομένων του Πανεπιστημίου με το διαδίκτυο.

Τα δεδομένα κίνησης εξετάστηκαν με τη μορφή χρονολογικών σειρών ως προς τον βαθμό αυτό-ομοιότητας που παρουσιάζουν. Όσον αφορά την Ethernet κίνηση βάσει των εκτιμήσεων του βαθμού αυτό-ομοιότητας προέκυψε σαφής ένδειξη για την ύπαρξη αυτό-ομοιότητας στα δεδομένα. Μάλιστα παρατηρήθηκε ότι το χαρακτηριστικό της αυτό-ομοιότητας γινόταν περισσότερο έντονο καθώς αυξανόταν ο βαθμός χρήσης του δικτύου. Από το τελευταίο συμπέρασμα επιβεβαιώνεται μια πρόταση η οποία έχει ήδη ανακαλυφθεί από προηγούμενες σχετικές έρευνες και έχει διατυπωθεί στη βιβλιογραφία. Σε αντίθεση, ο βαθμός αυτό-ομοιότητας της εξωτερικής κίνησης εισόδου, δηλαδή της κίνησης η οποία εισέρχεται από τον «έξω κόσμο» στο δίκτυο δεδομένων του Πανεπιστημίου εκτιμήθηκε σε χαμηλά επίπεδα. Το γεγονός αυτό ίσως οδηγεί στο συμπέρασμα ότι η Ethernet κίνηση παρουσιάζεται περισσότερο ισχυρά αυτό-όμοια σε σχέση με την WAN κίνηση.

Επίσης, παρατηρήθηκε ότι δεν υπάρχει σημαντική διαφορά όσον αφορά τον βαθμό αυτό-ομοιότητας μεταξύ της κίνησης πακέτων και ψηφιοσυλλαβών. Η διαφορά που εντοπίζεται οφείλεται στο γεγονός ότι τα πακέτα δεν αποτελούνται από σταθερό αριθμό ψηφιοσυλλαβών. Παρ' όλα αυτά, το συμπέρασμα που προκύπτει είναι ότι ο βαθμός αυτό-ομοιότητας μιας χρονολογικής σειράς δεδομένων κίνησης μετρημένης σε πακέτα απέχει ελάχιστα από τον βαθμό αυτό-ομοιότητας της ίδιας χρονολογικής σειράς δεδομένων κίνησης μετρημένης σε ψηφιοσυλλαβές.

Όσον αφορά τις δύο μεθόδους (αθροιστικών χρονολογικών σειρών και R/S στατιστικής) εκτίμησης του βαθμού αυτό-ομοιότητας πρέπει να σημειωθεί ότι αποτελούν ευρεστικές μεθόδους και χρησιμοποιούνται με σκοπό να παρέχουν μια ένδειξη για την ύπαρξη ή όχι αυτό-ομοιότητας στα δεδομένα και όχι προκειμένου να δώσουν μια στατιστικά ακριβή εκτίμηση του βαθμού αυτό-ομοιότητας. «Συγκρίνοντας» τα αποτελέσματα των δύο μεθόδων για κάθε μια από τις

χρονολογικές σειρές δεδομένων που εκτιμήθηκαν, προέκυψε ότι η ύπαρξη διαφοράς στις εκτιμήσεις τους δεν δημιουργεί προβλήματα «συγκρούσεων». Όταν η εκτίμηση της μιας μεθόδου υποδεικνύει μεγάλο (ή μικρό) βαθμό αυτό-ομοιότητας, τότε αναμένεται αντίστοιχος βαθμός αυτό-ομοιότητας και για την εκτίμηση της άλλης μεθόδου.

## 2. ΠΡΟΤΑΣΕΙΣ ΜΕΛΛΟΝΤΙΚΗΣ ΕΡΕΥΝΑΣ

Προκειμένου να εξαχθούν ασφαλέστερα συμπεράσματα σχετικά με τον χαρακτηρισμό της εξωτερικής κίνησης και της κίνησης των τοπικών δικτύων του δικτύου δεδομένων του Πανεπιστημίου τα οποία θα είναι ικανά γενικευτούν και να συμβάλλουν στην καλύτερη επίλυση ζητημάτων τα οποία σχετίζονται με την απόδοση (performance) και τον σχεδιασμό (planning) δικτύων υπολογιστών, προτείνονται τα εξής:

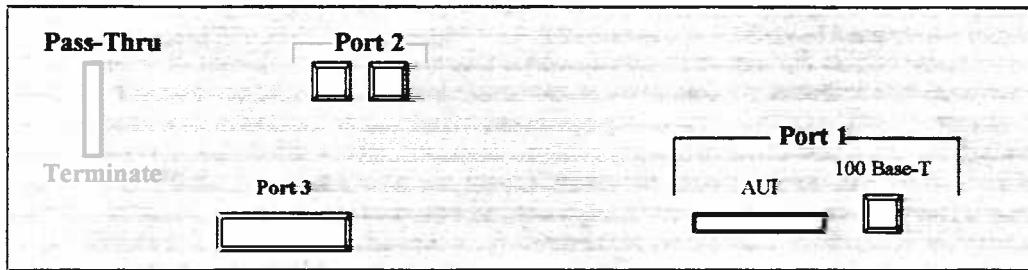
Κατ' αρχήν θα πρέπει να διερευνηθούν οι συνθήκες κάτω από τις οποίες η αυτό-ομοιότητα πρέπει οπωσδήποτε να λαμβάνεται υπόψη σε θέματα απόδοσης και σχεδιασμού. Το θέμα αυτό αποτελεί τρέχον αντικείμενο έρευνας, ενώ τα αποτελέσματα σχετικών ερευνών παρουσιάζονται ήδη στη βιβλιογραφία.

Επίσης, για την καλύτερη κατανόηση των χαρακτηριστικών της κίνησης δικτύων δεδομένων, προτείνεται η κατάτμησή της σε επιμέρους σύνολα κίνησης. Η κατάτμηση θα μπορούσε να πραγματοποιηθεί βάσει πηγών προορισμού, τύπο υπηρεσίας και χρησιμοποιούμενο πρωτόκολλο. Έτσι, εξετάζοντας για την ύπαρξη αυτό-ομοιότητας σε κάθε μια από τις επιμέρους συνιστώσες της συνολικής κίνησης θα ήταν ενδιαφέρον να παρακολουθήσουμε τον τρόπο με τον οποίο αυτές συνδυάζονται καθώς και το μέγεθος της «συνεισφοράς» τους στον σχηματισμό αυτό-ομοιότητας της συνολικής κίνησης.

## ΠΑΡΑΡΤΗΜΑ 1

### ΛΕΙΤΟΥΡΓΙΚΗ ΠΕΡΙΓΡΑΦΗ ΤΗΣ ΣΥΣΚΕΥΗΣ HP J3458A FAST ETHERNET LANPROBE

Η συσκευή LANprobe J3458A της Hewlett Packard χρησιμοποιήθηκε για τη συλλογή μέρους των μετρήσεων της κίνησης του δικτύου δεδομένων του Πανεπιστημίου. Στο Σχήμα 1 παρουσιάζονται οι βασικές θύρες της συσκευής.



Σχήμα 1 Βασικές θύρες της συσκευής LANprobe J3458A της Hewlett Packard.

Η ρύθμιση (configuration) του LANprobe πραγματοποιείται διασυνδέοντας ένα τερματικό με το LANprobe και χρησιμοποιώντας το κυρίως μενού που εμφανίζεται. Οι τιμές των παραμέτρων που τίθενται κατά τη διαδικασία της ρύθμισης της συσκευής μέσω του κυρίως μενού, εξαρτώνται από τον τρόπο χρήσης της συσκευής. Οι επιλογές που διατίθενται είναι οι εξής:

## 1. NO OPTIONAL INTERFACE.

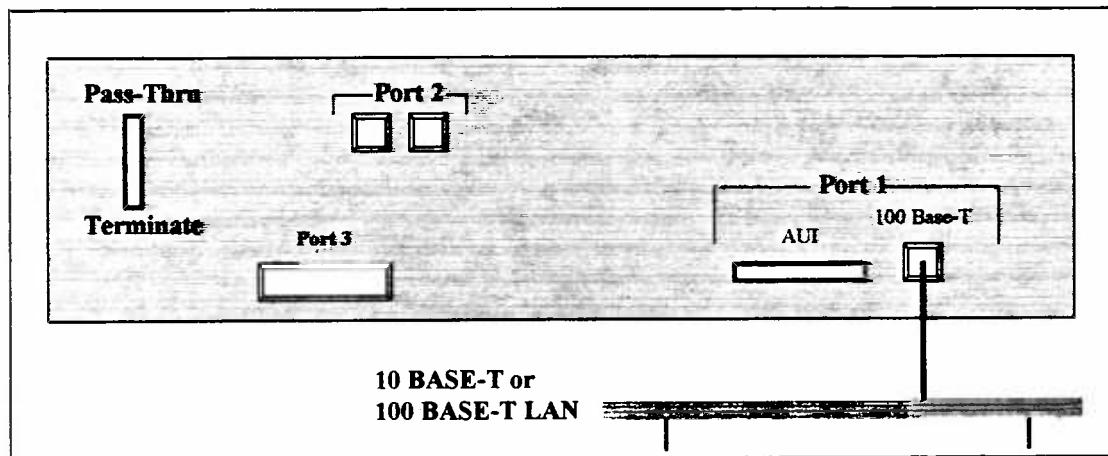
Η θύρα 1 (port 1) της LANprobe συσκευής συνδέεται με το Ethernet segment, προκειμένου να συλλεχθούν τα στοιχεία της εσωτερικής του κίνησης. Από αυτή τη θύρα, κατά τη διαδικασία ρύθμισης (configuration) του LANprobe, τίθενται τα στοιχεία *IP address*, *Default gateway IP address*, καθώς και *Subnet mask*. Οι διαφορετικές υλοποιήσεις της θύρας 1 (AUI, 100 Base-T), αφορούν τη σύνδεση του LANprobe σε 10 MB/s ή 10Base-T/100Base-TX Ethernet δίκτυα.

### 1.1 ΣΥΝΔΕΣΗ ΤΟΥ LANPROBE ΜΕ 10MB/S ETHERNET ΔΙΚΤΥΑ.

Η AUI θύρα του LANprobe συνδέεται με το δίκτυο. Η παράμετρος Physical Connector της επιλογής Modify/View Interface Values του κυρίως μενού τίθεται ως AUI.

### 1.2 ΣΥΝΔΕΣΗ ΤΟΥ LANPROBE ΜΕ 10BASE-T/100BASE-TX ΔΙΚΤΥΑ.

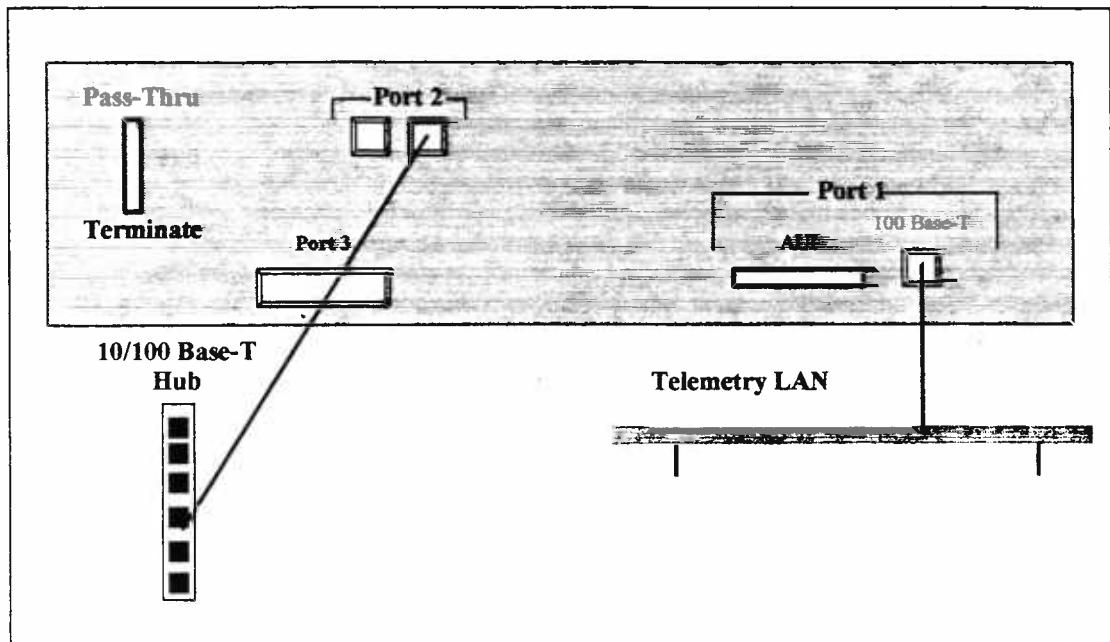
Η 10Base-T/100Base-TX (RJ-45) θύρα του LANprobe συνδέεται με ένα 10Base-T ή 100Base-TX δίκτυο. Η παράμετρος Physical Connector της επιλογής Modify/View Interface Values του κυρίως μενού τίθεται ως RJ-45.



Σχήμα 2 Σύνδεση του LANprobe με 10 Base-T ή 100 Base-T δίκτυο.

## 2. CONNECTING 10BASE-T/100BASE-TX HALF-DUPLEX OPTION 200 AND 201.

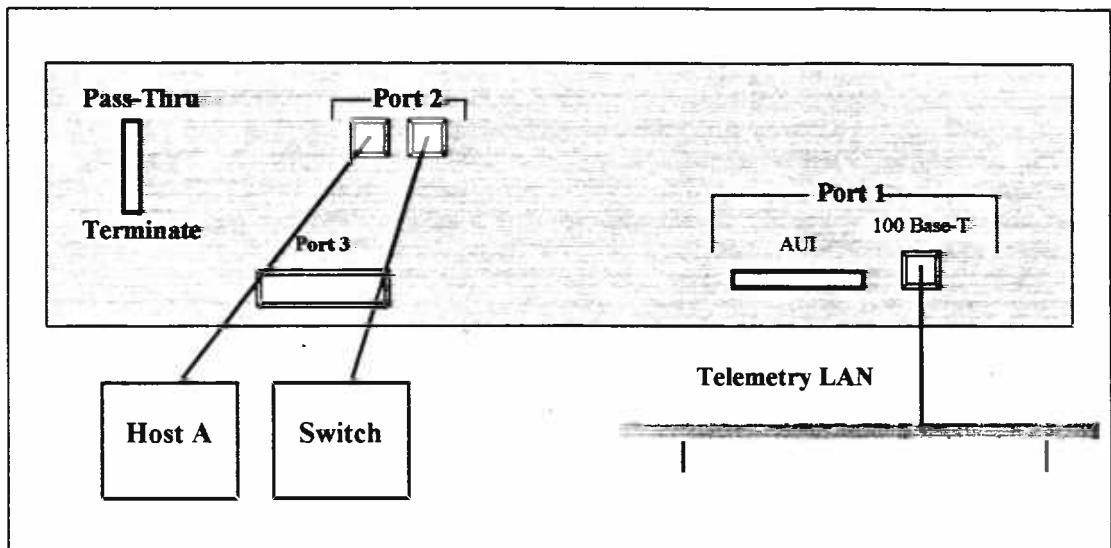
**2.1 HALF-DUPLEX ΣΥΝΔΕΣΗ ΜΕΤΑΞΥ 10BASE-T/100BASE-TX ΔΙΚΤΥΟΥ ΚΑΙ HUB.**  
Η 10Base-T/100Base-TX (RJ-45) θύρα του LANprobe συνδέεται με ένα 10Base-T ή 100Base-TX δίκτυο. Μια από τις δύο θύρες RJ-45 (Option 200 ή 201), συνδέεται με ένα 10Base-T ή 100Base-TX Ethernet Hub. Η ένδειξη *Pass-Thru / Terminate* του LANprobe τίθεται ως *Terminate*. Με αυτόν τον τρόπο ελέγχεται η κίνηση μεταξύ των υπολογιστικών συσκευών που είναι συνδεδεμένες με το Hub.



Σχήμα 3 Σύνδεση του LANprobe σε Hub και 10 Base-T ή 100 Base-T δίκτυο.

## 2.2 HALF-DUPLEX ΣΥΝΔΕΣΗ ΜΕΤΑΞΥ HOST ΚΑΙ ΜΕΤΑΓΩΓΕΑ.

Η 10Base-T/100Base-TX (RJ-45) θύρα του LANprobe συνδέεται με ένα 10Base-T ή 100Base-TX δίκτυο. Μια από τις δύο θύρες RJ-45 (Option 200 ή 201), συνδέεται στον κόμβο, ενώ η άλλη στη θύρα του μεταγωγέα. Η ένδειξη *Pass-Thru / Terminate* του LANprobe τίθεται ως *Pass-Thru*. Με αυτόν τον τρόπο ελέγχεται η κίνηση μεταξύ του Host A και του μεταγωγέα.



Σχήμα 3 Σύνδεση του LANprobe μεταξύ Host και μεταγωγέα.

## 3. CONNECTING 100BASE-TX/100BASE-FX OPTION 202.

Με βάση την επιλογή 202 (Option 202) του LANprobe επιτυγχάνεται η συλλογή στοιχείων επικοινωνίας μεταξύ ενός Host A και ενός μεταγωγέα. Η διαφορά με την επιλογή 201 (Option 201) έγκειται στο ότι το LANprobe συνδέεται με 100Base-TX/100Base-FX δίκτυα.

## ΠΑΡΑΡΤΗΜΑ 2

# Ο ΚΩΔΙΚΑΣ ΤΗΣ ΥΛΟΠΟΙΗΣΗΣ ΤΩΝ ΜΕΘΟΔΩΝ ΕΚΤΙΜΗΣΗΣ ΤΟΥ ΒΑΘΜΟΥ ΑΥΤΟ-ΟΜΟΙΟΤΗΤΑΣ ΤΩΝ *ΑΘΡΟΙΣΤΙΚΩΝ ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ & ΤΗΣ* *R/S ΣΤΑΤΙΣΤΙΚΗΣ*

### 'AggregatedSeries()

Μορφοποιεί το στυλ εμφάνισης των επικεφαλίδων και καλεί την Aggregated\_Series\_I() περνώντας της ως παράμετρο το πλήθος της υπό ανάλυσης χρονολογικής σειράς, η οποία θεωρείται ότι βρίσκεται στην 1η και 2η θέση του τρέχοντος φύλλου εργασίας.

### Sub AggregatedSeries()

N = FindN()

Range("Q1:HZ" & N + 15).ClearContents

Range("A1:HZ1").WrapText = False

Range("A1:HZ1").Font.Size = 12

Range("A1:HZ1").Font.Bold = False

Range("A1:HZ1").Font.Color = RGB(0, 0, 255)

Range("A1:HZ1").Font.Name = "Times New Roman"

A\_S\_I(N)

End Sub

### 'RS\_Statistic()

Μορφοποιεί το στύλ εμφάνισης των επικεφαλίδων και καλεί την R\_S\_I()  
περνώντας της ως παράμετρο το πλήθος της υπό ανάλυσης χρονολογικής  
σειράς, η οποία θεωρείται ότι βρίσκεται στην 1η και 2η θέση του  
τρέχοντος φύλλου εργασίας.

### Sub RS\_Statistic()

N = FindN()

Range("Q1:HZ" & N + 15).ClearContents

Range("A1:HZ1").WrapText = False

Range("A1:HZ1").Font.Size = 12

Range("A1:HZ1").Font.Bold = False

```
Range("A1:HZ1").Font.Color = RGB(0, 0, 255)
Range("A1:HZ1").Font.Name = "Times New Roman"
R_S_I(N)
End Sub
```

### 'R\_S\_I()

'Δέχεται ως όρισμα το πλήθος N των στοιχείων της χρονολογικής σειράς η οποία  
'θεωρείται ότι βρίσκεται στην 1η και 2η στήλη του τρέχοντος φύλλου εργασίας. Για  
'κάθε επίπεδο άθροιστης Ni δημιουργεί k νέες χρονολογικές σειρές, όπου k=N/Ni. Για  
'κάθε μία από τις k χρονολογικές σειρές και για κάθε επίπεδο άθροιστης Ni, υπολογίζει  
'την τιμή R/S. Εμφανίζει τα σημεία των διαγραμμάτων POX τύπου 1 και 2, την  
'εκτίμηση της παραμέτρου H που προκύπτει για κάθε επίπεδο άθροιστης Ni, καθώς  
'και την τελική εκτίμηση της H.

### Sub R\_S\_I(N)

lin0 = 1

'αριθμός στήλης εμφάνισης των χρονολογικών σειρών που θα δημιουργηθούν  
col0 = 17

'αριθμός στήλης στην οποία αναμένεται η υπό ανάλυση χρονολογική σειρά  
colt = 2

Pli = 400

Ni = Pli 'Ni = επίπεδο άθροιστης

seg = 1 'αρχικοποίηση του αριθμού των X.S. που θα δημιουργηθούν για το σύνολο  
των επιπέδων άθροιστης

tim = 1 'αρχικοποίηση του αριθμού των επιπέδων άθροιστης που θα χρησιμοποιηθούν  
'εμφάνισε τις επικεφαλίδες

Cells(lin0, colt - 1).Value = "Time Series"

Cells(lin0, colt + 3).Value = "R/S Est.H"

Cells(lin0, colt + 9).Value = " POX"

Cells(lin0, colt + 10).Value = "Diagram1"

Cells(lin0, colt + 11).Value = " Line 1"

Cells(lin0, colt + 12).Value = "Line 1/2"

Cells(lin0, colt + 13).Value = " POX"

Cells(lin0, colt + 14).Value = "Diagram2"

lin = 0

MeanHEst = 0

'όσο το επίπεδο άθροιστης δεν υπερβαίνει το πλήθος των στοιχείων της X.S.

Do While Ni <= N

lin = lin + k

k = div(N, Ni)

For i = 1 To k 'για το επίπεδο άθροιστης Ni δημιουργούνται k X.S.

Cells(lin0, col0 + seg + i).Value = " " & (i - 1) \* Ni + 1 & " έως " & (i - 1) \* Ni + Ni  
& " "

cellval = 0

'υπολόγισε τον μέσο όρο των Ni επόμενων στοιχείων της αρχικής X.S.

aver = mean(colt, lin0 + (i - 1) \* Ni + 1, lin0 + (i - 1) \* Ni + Ni)

For j = 1 To Ni 'υπολόγισε τα στοιχεία της i X.S.

cellval = cellval + Cells(lin0 + j + (i - 1) \* Ni, colt).Value

Cells(lin0 + j, col0 + seg + i) = cellval - j \* aver

**Next j**

'βρες την R/S τιμή της i X.Σ.

RSVal = RS(col0 + seg + i, lin0 + 1, lin0 + Ni)

RSValMean = RSValMean + RSVal

'βρες τη λογαριθμημένη τιμή R/S της i X.Σ.

**If RSVal = 0 Then**

logRSVal = 0

**Else**

logRSVal = Log(RSVal) / Log(10#)

**End If**

'κατασκεύασε τα σημεία του POX διαγράμματος τύπου 1

Cells(lin0 + i + lin, colt + 9).Value = Log(Ni) / Log(10#)

'εξέτασε αν H=1 ή H=0 ή 0<H<1, και δώσε κατάλληλη τιμή στο log(R/S)

**If logRSVal = (Log(Ni) / Log(10#)) - (Log(2) - Log(10#)) Then**

Cells(lin0 + i + lin, colt + 10).Value = (Log(Ni) / Log(10#)) - (Log(2) - Log(10#))

**Else**

**If logRSVal = 0 Then**

Cells(lin0 + i + lin, colt + 10).Value = 0

**Else**

Cells(lin0 + i + lin, colt + 10).Value = logRSVal

**End If**

**End If**

**Next i**

'κατασκεύασε τα σημεία την γραμμών με κλίση 1 και 1/2

Cells(lin0 + tim, colt + 11).Value = (Log(Ni) / Log(10#)) - (Log(2) / Log(10#))

Cells(lin0 + tim, colt + 12).Value = (1 / 2) \* ((Log(Ni) / Log(10#)) - (Log(2) / Log(10#)))

'κατασκεύασε τα σημεία του POX διαγράμματος τύπου 2

Cells(lin0 + tim, colt + 13).Value = Log(Ni) / Log(10#)

**If RSValMean = 0 Then**

logRSValMean = 0

**Else**

logRSValMean = Log(RSValMean / k) / Log(10#)

**End If**

'εξέτασε αν H=1 ή H=0 ή 0<H<1, και δώσε κατάλληλη τιμή στο log(ER/S)

**If logRSValMean = (Log(Ni) / Log(10#)) - (Log(2) - Log(10#)) Then**

Cells(lin0 + tim, colt + 14).Value = (Log(Ni) / Log(10#)) - (Log(2) - Log(10#))

**Else**

**If logRSValMean = 0 Then**

Cells(lin0 + tim, colt + 14).Value = 0

**Else**

Cells(lin0 + tim, colt + 14).Value = logRSValMean

**End If**

**End If**

'υπολόγισε την εκτίμηση της H η οποία προκύπτει από τις k X.Σ.

HEst = Cells(lin0 + tim, colt + 14).Value / (Log(Ni) / Log(10#) - Log(2) / Log(10#))

Cells(lin0 + tim, colt + 3).Value = HEst

MeanHEst = MeanHEst + HEst

RSValMean = 0

seg = seg + k + 1

'τρόπος εξέλιξης του επιπέδου άθροισης Ni

Ni = Ni + 400

tim = tim + 1

**Loop**

'υπολόγισε την εκτίμηση της H

**If** (tim - 1) <> 0 **Then**

MeanHEst = MeanHEst / (tim - 1)

**Else**

MeanHEst = 0

**End If**

Cells(lin0 + tim + 1, col0 + 3).Value = MeanHEst

Cells(lin0 + tim + 1, col0 + 3).Font.Color = RGB(0, 0, 255)

'Range("S1:HZ" & N + 15).ClearContents

MsgBox "Η εκτίμηση της Hurst παραμέτρου είναι " & MeanHEst & ".", , "Μέθοδος R/S Στατιστικής"

**End Sub**

**'A\_S\_I()**

'Δέχεται ως όρισμα το πλήθος N των στοιχείων της χρονολογικής σειράς η οποία θεωρείται ότι βρίσκεται στην 1η και 2η στήλη του τρέχοντος φύλλου εργασίας.

'Δημιουργεί c το πλήθος m-αθροιστικές χρονολογικές σειρές των N/m στοιχείων η κάθε μία. Υπολογίζει την διακύμανση της υπό ανάλυση χρονολογικής σειράς

'(logvarX) και τις διακυμάνσεις των των m - αθροιστικών χρονολογικών σειρών

'(logvariance). Επτι προκύπτουν για κάθε m-αθροιστική χρονολογική σειρά

'εκτιμήσεις του β και H, από την τις σχέσεις  $\beta = (\text{logvarX} - \text{logvariance}) / \log(m)$  και

' $H = 1 - \beta / 2$ . Η τελική εκτίμηση της H προκύπτει από τον μέσο όρο των εκτιμήσεων H

'οι οποίες αντιστοιχούν σε κάθε m-αθροιστική χρονολογική σειρά.

'

**Sub A\_S\_I(N)**

lin0 = 1

col0 = 17

colt = 2

'Range("S1:HZ" & N + 15).ClearContents

m = 400 'αρχικοποίηση του επιπέδου άθροισης

c = 1

k = div(N, m)

Cells(lin0, colt - 1).Value = "Time Series"

'όσο το επίπεδο άθροισης δεν υπερβάνει το ήμισυ των παρατηρήσεων

'της υπό ανάλυσης X.Σ.

**Do While** (m \* 2) <= N

c = c + 1

Cells(lin0, col0 + c).Value = "x(" & m & ")"

'υπολόγισε το j στοιχείο της αθροιστικής χρονολογικής σειράς

'επιπέδου άθροισης m

**For** j = 1 **To** div(N, m)

cell = 0

'άθροισε τα m επόμενα στοιχεία της υπό ανάλυση χρονολογικής σειράς

**For** i = m \* (j - 1) + 1 **To** m \* (j - 1) + m

cell = cell + Cells(lin0 + i, colt).Value

**Next i**

'εμφάνισε το j στοιχείο της αθροιστικής χρονολογικής σειράς  
'επιπέδου άθροισης m  
Cells(lin0 + j, col0 + c).Value = cell / m

**Next j**

'τρόπος εξέλιξης του επίπεδου άθροισης m  
m = m + 400

**Loop**

'επανα-αρχικοποίηση του επιπέδου άθροισης m. Η τιμή που εκχωρείται  
'στο επίπεδο άθροισης, πρέπει να είναι η ίδια με αυτή που εκχωρήθηκε  
'στην αρχή της διαδικασίας

m = 400

i = 1

'Cells(lin0, col0 + c + 1).Value = "log(m)"  
'Cells(lin0, col0 + c + 2).Value = "logVar(x(m))"  
Cells(lin0, col0 + c + 3).Value = "Est. β"  
'Cells(lin0, col0 + c + 4).Value = "Est. H"  
'εμφάνισε τους τίτλους των στηλών  
Cells(lin0, col0 + 2).Value = "H Aggr. Series"  
Cells(lin0, col0 + 7).Value = "Line -1"  
Cells(lin0, col0 + 5).Value = "Var-Time"  
Cells(lin0, col0 + 6).Value = "Diagram."

'βρες τη λογαριθμημένη διακύμανση της υπό ανάλυση χρονολογικής σειράς

If var(colt, lin0 + 1, lin0 + N) < 0 Then

logvarX = Log(var(colt, lin0 + 1, lin0 + N)) / Log(10#)

Else

logvarX = 0

End If

'αρχικοποίηση της εκτίμησης της H

EstH = 0

**Do While** (m \* 2) <= N

'βρες τη διακύμανση της m-αθροιστικής χρονολογικής σειράς  
variance = var(col0 + i + 1, lin0 + 1, lin0 + div(N, m))

'βρες το λογάριθμο της διακύμανσης της m-αθροιστικής χρονολογικής σειράς

If variance = 0 Then

logvariance = 0

Else

logvariance = Log(variance) / Log(10#)

End If

'κατασκεύασε τα σημεία του διαγράμματος διακύμανσης-χρόνου

Cells(lin0 + i, col0 + 5).Value = Log(m) / Log(10#)

'αν υπάρχει ακριβής αυτό-ομοιότητα, δηλ. η διακύμανση της m-αθροιστικής

'χρονολογικής σειράς είναι ίση με την διακύμανση της υπό ανάλυση X.Σ.

If logvariance = logvarX Then

Cells(lin0 + i, col0 + 6).Value = logvarX

Else

'αν δεν υπάρχει καθόλου αυτό-ομοιότητα, δηλ. β=2 το οποίο σημαίνει ότι

'για την διακύμανση της m-αθροιστικής χρονολογικής σειράς logvariance

'ισχύει: logvariance=logvarX-2\*log(m)

If logvariance = (logvarX - 2 \* (Log(m) / Log(10#))) Then

```
Cells(lin0 + i, colt + 6).Value = (logvarX - 2 * (Log(m) / Log(10#)))
Else
    Cells(lin0 + i, colt + 6).Value = logvariance
End If
End If
'κατασκεύασε τα σημεία της γραμμής με κλίση -1
Cells(lin0 + i, colt + 7).Value = logvarX - (Log(m) / Log(10#))
'υπολόγισε το β
Cells(lin0 + i, col0 + c + 3).Value = (logvarX - Cells(lin0 + i, colt + 6).Value) /
Cells(lin0 + i, colt + 5).Value
EstH = EstH + 1 - (Cells(lin0 + i, col0 + c + 3).Value / 2)
'εμφάνισε την εκτίμηση της H για επίπεδο άθροισης m
Cells(lin0 + i, colt + 2).Value = 1 - (Cells(lin0 + i, col0 + c + 3).Value / 2)
i = i + 1
'tρόπος εξέλιξης του επιπέδου άθροισης m. Πρέπει να προσεχθεί αυτός ο
'tρόπος εξέλιξης του επιπέδου άθροισης να είναι ίδιος με αυτόν που
'χρησιμοποιήθηκε νωρίτερα στην ίδια διαδικασία
m = m + 400
Loop
'υπολόγισε την τελική εκτίμηση της H
If (c - 1) <> 0 Then
    EstH = EstH / (c - 1)
Else
    EstH = 0
End If
Cells(lin0 + i + 1, colt + 2).Value = EstH
Cells(lin0 + i + 1, colt + 2).Font.Color = RGB(0, 0, 255)
MsgBox "Η εκτίμηση της Hurst παραμέτρου είναι " & EstH & ". ", "Μέθοδος
Διαγραμμάτων Διακύμανσης - Χρόνου"
End Sub
```

'

**'RS()**

'Επιστρέφει την RS στατιστική των τιμών που ορίζονται στο τρέχον λογιστικό φύλλο  
'από τη στήλη c (1o όρισμα) και τις γραμμές l1 (κάτω όριο-2o όρισμα) και l2 (πάνω  
'όριο-3o όρισμα). Θεωρείται ότι η βασική χρονολογική σειρά βρίσκεται στη 2η στήλη  
'του φύλου εργασίας.

'

```
Function RS(c, l1, l2)
col0 = 2
N = l2 - l1 + 1
If N >= 1 Then
    sd = Sqr(var(col0, l1, l2))
    If sd <> 0 Then
        RS = (maxim(c, l1, l2) - minim(c, l1, l2)) / sd
    Else
        RS = 10 ^ ((Log(N) / Log(10#)) - (Log(2) / Log(10#)))
    End If
    Else
```

```
RS = 0
End If
End Function
```

### 'minim()

'Επιστρέφει τον ελάχιστο των τιμών που ορίζονται στο τρέχον λογιστικό φύλλο από τη στήλη c (1o όρισμα) και τις γραμμές l1 (κάτω όριο-2o όρισμα) και l2 (πάνω όριο-3o όρισμα).

```
Function minim(c, l1, l2)
min = Cells(l1, c).Value
For i = l1 + 1 To l2
    If Cells(i, c).Value < min Then
        min = Cells(i, c).Value
    End If
Next i
minim = min
End Function
```

### 'maxim()

'Έπιστρέφει τον μέγιστο των τιμών που ορίζονται στο τρέχον λογιστικό φύλλο από τη στήλη c (1o όρισμα) και τις γραμμές l1 (κάτω όριο-2o όρισμα) και l2 (πάνω όριο-3o όρισμα).

```
Function maxim(c, l1, l2)
max = Cells(l1, c).Value
For i = l1 + 1 To l2
    If Cells(i, c).Value > max Then
        max = Cells(i, c).Value
    End If
Next i
maxim = max
End Function
```

### 'mean()

'Έπιστρέφει τον μέσο όρο των τιμών που ορίζονται στο τρέχον λογιστικό φύλλο από τη στήλη c (1o όρισμα) και τις γραμμές l1 (κάτω όριο-2o όρισμα) και l2 (πάνω όριο-3o όρισμα).

```
Function mean(c, l1, l2)
If l1 <= l2 Then
    mean = 0
    For i = l1 To l2
        mean = mean + Cells(i, c).Value
    Next i
    mean = mean / (l2 - l1 + 1)
```

**End If**  
**End Function**

'  
**var()**

Έπιστρέφει την διακύμανση (variance) των τιμών που ορίζονται στο τρέχον λογιστικό φύλλο από τη στήλη c (1o όρισμα) και τις γραμμές l1 (κάτω όριο-2o όρισμα) και l2 (πάνω όριο-3o όρισμα).

'  
**Function var(c, l1, l2)**  
**If Not equal(c, l1, l2) Then**  
    **var = 0**  
    **m = mean(c, l1, l2)**  
    **For i = l1 To l2**  
        **var = var + ((Cells(i, c).Value - m) \* (Cells(i, c).Value - m))**  
    **Next i**  
    **If l2 <> l1 Then**  
        **var = var / (l2 - l1)**  
    **Else**  
        **var = 0**  
    **End If**  
    **Else**  
        **var = 0**  
    **End If**  
**End Function**

'  
**equal()**

Έπιστρέφει την τιμή true σε περίπτωση που οι τιμές που ορίζονται στο τρέχον λογιστικό φύλλο από τη στήλη c (1o όρισμα) και τις γραμμές l1 (κάτω όριο-2o όρισμα) και l2 (πάνω όριο-3o όρισμα) είναι όλες ίδιες. Σε αντίθετη περίπτωση έπιστρέφει την τιμή false.

'  
**Function equal(c, l1, l2)**  
**If l1 <= l2 Then**  
    **i = l1**  
    **Do While (Cells(i, c).Value = Cells(i + 1, c).Value) And (i < l2)**  
        **i = i + 1**  
    **Loop**  
    **If i = l2 Then**  
        **equal = True**  
    **Else**  
        **equal = False**  
    **End If**  
    **Else**  
        **equal = True**  
    **End If**  
**End Function**

'  
**'div()**

'Επιστρέφει το πηλίκο διαίρεσης, με διαιρετέο το 10 όρισμα και διαιρέτη το 20.

'  
**Function** div(dr, d)

div = 0

diair = d

**Do While** diair <= dr

div = div + 1

diair = diair + d

**Loop**

**End Function**

'  
**'FindN()**

'Έπιστρέφει τον αριθμό των παρατηρήσεων της χρονολογικής σειράς, η οποία βρίσκεται στη δεύτερη στήλη του φύλλου εργασίας.

'  
**Function** FindN()

**Dim** rowNum **As Long**

rowNum = 2

colNum = 2

**Do While** ActiveSheet.Cells(rowNum, colNum).Value <> ""

rowNum = rowNum + 1

**Loop**

FindN = rowNum - 2

**End Function**

rowNum = 1  
col0 = 1 'στηρίζεται στην αριθμητική σειρά της στήλης X.  
'Τοποθετείται στην πρώτη στήλη της Χ. Ε. και δεν μεταφέρεται  
Cells(rowNum, col0 + 1).Value = "Σύνολο Σειρών"  
Cells(rowNum, col0 + 2).Value = "Παρατηρήσεις"  
Cells(rowNum, col0 + 3).Value = "Ημέρες"  
'Παρατηρεί το πρώτο στοιχείο της στήλης X.  
Cells(rowNum + 1, col0).Value = "#"  
Cells(rowNum + 1, col0 + 1).Value = "#"  
Cells(rowNum + 1, col0 + 2).Value = "#"  
Cells(rowNum + 1, col0 + 3).Value = "#"  
step = 2 'διαχ. μετ. Μ.Σ.  
rowNum = Cells(rowNum + 1, 1).Value + 1 'Ανανεώνεται η στάση στην στήλη Χ.  
rowNum = 1  
'Ενα τέλιο μαρτύρισμα για την αποτελεσματικότητα της έργου μου.  
maxRow = maxRow + 1, minRow = minRow + 1, maxCol = maxCol + 1, minCol = minCol + 1  
**Do While** rowNum < maxRow  
 row0 = row0 + 1

## ΠΑΡΑΡΤΗΜΑ 3

# Ο ΚΩΔΙΚΑΣ ΤΗΣ ΜΑΚΡΟΕΝΤΟΛΗΣ ΔΗΜΙΟΥΡΓΙΑΣ ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ ΑΠΟ ΤΑ ΔΕΔΟΜΕΝΑ ΚΙΝΗΣΗΣ ΤΗΣ ΕΦΑΡΜΟΓΗΣ CFLOWD

'  
**'ExhortTimeSeries()**

'Διαβάζει τον χρόνο έναρξης, τον χρόνο λήξης, τον αριθμό των bytes, και τον αριθμό των πακέτων κάθε ροής και δημιουργεί τις Χ.Σ. των πακέτων και των bytes που αντιστοιχούν στο σύνολο των ροών. Τα στοιχεία των πακέτων των ροών πρέπει να τοποθετηθούν στη 2η στήλη του φύλλου εργασίας, τα στοιχεία των bytes στην 3η, οι χρόνοι έναρξης στην 4η και μάλιστα ταξινομημένοι σε αύξουσα σειρά και οι χρόνοι λήξης στην 5η. Οι Χ.Σ. που προκύπτουν τοποθετούνται στις στήλης 6-7.

**Sub ExportTimeSeries()**

j = 1

i = j

n = FindN() 'εντόπισε το πλήθος των ροών

lin0 = 1

col0 = 7 'στήλη στην οποία τοποθετείται η Χ.Σ των πακέτων

'Τοποθέτησε τίτλους για τις Χ.Σ. που θα δημιουργηθούν

Cells(lin0, col0 - 1).Value = "Time Series"

Cells(lin0, col0).Value = "Packets"

Cells(lin0, col0 + 1).Value = "Bytes"

'Τοποθέτησε το πρώτο στοιχείο των δύο Χ.Σ.

Cells(lin0 + 1, col0).Value = 0

Cells(lin0 + 1, col0 + 1).Value = 0

Cells(lin0 + 1, col0 - 1).Value = Cells(lin0 + 1, 4).Value

step = 2 'βήμα της Χ.Σ.

t = Cells(lin0 + 1, 4).Value + step 'τρέχουσα χρονική στιγμή της Χ.Σ.

nobs = 1

'όσο το t είναι μικρότερο από τον μεγαλύτερο χρόνο λήξης όλων των ροών

maxt = maxim(5, lin0 + 1, lin0 + n)

**Do While** t <= maxt

nobs = nobs + 1

```
obsp = 0
obsb = 0
'Λαμβάνοντας υπ' όψη όλες τις χρονικές επικαλύψεις των
'ροών τη χρονική στιγμή t, υπολόγισε τον συνολικό αριθμό
'πακέτων και των bytes για τη συγκεκριμένη χρονική στιγμή
Do While (Cells(lin0 + i, 4).Value < t) And (i <= n)
If (t <= Cells(lin0 + i, 5).Value) Then
    obsp = obsp + Find(t, i, 2)
    obsb = obsb + Find(t, i, 3)
End If
    i = i + 1
Loop
    i = j
'τοποθέτησε τις τιμές των δύο X.Σ.
    Cells(lin0 + nobs, col0).Value = obsp
    Cells(lin0 + nobs, col0 + 1).Value = obsb
    Cells(lin0 + nobs, col0 - 1).Value = t
    t = t + step
Loop
End Sub
```

#### 'Find()

Έπιστρέφει τον αριθμό των πακέτων (αν n=2) ή τον αριθμό 'των bytes (αν n=3) που  
'αντιστοιχούν στην i-οστή ροή στον χρόνο t. Υποθέτουμε ότι στον χρόνο έναρξης  
'κάθε ροής ts αντιστοιχούν 0 πακέτα και 0 bytes. Επίσης, υποθέτουμε ότι στον χρόνο  
'λήξης κάθε ροής tf αντιστοιχούν p το πλήθος πακέτα και b το πλήθος bytes, όπου p  
'και b το πλήθος των πακέτων και των bytes της κάθε ροής. Τέλος, υποθέτουμε ότι ο  
'αριθμός των πακέτων xp που αντιστοιχούν στο χρόνο t όπου ts<t<tf, είναι  
'xp=p\*t/(tf-ts). Αντίστοιχα, για τον αριθμό των bytes που αντιστοιχούν στο χρόνο t  
'όπου ts<t<tf, ισχύει xb=b\*t/(tf-ts).

#### Function Find(t, i, n)

```
lin0 = 1
Find = (Cells(lin0 + i, n).Value * (t - Cells(lin0 + i, 4))) / (Cells(lin0 + i, 5) -
Cells(lin0 + i, 4))
```

#### End Function

#### 'FindN()

Έπιστρέφει τον αριθμό των παρατηρήσεων της χρονολογικής σειράς, η οποία  
'βρίσκεται στη δεύτερη στήλη του φύλλου εργασίας.

#### Function FindN()

```
Dim rowNum As Long
rowNum = 2
colNum = 2
```

**Do While** ActiveSheet.Cells(rowNum, colNum).Value <> ""

```
rowNum = rowNum + 1
```

## Loop

FindN = rowNum - 2

**End Function**

'maxim()

Έπιστρέφει τον μέγιστο των τιμών που ορίζονται στο τρέχον λογιστικό φύλλο από τη 'στήλη c (1o όρισμα) και τις γραμμές 11 (κάτω όριο-2o όρισμα) και 12 (πάνω όριο-3o όρισμα).

**Function** maxim(c, l1, l2)

Max = Cells(11, c).Value

**For i = 11 + 1 To 12**

If Cells(i, c).Value > Max Then

**Max = Cells(i, c).Value**

End If

Next i

maxim = Max

**End Function**

## ΠΑΡΑΡΤΗΜΑ 4

### Ο ΚΩΔΙΚΑΣ ΤΗΣ ΜΑΚΡΟΕΝΤΟΛΗΣ ΔΗΜΙΟΥΡΓΙΑΣ ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ ΑΠΟ ΤΑ ΔΕΔΟΜΕΝΑ ΚΙΝΗΣΗΣ ΤΗΣ ΕΦΑΡΜΟΓΗΣ FLOWCOLLECTOR

#### 'CiscoTimeSeries()

'Διαβάζει στην 1η στήλη το δευτερόλεπτο για το οποίο υπάρχουν στοιχεία ροών στις 'τρεις επόμενες στήλες. Συγκεκριμένα στη 2<sup>η</sup> στήλη υπάρχει το port πηγής της ροής, 'στην 3η στήλη υπάρχει ο συνολικός αριθμός πακέτων της ροής ενώ στην 4η στήλη 'υπάρχει ο συνολικός αριθμός bytes της ροής. Ένα παράδειγμα μορφής εισόδου είναι 'το εξής:

' 1η στήλη 2η στήλη 3η στήλη 4η στήλη

	41	SPort	Pkts	Octets
'		80	2	734
'	43	SPort	Pkts	Octets
'		53	2	80
'		80	2	80
'		80	3	1101
'		80	2	502
'	44	SPort	Pkts	Octets
'		53	2	102
'		53	4	1628
'		53	5	1664

'Οι Χ.Σ. που δημιουργούνται βάσει των στηλών 1-4, τοποθετούνται στις στήλες 6-9.  
'Συγκεκριμένα στην 6η στήλη τοποθετείται η Χ.Σ. των πακέτων, στην 7η η Χ.Σ. των 'bytes, στην 8η η Χ.Σ. των πακέτων της κίνησης του ηλεκτρονικού ταχυδρομείου (SPort = 80) και στην 9η η Χ.Σ. των ψηφιοσυλλαβών της κίνησης του ηλεκτρονικού ταχυδρομείου.

**Sub CiscoTimeSeries()**

emp = Cells(1, 5).Value

lin = 0

**For i = 2 To 65000** 'για τα πρώτα 65000 στοιχεία της 1ης στήλης

'αν στην επόμενη γραμμή υπάρχουν στοιχεία ροών

**If Cells(i, 3).Value = "Pkts" Then**

**If lin = 0 Then** 'αν βρισκόμαστε στην 1η γραμμή

diaf = 0 'tότε η διαφορά με το προηγούμενο δευτερόλεπτο είναι 0

**Else**

curnt = Cells(i, 1).Value

'βρες τη διαφορά με το προηγούμενο δευτερόλεπτο

diaf = curnt - prev - 1

**If diaf < -1 Then**

diaf = diaf + 60

**End If**

**End If**

'τοποθέτησε μηδενικές τιμές στις παρατηρήσεις των Χ.Σ. που

'αντιστοιχούν σε δευτερόλεπτα για τα οποία δεν υπάρχει κίνηση

**For k = 1 To diaf**

lin = lin + 1

Cells(lin, 6).Value = 0

Cells(lin, 7).Value = 0

Cells(lin, 8).Value = 0

Cells(lin, 9).Value = 0

**Next k**

prev = Cells(i, 1).Value

i = i + 1

'αν το τρέχον δευτερόλεπτο δεν είναι ίδιο με το δευτερόλεπτο

'της προηγούμενης γραμμής, τότε μηδένισε τους μετρητές

**If diaf <> -1 Then**

lin = lin + 1

```
pkts = 0  
octets = 0  
mpkts = 0  
mocets = 0
```

**End If**

'όσο υπάρχουν στοιχεία ροής για το συγκεκριμένο δευτερόλεπτο

**Do While Cells(i, 3).Value <> emp**

'αν πρόκειται για ροή ηλεκτρονικού ταχυδρομείου,

'ενημέρωσε τους μετρητές πακέτων και ψηφιοσυλλαβών

**If Cells(i, 2).Value = 80 Then**

```
mpkts = mpkts + Cells(i, 3).Value
```

```
mocets = mocets + Cells(i, 4).Value
```

**End If**

'ενημέρωσε τους μετρητές των πακέτων και ψηφιοσυλλαβών

```
pkts = pkts + Cells(i, 3).Value
```

```
octets = octets + Cells(i, 4).Value
```

```
i = i + 1
```

**Loop**

'τοποθέτησε τις παρατηρήσεις των Σ.Σ.

```
Cells(lin, 6).Value = pkts
```

```
Cells(lin, 7).Value = octets
```

```
Cells(lin, 8).Value = mpkts
```

```
Cells(lin, 9).Value = mocets
```

**End If**

**Next i**

**End Sub**

## ΒΙΒΛΙΟΓΡΑΦΙΑ

### ΒΙΒΛΙΑ

- [BER94] Beran J., Statistics for Long-Memory Processes, Chapman & Hall, 1994.
- [HEL96] Held, LAN Management with SNMP and RMON, Wiley, 1996.
- [STAL97] Stallings, High-Speed Networks, Prentice Hall, 1997.
- [ROS91] Rose M. T., The Simple Book, Prentice Hall, 1991.

### ΕΠΙΣΤΗΜΟΝΙΚΑ ΑΡΘΡΑ

- [GROS96] M. Grossglauser, and J. Bolot, "On the Relevance of Long-range Dependence in Network Traffic." Proceeding, SIGCOMM'96, August 1996.
- [JAI86] R. Jain and S. A. Routhier, "Packet trains: Measurements and a New Model for Computer Network Traffic," IEEE Journal on Selected Areas Communications, vol. SAC-4, pp. 986-995, 1986.
- [LELA94] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the Self-Similar Nature of Ethernet Traffic (Extended Version)," IEEE/ACM Transactions on Networking, vol. 2, pp. 1-15, 1994.
- [NORR94] I. Norros, "A storage Model with Self-similar input" Queueing Systems, Vol. 16, 1994.
- [PRU95] P. Pruthi and a A. Erramilli, "Heavy-Tailed On/Off Source Behavior and Self-Similar Traffic," in Proc. ICC '95, Seattle, WA, 1995, pp. 445-450.
- [RYU96] B. Ryu, and A. Elwalid, "The Importance of Long-range Dependence of VBR Video Traffic in ATM Traffic Engineering: Myths and Realities." Proceeding, SIGCOMM'96, August 1996.
- [RYU97] B. Ryu, and S. Lowen, "Point Process Approaches for Modeling and Analysis of Self-Similar Traffic, Part II-Applications." Proceedings, International Conference on Telecommunications Systems, Modeling, and Analysis, March 1997.
- [TAQ96] M. S. Taqqu and V. Teverovsky, and W. Willinger, "Is Network Traffic Self-Similar or Multifractal?" Fractals, 1996.
- [WILL94] W. Willinger, M. S. Taqqu, R. Sherman, and D. V. Wilson, "Self-similar Traffic Modelling for High Speed Networks." ConneXions,



November 1994.

- [WILL97] W. Willinger, M. S. Taqqu, R. Sherman, and D. V. Wilson, "Self-similarity through High-Variability: Statistical Analysis of Ethernet LAN Traffic at the Source Level," IEEE/ACM Transactions on Networking, vol. 5, pp. 71-86, 1997.
- [WILL97b] W. Willinger, M. S. Taqqu, and R. Sherman, "Proof of a fundamental result in self-similar traffic modeling," Computer Communication Review, vol. 27, pp. 5-23, 1997.

## REQUESTS FOR COMMENTS (RFCs)

- [RFC1757] RFC 1757, "Remote Network Monitoring Management Information Base", February 1995.
- [RFC2021] RFC 2021, "Remote Network Monitoring Management Information Base Version 2 using SMIv2", January 1997.

## ΠΗΓΕΣ ΑΠΟ TO WORLD WIDE WEB

- <http://engr.ans.net/cflowd/>
- <http://www.cisco.com/warp/public/732/netflow/index.html>



